



41^{ÈME} ÉDITION - LA ROCHELLE

30 MAI - 2 JUIN

INFORSID 2023

INFormatique des ORganisations et Systèmes d'Information et de Décision

SI ET RESPONSABILITÉ SOCIÉTALE ET ENVIRONNEMENTALE



Président du Comité de Programme
Christophe PONSARD
CETIC, Université de Namur

Président du Comité d'Organisation
Cyril FAUCHER
L3i, La Rochelle Université

@INFORSID2023



aYaline

nēosoft



Préface

Les Systèmes d'Information (SI) sont apparus avec l'humanité et sa soif de collecter, stocker, traiter et diffuser l'information. Nos dernières décennies ont connu un développement technologique exponentiel dopant nos SI qui sont devenus capables du partage instantané d'information au niveau planétaire, de la gestion intégrée de grandes quantités de données, d'apprentissage automatique et de diverses formes d'intelligence artificielle. Notre vie quotidienne et professionnelle en a tiré de nombreux bénéfices mais au prix de conséquences sociétales croissantes en matière d'emploi, de fracture numérique, de dés-information et d'impacts environnementaux.

Dans ce contexte, la gestion responsable du numérique s'est rapidement imposée à moi comme un thème central pour cette 41^{ème} édition du congrès INFORSID qui se tient à La Rochelle, une ville pionnière et parmi les plus investies de France en matière de soutenabilité. En tant que président du comité de programme, j'ai le grand plaisir de vous présenter les actes qui résultent du travail de sélection de ce comité mais surtout du vôtre, la communauté INFORSID qui à largement répondu à cette thématique comme le reflète le programme au travers de sessions spécifiques (durabilité, applications sociétales) ainsi qu'au sein de la quasi-totalité de nos sessions traditionnelles telles que l'ingénierie des exigences, les modèles et méthodes, l'interaction avec les utilisateurs, la sécurité et les données !

Sur 40 soumissions initiales, 29 ont finalement été acceptées avec une belle diversité puisque 11 nationalités sont représentées (par ordre d'importance : France, Belgique, Espagne, Tunisie, Maroc, Cameroun, Pays-Bas, Autriche, Australie, USA) et avec de l'ordre du tiers d'auteurs. Le programme se compose de 6 articles longs détaillant des contributions élaborées, de 8 articles courts focalisés sur des travaux en cours ou sur des applications des SI à des problématiques sociétales. 10 articles sont aussi issus de conférences et revues internationales de 2022 avec comme objectif de les faire partager avec la communauté INFORSID. Enfin, 5 démonstrations sont également introduites lors des sessions avec la possibilité d'interagir plus activement avec leurs auteurs sur des stands proposés lors d'un temps de midi.

Le programme est complété par plusieurs autres interventions donnant des éclairages spécifiques sur la thématique de la responsabilité sociétale des SI. Durant le premier exposé invité, le professeur Eric Dubois du LIST vous présentera, avec l'appui d'un cas d'utilisation luxembourgeois, les rôles et défis pouvant être relevés par l'utilisation d'un jumeau numérique dans le domaine de la transition énergétique d'un district urbain. Dans un second exposé invité, le professeur Sjaak Brinkkemper de l'université d'Utrecht vous présentera les possibilités d'un système de génératif de rapports basé sur des données

multimodales dans le contexte de SI hospitaliers. Enfin, et pour faire le lien avec notre ville hôte, David Berthiaud, directeur de la transformation numérique à La Rochelle, vous présentera une infrastructure numérique pour piloter, par la donnée, le projet de territoire pour atteindre la neutralité carbone en 2040. Anaïs Schmitt coordinatrice de projet Urban & Coastal Lab de La Rochelle (UCLR) vous présentera une plateforme de mutualisation de données de projets ciblant les zones littorales, depuis la collecte des données jusqu'à un système d'aide à la décision.

La conférence vous propose aussi plusieurs ateliers avec une large dimension participative sur des sujets à nouveau très en phase avec notre thématique: les systèmes d'information et de décision et démocratie, l'enseignement des SI Responsables, égalité en actions! et l'exploration des traces dans un monde du tout numérique. Une introduction au concept de Green Agile est aussi au programme afin de vous donner une envie d'expérimenter avec de nouvelles pratiques plus responsables en entreprise, en équipe ou en management! Enfin, un prix de thèse est aussi décerné à la meilleure thèse soutenue depuis le début de l'année dans les thématiques de l'association.

Pour clôturer cette préface, je tiens à remercier de nombreuses personnes:

- les membres du bureau qui m'ont fait confiance pour l'organisation scientifique et accompagné tout au long du processus menant à ces jours de rassemblement, et tout particulièrement Agnès Front;
- les auteurs pour leur travaux explorant et validant le champ des SI sous des angles sans cesse renouvelés;
- les membres du comité de programme pour leurs revues de qualité et riches en retours très appréciés des auteurs, parmi eux les « sages » du conseil de programme qui ont permis d'aboutir à ce programme;
- les porteurs et porteuses d'ateliers pour leur investissement dans l'animation de la communauté;
- les conférenciers invités pour la mise en perspective qu'ils nous offrent au moyen de leur expertise et de leur expérience;
- Cyril Faucher pour avoir mené toute l'organisation à bon port;
- vous tous qui participez et faites vivre la communauté INFORSID.

Très belle conférence à tou.te.s!

Christophe PONSARD
Président du comité de programme

Comités

Le comité de la 41ème édition d'INFORSID est composé par les responsables de l'organisation ainsi que les membres du comité de programme et les membres du conseil du comité de programme. Le congrès INFORSID 2023 est organisé par le laboratoire d'informatique L3i de La Rochelle Université.

Comité de programme

Pascal ANDRE, LS2N, Nantes Université

Pierre-Emmanuel ARDUIN, DRM, Université Paris-Dauphine

Cécile CAYERE, LIUPPA, Université de Pau et des Pays de l'Adour

Ronan CHAMPAGNAT, L3i, La Rochelle Université

François CHAROY, LORIA, Université de Lorraine

Max CHEVALIER, IRIT, Université Paul Sabatier

Mario CORTES-COMAX, LIG, Université Grenoble Alpes

Cédric DU MOUZA, CEDRIC, CNAM

Annabelle GILLET, LIB, Université de Bourgogne

Samedi HENG, UCLouvain et Service fédéral des Pensions

André MIRALLES, IRSTEA

Damien MONDOU, L3i, La Rochelle Université

Anthony SIMONOFSKI, FSESG, Université de Namur

Lynda TAMINE, IRIT, Université Paul Sabatier

Mounir TOUZANI, INRAE Toulouse

Nathalie VALLES-PARLANGÉAU, LIUPPA, Université de Pau et des Pays de l'Adour

Marlène VILLANOVA, LIG, Université Grenoble Alpes

Conseil du comité de programme

Lylia ABROUK, LIB, Université de Bourgogne

Khalid BENALI, LORIA, Université de Lorraine

Omar BOUSSAID, ERIC, Université Lumière Lyon 2

Cécile FAVRE, Université Lyon 2

Agnès FRONT, LIG, Université Grenoble Alpes

Robert VISEUR, Université de Mons

Akram IDANI, LIG, Université Grenoble Alpes

Président du comité de programme: Christophe PONSARD, CETIC, Université de Namur

Comité d'organisation

Damien MONDOU, L3i, La Rochelle Université
Muhammad Muzzamil LUQMAN, L3i, La Rochelle Université
Dominique LIMOUSIN, L3i, La Rochelle Université
Nathalie RENAUDIN, L3i, La Rochelle Université
Ahmed HAMDY, L3i, La Rochelle Université
Karell BERTET, L3i, La Rochelle Université
Petra GOMEZ, L3i, La Rochelle Université
Antoine DOUCET, L3i, La Rochelle Université
Mélanie MALINAUD, L3i, La Rochelle Université
Mourad RABAH, L3i, La Rochelle Université
Cyrille SUIRE, L3i, La Rochelle Université
Jamal MALKI, L3i, La Rochelle Université
Christophe DEMKO, L3i, La Rochelle Université
Alain BOUJU, L3i, La Rochelle Université
Ronan CHAMPAGNAT, L3i, La Rochelle Université
Laetitia BARREAU, L3i, La Rochelle Université
Rouaa WANNOUS, L3i, La Rochelle Université
Cécile CAYERE, LIUPPA, Université de Pau et des Pays de l'Adour

Infographie: Chloé GOUDOUNESQUE, Dir. des Pédagogies Innovantes, La Rochelle Université

Président du comité d'organisation: Cyril FAUCHER, L3i, La Rochelle Université

Ateliers

Systèmes d'information et de décision et démocratie

Maryse SALLES, IRIT, Université Toulouse 1 Capitole
Raphaëlle BOUR, IRIT, Université Toulouse 1 Capitole
Nathalie VALLES-PARLANGÉAU

Enseigner les SI responsables

Rébecca DENECKERE, Université Paris 1 Panthéon-Sorbonne
Elena KORNYSHOVA, CNAM
Nathalie VALLES-PARLANGÉAU, IUT de Bayonne et du Pays-Basque

Egalité en actions !

Cécile FAVRE, Université Lyon 2
Claudia RONCANCIO, Ensimag, Université Grenoble Alpes

Exploration des traces dans un monde du tout numérique : enjeux et perspectives

Damien MONDOU, L3i, La Rochelle Université
Ronan CHAMPAGNAT, L3i, La Rochelle Université
Didier VYE, UMR LIENSs CNRS et La Rochelle Université
Cyril FAUCHER, L3i, La Rochelle Université

Keynotes

Jumeaux numériques et systèmes d'information pour des villes "intelligentes"

Orateur : Prof. Eric Dubois, Luxembourg Institute of Science and Technology

Résumé: Le concept de jumeau numérique est bien connu depuis de nombreuses années avec un déploiement dans différents domaines d'application. Plus récemment, un agenda en matière de recherche et innovation est apparu concernant leur application au niveau des villes et communautés.

Après avoir introduit le concept de jumeau numérique en général, puis au niveau plus spécifique des villes et communautés, l'exposé s'attachera à développer les thèmes suivants :

- Rôle sociétal et économique, et défis environnementaux et technologiques associés à l'adoption d'un jumeau numérique au niveau d'une ville et d'une région
- Agenda de recherche en matière de jumeau numérique et système d'information en liaison avec la stratégie UE
- Illustration au travers d'un cas d'utilisation Luxembourgeois du rôle d'un jumeau numérique dans le domaine de la transition énergétique d'un district urbain

Biographie : Avant de rejoindre le Luxembourg Institute of Science and Technology - LIST- (anciennement CRP Henri Tudor) en 2000, le Prof. Dr. Eric Dubois a occupé des postes dans le secteur privé (Philips Research Laboratory, S.W.I.F.T) et dans la recherche publique (Universités de Lorraine et de Namur). Au LIST, jusqu'en 2020, Eric Dubois était le directeur du département "IT for Innovative Services", il est maintenant impliqué dans la nouvelle initiative stratégique Nation Wide Digital Twin au LIST. Ses domaines d'intérêt en R&D comprennent : les systèmes d'information et de services, l'ingénierie des exigences, l'analyse des risques de sécurité et la conformité, les architectures d'entreprises. Il est (co-)auteur de plus de 120 publications dans ces domaines et a supervisé 6 thèses de doctorat.

Reducing the Administrative Burden in Healthcare: Speech and Action Recognition in Generative Medical Reporting

Orateur : Prof. Sjaak Brinkkemper, Utrecht University, the Netherlands

Résumé : The administrative burden in routine healthcare processes is high, but for communication among care providers the reporting about patient consultations in hospital information systems (HIS) is very essential. In order to reduce this burden we have started the Care2Report research program that aims at a generative HIS reporting based on multimodal (audio, video, bluetooth) recording of a consultation, followed by knowledge representation, ontological conversation interpretation, and finally the generation and uploading of the report in the patient medical record in the HIS. Novel generative pre-trained transformers (e.g. ChatGPT, YouChat, ChatSonic) provide innovative means for conversation dialogue interpretation and abstractive summarization.

In this keynote I will present the aims and goals of the Care2Report research program, the various linguistic intelligence pipelines, its current functional and technical architecture, and the achievements so far. The linguistic pipeline research will be illustrated by (i) a generic method for the design of trusted cloud pipelines in medical reporting, (ii) the generation of medical guideline ontologies for the matching of the consultation audio transcript, and (iii) the automated pseudonimisation of privacy related data by means of named entity recognition. We end with an outlook of the current research projects and experiments in healthcare institutions.

Biographie : Prof. Sjaak Brinkkemper is full professor of Software Production at the Department of Information and Computing Sciences of Utrecht University, the Netherlands. He leads a group of about twenty-five researchers specialized in product software development and software entrepreneurship. The main research themes of the group are methodology of software production, implementation and adoption, and techno-economic aspects of the software industry. Brinkkemper has published about 10 books and over 300 papers in his research interests: medical informatics, health-care software production, requirements engineering, software architecture, and method engineering.

Une plateforme numérique pour piloter par la donnée le projet de territoire pour atteindre la neutralité carbone en 2040

Orateur : David Berthiaud, Directeur de la transformation numérique, ville et agglomération de La Rochelle

Résumé : Le Consortium du programme « La Rochelle Territoire Zéro Carbone » déploie une infrastructure de gestion des données du territoire et d'un portail web pour les exploiter. Cet outil aidera à construire la trajectoire pour atteindre l'objectif de neutralité carbone à l'horizon 2040. Cette plateforme logicielle appelée Terreze permettra de collecter, traiter, analyser et partager les informations utiles et utilisables entre et pour les acteurs territoriaux. Un certain nombre d'exigences sont mis en œuvre : l'éco-conception, l'open source, l'interopérabilité, la scalabilité, la sécurité et le privacy by design.

Urban & Coastal Lab La Rochelle (UCLR) depuis la collecte des données de la recherche jusqu'à un système d'aide à la décision

Oratrice : Anaïs Schmitt, chief data officer et coordinatrice du projet Urban and Coastal Lab La Rochelle, La Rochelle Université

Résumé : UCLR est un programme de La Rochelle Université qui a pour objectif de mettre en place une plateforme de données et de modèles, issus des différents laboratoires de recherche de La Rochelle Université. La mutualisation des données en un lieu unique est une opportunité de faire émerger des liens entre disciplines. L'accessibilité, la mise en cohérence et la valorisation des données et modèles peuvent aider à améliorer la compréhension de notre socio-écosystème et in fine, aider à sa gestion. Pour ce projet d'envergure les activités scientifiques ont été lancées (2022-2023) via 7 projets d'études hétérogènes, mis en place dans le but d'initier la collecte de données, de mettre en place une méthode de catalogue et d'approfondir les démarches scientifiques qui serviront à proposer des outils de valorisation des données des zones littorales.

Table des matières

Ingénierie des exigences

Long	Conception d'ontologies formelles : une approche par élicitation de connaissances <i>Ouassila Labbani Narsis et Christophe Nicolle</i> 1
Long	Exploitation de scénarios prospectifs pour l'analyse des exigences - Application à la conception des avions futurs <i>Brieuc Danet</i> 17
Int	L'avion et sa chaîne d'assemblage : des premières exigences à une conception globale <i>Anouck Chan, Anthony Fernandes Pires, Thomas Polacsek et Stephanie Roussel</i> 33
Int	Eliciter, raffiner et attribuer des buts aux bons acteurs à partir d'objectifs de haut niveau <i>Anouck Chan, Anthony Fernandes Pires et Thomas Polacsek</i> 35

Durabilité

Int	Outils de calcul de l'empreinte TIC des organisations : adaptation d'une étude européenne <i>Guillaume Bourgeois, Benjamin Duthil et Vincent Courboulay</i> 37
Démo	Un simulateur pour une adoption de Cloud plus éco-consciente <i>Evgeny Vasyuk, Arthur Vervaeet, Mar Callau-Zori et Raja Chiky</i> 39
Démo	RADIANCE: A tool for software behavior design and energy consumption categorization <i>Jorge Andrés Larracochea, Philippe Roose, Sergio Ilarri, Sébastien Laborie et Yudith Cardinale</i> .. 43

Modèles et méthodes

Long	CircuSChain: a Method to Continuously Improve Circularity in Supply Chains <i>Asiye Kurt, Cortes-Cornax Mario, Agnès Front, Cung Van-Dat et Mangione Fabien</i> 49
Court	Vers un Profil Environnemental des Processus Métier <i>Mario Cortes-Cornax, Claudia Roncancio et Raquel Oliveira</i> 65
Int	Faciliter la migration vers une architecture microservices : Approche basée sur la rétroingénierie dirigée par les modèles et l'apprentissage par renforcement <i>Mohammadhadi Dehghani, Shekoufeh Kolaoudou-Rahimi, Massimo Tisi et Dalila Tamzalit</i> 71

Applications sociétales

Court	Méthodologie d'accompagnement à la transition vers l'économie circulaire sur base d'un SI responsable <i>Louise Noël et Christophe Ponsard</i> 73
Court	Vers une optimisation de la diffusion de l'information dans une ville intelligente <i>Maiika Grim Yefsah, Mohamed Chachoua et Lea Jeantet</i> 79

Court	Aide alimentaire : digital et dignité ? Trois études de cas de systèmes d'information au service de la dignité des personnes <i>Pierre-Emmanuel Arduin et Doudja Saïdi-Kabeche</i>	85
Court	Vers la conceptualisation du périmètre de la Responsabilité Numérique des Entreprises <i>Aïste Rugeviciute</i>	91
Court	OntoFiC : une ontologie pour la détection de fraude financière et la modélisation des comportements des clients <i>Lylia Abrouk, Hamza Chergui, Hamid Ahaggach, Benjamin Auger et Dominique Cheron</i>	97

Sécurité

Int	Caractérisation des fausses nouvelles : une approche basée sur la modélisation conceptuelle <i>Belloir Nicolas, Wassila Ouerdane et Oscar Pastor</i>	103
Démo	ST-FRAUD : un outil de détection de transactions frauduleuses <i>Hamza Chergui, Nadine Cullot, Nicolas Cabioch et Lylia Abrouk</i>	105

SI & utilisateurs 1

Long	La Sécurité Psychologique Et Ses Conséquences Sur Les Equipes Scrum <i>Théo Moussaoui et Rebecca Deneckere</i>	111
Démo	Enrichissement et visualisation de trajectoires de visiteurs dans un musée <i>Jérémy Richard, Cyril Faucher et Karell Bertet</i>	127

SI & utilisateurs 2

Long	Éthique de la gestion du consentement au traitement de données personnelles : une analyse au prisme des dark patterns <i>Robert Viseur</i>	133
Long	Proposition d'une architecture utilisant le trace clustering pour recommander un parcours d'apprentissage <i>Wiem Hachicha, Leila Ghorbel, Ronan Champagnat, Mourad Rabah, Samuel Nowakowski et Corinne Amel Zayani</i>	149
Court	COVID-19 et (dés)information : l'open data face à ses limitations <i>Robert Viseur</i>	165

Réseaux Sociaux

Int	ChouBERT : Pré-entraînement d'un modèle de langue française pour le Crowdsensing avec des Tweets dans un contexte phytosanitaire <i>Shufan Jiang, Rafael Angarita, Stéphane Cormier, Julien Orensanz et Francis Rousseaux</i>	171
Int	Evaluer la polarisation des réseaux sociaux numériques par l'étude des frontières de communautés – Approche ERIS <i>Alexis Guyot, Annabelle Gillet, Eric Leclercq et Nadine Cullot</i>	173

SI & données

- Int** Interrogation des graphes temporels de propriétés
Landy Andriamampianina, Franck Ravat, Jiefu Song et Nathalie Vallès-Parlangeau 175
- Court** Alignement incrémental de données dans des environnements dynamiques
Oumaima El Haddadi, Max Chevalier, Bernard Dousset, Ahmad El Allaoui, Anass El Haddadi et Olivier Teste..... 177
- Démo** Un moteur de recherche d'événements pour explorer la presse numérique ou historique
Guillaume Bernard et Thomas Blot 183

Applications intelligentes

- Int** Exploration des Modèles de Processus Contextuels à l'Aide des Données de Capteurs : un Cas d'Activités Quotidiennes dans une Maison Intelligente
Ramona Elali, Rebecca Deneckere, Elena Kornyshova et Camille Salinesi 189
- Int** Du Concept de Vie Intelligente (Smart Life) à celui de l'Ingénierie de la Vie Intelligente (Smart Life Engineering)
Elena Kornyshova, Rebecca Deneckere, Eric Gressier-Soudan, John Murray et Sjaak Brinkkemper
191

Conception d'ontologies formelles : une approche par élicitation de connaissances

Ouassila Labbani Narsis, Christophe Nicolle

*CIAD UMR 7533, Université de Bourgogne, UB, F-21000 Dijon, France
{ouassila.narsis,cnicolle}@u-bourgogne.fr*

RÉSUMÉ. La construction d'ontologies formelles reste une démarche complexe pour les entreprises. Dans la littérature, cette construction se base sur les connaissances techniques et les spécialités des experts du domaine, sans plus de détails sur les méthodologies employées. Les éventuels problèmes de désaccords entre experts, d'expression des connaissances implicites rarement verbalisées, de qualification du résultat par des cas d'usages, ou simplement d'adhésion du groupe d'experts, restent actuellement sans réponse. Cet article propose une démarche méthodologique basée sur l'élicitation des connaissances pour la conception d'ontologies formelles, consensuelles et partagées. L'approche proposée est testée expérimentalement sur des projets de collaborations industrielles dans le domaine de la manufacture (associant des sources de connaissances d'entreprises multinationales) et dans le domaine de la viticulture (associant des connaissances explicites et des savoirs implicites acquis par l'observation).

ABSTRACT. Building formal ontologies remains a complex process for companies. In the literature, this process is based on the technical knowledge and expertise of domain experts, without further details on the used methodologies. Possible problems of disagreements between experts, expression of implicit knowledge related to high level know-how rarely verbalized, qualification of results by use cases, or simply adhesion of the group of experts, remain currently unsolved. This paper proposes a methodological approach based on knowledge elicitation for the conception of formal, consensual, and shared ontologies. The proposed approach is experimentally tested on industrial collaboration projects in the field of manufacturing (associating knowledge sources from multinational companies) and in the field of viticulture (associating explicit knowledge and implicit knowledge acquired through observation).

MOTS-CLÉS : Ingénierie de connaissances, Élicitation de connaissances, Gestion des connaissances, Conception d'ontologie.

KEYWORDS: Knowledge engineering, Knowledge elicitation, Knowledge management, Ontology design.

1. Introduction et motivation

La transformation numérique représente un défi technologique majeur pour les entreprises. Elle se traduit par l'introduction de nouveaux outils et technologies trop souvent en rupture avec les usages. Cette transformation s'accompagne souvent d'une modification de l'organisation, des processus, mais aussi du rapport du salarié avec sa fonction. Ce dernier point peut sérieusement affecter la préservation et la transmission des savoir-faire acquis par expérience depuis plusieurs années. Le savoir-faire d'une entreprise est sa principale richesse et la captation numérique de cette richesse devient un enjeu stratégique. Cette captation est un processus articulé en plusieurs étapes : il faut identifier les savoir-faire, les recueillir, les homogénéiser, les combiner puis les formaliser. Cette dernière étape permet d'obtenir une description numérique compréhensible par l'humain et sans ambiguïté pour la machine. Dans ce processus, l'utilisation de la modélisation sémantique joue un rôle très important, d'où l'usage de formalismes et de langages basés sur la logique tels que les *ontologies*. En informatique, une ontologie désigne une description formelle et explicite des concepts et de leurs relations. Décrite en logique de description, son rôle est de capturer les connaissances du domaine de manière générique et d'en fournir une compréhension partagée et communément acceptée (Maedche *et al.*, 2000). La construction d'ontologies permet d'une part d'atteindre un consensus dans la compréhension commune des connaissances, et d'autre part, de faciliter la mise en place d'un modèle formel représentatif et flexible, indispensable à la description numérique du savoir-faire d'une entreprise. Néanmoins, créer une ontologie est un processus complexe. Au-delà d'une compréhension du domaine concerné et des aspects techniques de l'ingénierie ontologique, il requiert une volonté de collaboration qui dépasse les frontières disciplinaires (Kanza *et al.*, 2018). Malgré les efforts pour le développement de méthodologies et outils d'aide à la construction d'ontologies (Gruber, 1995 ; Jones *et al.*, 1998 ; Missikoff, Schiappelli, 2005), la démarche reste empirique sans méthode normalisée. Les approches existantes ne prennent pas en compte les enjeux humains, comme la gestion des savoirs implicites, la collaboration et la gestion des désaccords entre experts, ou encore l'adhésion des collaborateurs aux évolutions technologiques. Avant d'être un problème de formalisation, construire une ontologie est tout d'abord un problème de conception et de modélisation qui doit s'effectuer à partir de la description linguistique et informelle des savoir-faire des experts et des connaissances du domaine. Pour répondre à cet enjeu, nous proposons une démarche permettant de construire des ontologies formelles basées sur les savoir-faire. Nous montrons l'apport de l'élicitation de connaissances dans cette démarche, et comment structurer et formaliser les connaissances afin d'uniformiser l'hétérogénéité sémantique pour construire un espace numérique d'expression et de partage des connaissances.

2. Élicitation de connaissances pour le recueil des savoir-faire

La problématique d'acquisition et de formalisation des connaissances est au centre des recherches menées dans le domaine de l'*ingénierie de connaissances* (Studer *et al.*, 1998a). Cette discipline tire ses origines de l'intelligence artificielle et de la psy-

chologie cognitive. Elle propose des concepts, méthodes et techniques permettant d'acquérir, de modéliser, et de formaliser les connaissances d'un domaine dans un objectif d'opérationnalisation, de structuration et de communication. En ingénierie de connaissances, l'*élicitation de connaissances* correspond au processus permettant d'aider un expert métier à exprimer et formaliser ses connaissances dans un objectif de les conserver, les utiliser et les transmettre. C'est un long processus qui doit être adapté selon la nature des connaissances à acquérir, le domaine d'application, et les raisonnements à mettre en œuvre. C'est la phase la plus coûteuse et la plus délicate dans un processus de construction d'un système à bases de connaissances.

En milieu industriel, recueillir les savoir-faire est une étape importante dans la construction de solutions numériques, et nécessite un accès direct aux connaissances et à l'expertise humaine. De nombreuses techniques d'élicitation de connaissances ont été proposées dans la littérature (Gavrilova, Andreeva, 2012; Iqbal, Suaib, 2014; Khan¹ *et al.*, 2014). La figure 1 représente une taxonomie des techniques les plus utilisées (Chen, 2012). La classification proposée est basée sur la nature des techniques selon les types de communication utilisés par l'analyste pour obtenir les connaissances des experts. Ces techniques d'extraction de connaissances sont pour la plupart issues du domaine de la psychologie du travail. Certaines sont dites *directes*, et consistent à interroger l'expert et à le faire s'exprimer oralement, comme les interviews. Tandis que les autres, dites *indirectes*, sont fondées sur l'interprétation des éléments recueillis, capturés indirectement à partir d'autres sources, comme pour l'analyse documentaire.

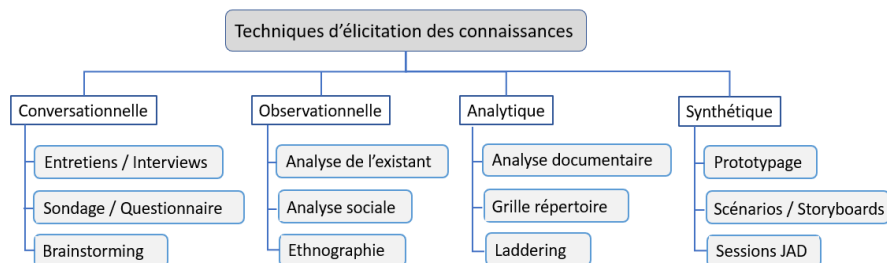


FIGURE 1. Taxonomie des techniques d'élicitation de connaissances

3. Enjeux des techniques d'élicitation de connaissances

L'élicitation de connaissances est une étape critique dans le processus de développement des systèmes à base de connaissances. Il est donc impératif d'appliquer les techniques appropriées pour mener à bien ce processus selon la nature des connaissances et du raisonnement à mettre en œuvre. Il existe évidemment des combinaisons de ces techniques où différentes approches sont utilisées à différentes étapes du

processus d'acquisition des connaissances. Néanmoins, le manque de méthodologies précises pour le choix et l'adaptation de ces techniques, ainsi que le grand écart entre le domaine de l'expert et celui de l'analyste rendent cette tâche difficile et source de plusieurs erreurs.

La conversation est l'une des formes d'interaction sociale les plus répandues. Les techniques conversationnelles sont donc très utiles pour l'acquisition des connaissances. Elles permettent également de découvrir les opinions, les sentiments et les objectifs des différents individus. Cependant, ces techniques demandent beaucoup de travail, comme pour l'organisation des réunions ou la production et l'analyse des transcriptions à partir des enregistrements (Christel, Kang, 1992 ; Sharma, Pandey, 2014). De plus, leur efficacité dépend fortement du comportement et de l'attitude des experts et de l'analyste où le contexte de la conversation joue un rôle très important (Byrd *et al.*, 1992).

Les techniques observationnelles sont très pratiques lorsque l'analyste manque d'expérience dans le domaine, et peuvent être utilisées pour la collecte de connaissances tacites liées à la pratique. Néanmoins, tout comme les techniques conversationnelles, les techniques d'observation sont également difficiles à maîtriser et nécessitent beaucoup de temps et d'organisation (Hudlicka, 1996).

La réutilisation de connaissances déjà disponibles permet de gagner du temps et de réduire les coûts. Les techniques analytiques sont alors utilisées pour la collecte des connaissances existantes qui ne sont pas directement exprimées, et provenant de sources différentes. Cependant, ces techniques sont principalement basées sur des données empiriques et de la documentation. Elles font appel à des connaissances antérieures et peuvent restreindre la vision globale du système. La possibilité de reproduire des erreurs est donc une menace constante pour ce type de techniques, et il ne faut les utiliser que dans des situations particulières et bien maîtrisées.

Il est important de souligner que les techniques présentées sont basées sur l'hypothèse implicite que les experts et l'analyste sont coopératifs et sincères, que les experts sont disposés à partager leurs connaissances, et que l'analyste possède l'expérience et les compétences nécessaires pour mener à bien une session d'élicitation. Cependant, dans la réalité, l'élicitation de connaissances est un processus d'interaction sociale complexe qui doit faire face à plusieurs difficultés :

- **Réticence de l'expert.** Le processus d'acquisition de connaissances met l'expert dans des conditions différentes auxquelles il est habitué. Il peut être donc réticent à ce processus, ou agir différemment, perturbé par la présence de l'analyste.

- **Explicitation des connaissances implicites.** Certaines connaissances ne sont pas directement verbalisables par les experts, et sont difficilement accessibles et communicables.

- **Sélection et accès aux connaissances génériques.** Ces connaissances correspondent à un domaine particulier et sont évidentes pour les experts du même domaine. Elles sont indispensables au raisonnement, peuvent être rendues explicites, mais sont rarement exprimées.

– **Distorsions des connaissances expertes.** L'analyste est souvent étranger au domaine de l'expert. Un décalage peut alors se créer entre les connaissances de l'expert et celles transcrites du fait d'une mal compréhension des propos de l'expert.

– **Hétérogénéité sémantique.** Plusieurs experts du même domaine sont concernés par le processus d'élicitation. Il n'est donc pas exclu de se retrouver dans des situations où les experts ne sont pas d'accord. L'expert peut aussi manquer de précision dans l'expression de ses connaissances ou se tromper, ce qui peut engendrer des incohérences ou des ambiguïtés sémantiques.

À ces difficultés s'ajoute le fait que la plupart des techniques d'élicitation de connaissances sont coûteuses, très complexes, et doivent faire face à une réticence générale, particulièrement dans l'industrie. De plus, il n'existe pas de méthodologies claires pour l'application de ces techniques qui sont souvent effectuées d'une manière ad-hoc, sans un processus prédéfini.

4. Élicitation de connaissances basée sur la didactique professionnelle

L'un des défis majeurs posés en élicitation de connaissances est de tendre vers le recueil le plus exhaustif possible des connaissances métiers et des exigences de la solution cible. Il est donc nécessaire de combiner plusieurs techniques afin de bien comprendre le domaine, structurer et adapter les entretiens, faire face aux éventuelles ambiguïtés et désaccords entre les experts, et assurer leur adhésion aux représentations numériques résultantes. La réussite d'un processus d'élicitation de connaissances dépend fortement de la méthodologie appliquée et des solutions mises en place pour faire face aux éventuelles difficultés.

Comme discuté précédemment, les techniques d'élicitation se reposent essentiellement sur la participation des experts métiers. De ce fait, la volonté de ces derniers à partager leurs connaissances et leur adhésion au projet jouent un rôle principal dans la réussite de l'opération. De plus, la nature des interactions humaines et sociales que les experts du domaine peuvent avoir avec leur interlocuteur peuvent fortement affecter leur comportement et l'expression de leur savoir-faire.

Pour répondre à ces verrous, nous avons développé, en collaboration avec des chercheurs dans le domaine des Sciences Humaines et Sociales, une approche d'élicitation de connaissances inspirée de la *didactique professionnelle*, et basée sur les apports de la psychologie du travail et du développement (Pastré, 2002). Notre approche est une combinaison des techniques conversationnelles, observationnelles et analytiques, et peut être adaptée selon le domaine d'application, la nature des connaissances à collecter, et les besoins du système à développer.

Notre démarche cible les connaissances implicites et explicites. Elle prend en compte l'état psychologique des experts lors des entretiens afin de garantir la clarté, la complétude et la cohérence des propos, et de résoudre tout conflit apparent. Elle permet de répondre à plusieurs problématiques d'élicitation de connaissances, comme

l'explicitation des connaissances implicites, l'adhésion des experts métier, et la résolution d'éventuels conflits liés à l'hétérogénéité sémantiques des connaissances. Les principales phases de notre approche sont présentées dans la figure 2.

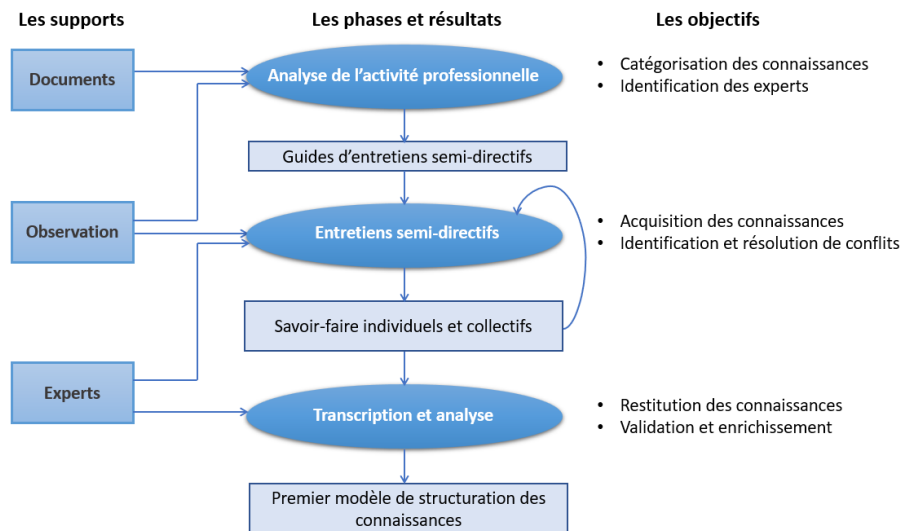


FIGURE 2. Approche d'élicitation de connaissances pour le recueil des savoir-faire

4.1. Analyse de l'activité professionnelle

Inspirée des approches analytiques, cette phase consiste à étudier les documents internes de l'entreprise et, dans certains cas, réaliser un état de l'art sur le domaine afin de collecter les connaissances génériques (réglementation, normes, processus, etc.). Ces connaissances sont indispensables au raisonnement, et ne sont toujours pas évidentes pour l'analyste. Il faut donc prendre conscience de leur existence et leur importance afin de les collecter et de les intégrer dans le système à développer.

Combinée à cette méthode analytique, nous proposons de réaliser une phase d'observation pré-entretien afin d'acquérir une meilleure compréhension du domaine d'application et des processus existants. L'objectif est de faire émerger un référentiel des connaissances et des modalités nécessaires pour accomplir une tâche précise. La finalité de cette phase est de construire, à partir de sources d'information hétérogènes (documents internes, état de l'art, observation), le dictionnaire des données de l'entreprise permettant de maîtriser une partie du vocabulaire métier du domaine indispensable pour préparer la phase des entretiens et faciliter la communication.

À l'issue de cette phase, le dictionnaire de données, le glossaire métier, ainsi que la liste des processus et des experts concernés par la collecte des savoir-faire, sont pré-

sentés aux experts métier pour validation. Pour cela, nous utilisons généralement des plateformes de collaboration de type Wiki ¹ afin de faciliter l'échange avec les experts. Cette phase permet d'assurer une compréhension commune et partagée du vocabulaire et des besoins du système à développer. Elle est indispensable dans un processus d'élicitation pour mieux préparer et définir les guides d'entretien et gagner en temps et en efficacité. Notre expérience dans ce domaine a montré une bonne recevabilité de la démarche et une forte motivation et collaboration par les experts métier.

4.2. Entretien semi-directif

Cette phase consiste à extraire la connaissance par la mise en place d'entretiens adaptés selon les besoins identifiés suite à la phase d'analyse. L'objectif est de définir des guides d'entretien adaptés, d'anticiper les temps nécessaires pour la réalisation de ces entretiens, et de définir en amont le planning selon la disponibilité des experts.

Dans les techniques conversationnelles, la conduite d'entretien est l'une des pratiques de collecte de connaissances les plus utilisées. Elle peut être directive, non-directive, ou semi-directive (Chevalier, Meyer, 2018). Dans notre approche, nous utilisons les entretiens semi-directifs basés sur un dialogue continu. Pour cela, un guide d'entretien est réalisé au préalable avec des questions directives afin de lister les points à aborder. Il se poursuit, selon la réponse de l'expert, par un ensemble de questions libres laissant la liberté à l'expert de s'exprimer. Dans notre approche, les guides d'entretien sont définis en partenariat avec des experts en psychologie de travail et en Sciences Humaines et Sociales afin de bien prendre en compte l'état psychologique des experts lors des entretiens, les aider à mieux exprimer leurs pensées, et faire face aux éventuels soucis de réticence.

Dans notre approche, une première étape d'entretiens individuels est réalisée. Elle permet de recueillir l'information générale concernant l'exercice des savoir-faire de manière individualisée. Dans cette phase d'entretien, les techniques d'observation entrent également en jeu lorsque la communication verbale devient insuffisante pour la collecte des connaissances tacites. Elles permettent également la vérification et la validation des connaissances recueillies.

Des entretiens collectifs pourraient être réalisés si l'équipe le juge pertinent. Ceux-ci seront effectués après analyse des entretiens individuels, et permettent de traiter des thèmes spécifiques, souvent relatifs à des sujets de confusion ou de désaccord entre experts. L'objectif de ce travail collaboratif est de lever l'ambiguïté et d'uniformiser les connaissances collectées. Dans notre approche, ce processus est réalisé par des analystes possédant des compétences interpersonnelles pour aider à établir un consensus entre des groupes hétérogènes d'experts. Ces compétences sociales sont aussi importantes que les techniques utilisées dans le processus d'élicitation.

1. Un wiki est une application web qui permet la création, la modification et l'illustration collaboratives de pages à l'intérieur d'un site web. Nous utilisons dans notre approche Semantic MediaWiki https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki

4.3. *Transcription et analyse*

La transcription des entretiens est la restitution des propos de l'expert en vue d'une analyse (Rioufreyt, 2016). Elle doit être fidèle, compréhensible et complète. Ainsi, pour garantir la clarté et l'uniformité des transcriptions, nous proposons la construction d'un modèle général et structuré (*template*) regroupant l'ensemble des métadonnées à associer pour chacun des entretiens transcrits. La mise en place d'un template commun pour la restitution aide également à la structuration et l'organisation des connaissances collectées.

Après la transcription sous forme textuelle des entretiens (enregistrés en audio), un volume important de propos des experts est collecté. Pour traiter le résultat, nous procédons à une *analyse thématique catégorielle* (Dany, 2016), une technique en sciences humaines et sociales permettant de décrire et de tirer des conclusions sur les caractéristiques d'une communication. Elle met l'accent sur l'identification, l'analyse, et l'interprétation des patterns de sujets ou thèmes au sein des textes examinés, en répondant aux questions basiques de type *Qui, Quoi, Quand, Comment*, etc.

Des analyses complémentaires, comme le calcul d'une Binomiale (théorème de Bernoulli, 1713) (Droesbeke *et al.*, 2002), peuvent être également réalisées selon les besoins du projet. Les avantages de ces méthodes d'analyse comprennent notamment la mise en évidence de points importants manquants pouvant servir de point de développement, d'approfondissement et de discussion.

L'analyse des entretiens permet également la collecte de règles logiques en lien avec le savoir-faire afin de les coder et les intégrer dans le modèle formel. Ces règles sont directement extraites du corpus d'entretiens, et correspondent à des phrases de type conditionnel comme : "*Si... alors*", "*Quand...*", "*Il faut...*".

Les connaissances collectées et le résultat de l'analyse sont restitués et présentés aux experts pour validation avant formalisation et transformation numérique de ces connaissances. Par souci de déontologie, la restitution est anonyme, et aucun entretien n'a été intégralement transmis hormis à l'équipe de recherche.

5. **Modélisation formelle des connaissances**

Une ontologie est une représentation formelle de la connaissance. Elle est définie comme une spécification formelle et explicite d'une conceptualisation partagée d'un domaine d'intérêt (Studer *et al.*, 1998b). Elle fournit une structure commune et partagée des concepts, leurs propriétés et leurs relations, et permet la représentation des connaissances d'un domaine sous une forme interprétable par la machine. Par conséquent, une ontologie est destinée à être utilisée comme un composant logiciel dans des systèmes informatiques répondant à des objectifs opérationnels différents.

Au cours des dernières années, il y a eu un réel intérêt pour l'application des ontologies dans plusieurs domaines afin de résoudre des problèmes de modélisation liés à

la gestion de connaissances, tels que l'intelligence artificielle, la linguistique, le web sémantique, l'intégration de bases de données hétérogènes, etc.

De nombreuses méthodologies de construction d'ontologies ont été proposées dans la littérature. Elles se distinguent selon le type d'ontologie à développer (générique ou du domaine) et des données en entrée (corpus de textes, bases de connaissances, schémas de base de données, etc.). La plupart des méthodologies existantes proposent des recommandations et directives générales sous forme de guides à respecter dans le processus de développement (Gruber, 1995; Gómez-Pérez *et al.*, 1995; Noy *et al.*, 2004). Certaines méthodologies proposent de suivre des phases de construction inspirées des processus de développement en entreprise (Gruninger, Fox, 1996), tandis que d'autres méthodologies plus complètes tentent de combiner plusieurs approches pour construire une démarche unifiée (Uschold, 1996; De Nicola *et al.*, 2005).

Il est généralement admis que, pour être efficace et économiquement faisable, les ontologies doivent être développées de manière collaborative. Une première méthodologie de développement collaboratif d'ontologies a été proposée par Holsapple *et al.* (Holsapple, Joshi, 2002). Depuis, Construire des ontologies de manière collaborative est devenu un paradigme central définissant l'*ingénierie d'ontologie collaborative*, et plusieurs méthodologies ont été proposées dans le domaine (Simperl, Luczak-Rösch, 2014; Hayes *et al.*, 2005). Cette catégorie se base sur le fait qu'une ontologie doit faire l'objet d'un consensus et être acceptée par la communauté du même domaine. Elle fait intervenir plusieurs personnes qui doivent travailler ensemble afin de créer une compréhension commune et partagée de la structure et du sens des concepts de l'ontologie.

6. Enjeux des méthodologies de construction d'ontologies

Malgré les efforts et le nombre croissant de travaux sur la construction d'ontologies (Fernández-López, 1999; Pinto, Martins, 2004; Gómez-Pérez *et al.*, 2004), nous constatons que les méthodologies proposées sont souvent limitées à des recommandations et ne constituent pas une méthodologie complète. Elles sont développées de manière indépendante et se limitent souvent à des études de cas pour le développement d'une seule ontologie liée à un projet particulier. Le principal obstacle à l'application de ces méthodologies est qu'il n'y a pas d'indication claire sur le degré de généralité de ces techniques et méthodes individuelles. Par conséquent, il est difficile de décider quelles méthodologies sont susceptibles de s'appliquer pour un domaine particulier.

Impliquer différents acteurs et experts du domaine dans le processus de construction d'ontologie est l'ambition de l'ingénierie d'ontologie collaborative. Elle vise à construire une ontologie riche et complète en impliquant plusieurs acteurs ou groupes. Néanmoins, elle est principalement basée sur l'hypothèse qu'un compromis est toujours possible en cas de divergence d'opinions et de connaissances, ce qui ne reflète toujours pas la réalité. Les groupes de participant dans le processus sont généralement géographiquement décentralisés et ayant des domaines de compétences variés, rendant ainsi difficile la définition d'un consensus qui doit se faire de manière progressive et

asynchrone. Euzenet (Euzenet, 1995 ; 1996) a identifié un certain nombre de problèmes liés à la construction d'ontologies dans un contexte distribué, comme la gestion des interactions et la communication entre les différentes personnes, le contrôle de l'accès aux données, et la détection et la correction d'erreurs. D'autres limites de la construction d'ontologie collaborative sont également discutés dans (Simperl, Luczak-Rösch, 2014). Les méthodologies existantes sont souvent spécifiques à des projets particuliers, et sont difficiles à généraliser.

De manière générale, les approches existantes de construction d'ontologies restent empirique, sans méthode normalisée. Elles ne prennent pas en compte certains aspects humains, ne proposent pas de solutions pour assister et orienter les experts dans l'expression de leur savoir-faire, et ne font pas face aux réticences et appréhensions potentielles face à cette transformation technologique. Pour résoudre ce problème, nous proposons dans la section suivante une méthodologie collaborative pour le développement d'ontologies centrée sur l'expert et basée sur la technique d'élicitation de connaissances que nous avons proposée dans la section 4.

7. Du savoir-faire à l'ontologie formelle

Une fois que les savoir-faire sont clairement identifiés, il est nécessaire de les exprimer dans un langage compréhensible par la machine. Le passage d'une description informelle à une modélisation formelle des connaissances est un processus complexe et risqué. Afin de répondre à cette problématique, nous proposons une démarche pour structurer, modéliser et formaliser les connaissances. La figure 3 en représente les principales étapes. Après l'élicitation des connaissances pour le recueil du savoir-faire, une structuration et une modélisation semi-formelle sont réalisées, suivi d'une formalisation sémantique de ces savoirs avec intégration des ontologies existantes du domaine. Notre proposition est une hybridation des deux principales approches utilisées en in-

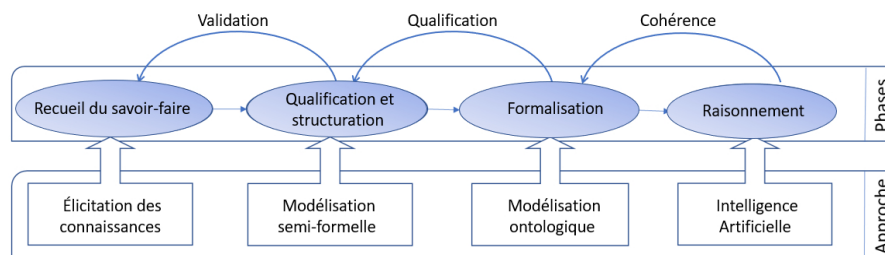


FIGURE 3. Approche de formalisation des savoir-faire

génierie de connaissances pour la construction de modèles. La première correspond à une démarche ascendante basée sur les techniques d'extraction et d'identification des connaissances, tandis que la deuxième est une démarche descendante qui privilège la réutilisation et l'adaptation de modèles existants (Aussenac-Gilles *et al.*, 2012).

7.1. Structuration et modélisation semi-formelle

L'objectif de cette phase est de passer d'une représentation informelle des savoir-faire, exprimés en langage naturel, à une structuration cohérente et une modélisation semi-formelle sous forme de cartes mentales et de diagrammes UML (Unified Modeling Language) (OMG, 2017). Ce standard de modélisation d'un domaine, selon le paradigme objets, propose un ensemble de diagrammes permettant l'expression des différentes facettes d'un savoir-faire : la définition des éléments statiques de la connaissance, la définition des éléments dynamiques associés (les usages) et une représentation des contraintes simples associée aux usages. Cette étape de modélisation est basée sur les résultats de la phase d'analyse et de restitution des connaissances présentée dans la section 4. Elle est réalisée en collaboration avec les experts métiers, qui sont les seuls à pouvoir valider les modèles proposés. La modélisation UML est principalement utilisée pour assurer la bonne compréhension des savoir-faire et les attentes des collaborateurs. Elle réduit considérablement le gap entre la description informelle des savoirs fournie par les experts, et leur formalisation numérique qui doit être compréhensible par la machine. De plus, UML est basé sur une représentation graphique facilement compréhensible par les participants, et largement utilisé en milieu industriel. Dans notre approche, le modèle UML proposé doit être suffisamment expressifs pour être compris par les experts du domaine non-informaticiens et suffisamment contraint pour correspondre à une logique informatique. Il permettra ainsi de mener des échanges entre les différents intervenants pour obtenir une modélisation validée par tous.

Nous utilisons dans notre approche les diagrammes de classes UML permettant de mettre en avant les concepts du domaine (classes), leurs propriétés (attributs) et leurs relations (association). Les diagrammes construits doivent être validés par les experts du domaine et révisés si nécessaires afin de représenter de manière cohérente et complète les connaissances recueillies lors de la phase d'élicitation de connaissances. Le langage UML est aussi maîtrisé par les personnels informatiques des entreprises, cela nous permet de les intégrer dans la démarche de transition numérique à partir de cette étape. Ils auront la charge, si ce n'est d'implémenter les modèles finaux présentés, au moins d'interconnecter la solution au système d'information numérique et de développer des services d'exploitations de ces connaissances.

7.2. Formalisation et modélisation ontologique

La modélisation UML répond aux besoins d'expression, mais pas au besoin d'une compréhension machine. Le langage UML ne permet qu'une représentation semi-formelle de la connaissance et ne prend pas en compte la mise en place des raisonnements associés au savoir-faire. Pour répondre à cette limitation, nous proposons de traduire cette représentation UML dans une équivalence sémantique formelle sous forme d'une ontologie (Mkhinini *et al.*, 2019). Plusieurs travaux scientifiques se sont intéressés à l'étude de la relation entre les modèles UML et les langages d'ontologie. Pour comprendre et se positionner par rapport à ces travaux, nous avons réalisé une revue de

littérature systématique (SLR : Systematic Literature Review) sur le sujet (Mkhinini *et al.*, 2020). Le résultat de cette étude a montré que la majorité des travaux existants proposent généralement une transformation d'un diagramme de classes UML vers une ontologie en se focalisant sur l'aspect syntaxique des éléments UML. De plus, ils ne prennent pas en compte l'intégration des ontologies existantes validées et utilisées par des experts du même domaine d'application. De notre expérience, la transformation des modèles UML en ontologie doit se baser, au-delà de la syntaxe, sur le contexte et le sens de chaque élément. Pour répondre à ce besoin et compléter le savoir décrit par les experts métier, nous sélectionnons et utilisons tout ou partie d'ontologies existantes disponibles dans des bases comme LOV² (Linked Open Vocabularies).

Pour ce faire, nous avons développé une approche basée sur la correspondance sémantique entre les modèles de connaissances permettant la réutilisation des ontologies existantes dans le domaine (Mkhinini *et al.*, 2019). Après validation par les experts métier des modèles UML générés à l'étape précédente, nous utilisons des règles de transformation inspirées des travaux existants pour passer des modèles UML représentant les savoir-faire à une représentation ontologique formelle. Nous proposons par la suite d'intégrer des ontologies existantes en identifiant les similarités sémantiques entre ces ontologies et les diagrammes de classes UML. Une étape d'alignement des concepts identifiés dans des sources externes (connaissance du domaine) et ceux définis en local (savoir-faire des experts de l'entreprise) est alors nécessaire pour mettre en correspondance les connaissances du domaine et les savoir-faire des experts métiers. Ainsi, l'ontologie finale définit le vocabulaire logique qui permet d'exprimer des faits et des connaissances du domaine et les connaissances métiers sur lesquelles nous pouvons raisonner. Cette approche permet l'identification des ensembles sémantiques correspondant à la description de chaque élément du savoir-faire en relation avec les connaissances du domaine. L'ontologie résultante servira de base pour construire des cas d'usages métiers et des outils numériques adaptés sur lesquels des règles de raisonnement pourront être ajoutées.

8. Expérimentation

Notre approche de formalisation d'ontologies basée sur l'élicitation de connaissances a été expérimentée dans le cadre de projets de collaborations industrielles.

Le premier projet, en collaboration avec une multinationale, dans le domaine du packaging de luxe³, avait pour principal objectif la capitalisation et la digitalisation des savoir-faire de l'entreprise pour la construction de nouvelles solutions numériques. Ce projet englobait des processus métiers réalisés sur 4 pays (France, Etats-Unis, Brésil, Chine), dont 4 langues et usages différents. Dans ce projet, nous avons appliqué notre démarche d'élicitation de connaissances pour interviewer 55 collaborateurs représentant différents services de l'entreprise sur plusieurs sites dans les 4 pays. Cette

2. <https://lov.linkeddata.es/dataset/lov/>

3. Le contrat de confidentialité ne permet pas de dévoiler le nom de l'entreprise.

expérience nous a confronté aux problèmes d'expression du savoir-faire, souvent basée sur une terminologie diversifiée et complexe. Après transcription et analyse des connaissances collectées, nous avons identifié le vocabulaire du domaine, élaboré une vision globale de l'organisation des différents services de l'entreprise, et proposé une modélisation UML pour la structuration et l'unification de la terminologie et des usages. Après validation par les experts, cette modélisation a été par la suite utilisée pour construire une ontologie formelle basée sur le savoir-faire, et à laquelle nous avons intégré des concepts d'ontologies du domaine. L'ontologie résultat a été utilisée pour démontrer la faisabilité industrielle de notre approche et fournir un espace numérique collaboratif d'expression des savoir-faire pour l'entreprise. À partir de cet environnement numérique, nous avons développé une application de raisonnement artificiel permettant d'extraire l'ensemble des projets, expériences, risques, coûts, ressources matérielles, matières premières et ressources humaines proches des éléments décrits dans un cahier des charges client. Il était ainsi possible d'anticiper au plus juste le risque de prise en charge de réalisation de la demande client.

La deuxième expérimentation concerne un projet de collaboration dans le domaine de la viticulture⁴. L'objectif de ce projet était de fournir une plateforme de traçabilité complète du cycle de vie de la vigne en se basant sur les connaissances d'experts et des données de remontées de capteurs. L'application de notre approche d'élicitation de connaissances dans ce projet a permis la collecte des savoir-faire métier (mise en culture de la vigne, lutte contre les maladies, méthodes de vinification, etc.) des viticulteurs et des vinificateurs afin de comprendre le cycle de vie de la vigne et du vin, et les différents raisonnements et pratiques mis en œuvre dans le cadre de leur profession. Ainsi, pour comprendre le vocabulaire du domaine et construire les guides d'entretiens semi-directifs, un travail de veille bibliographique a été réalisé pour en savoir plus sur la filière vitivinicole et les risques qui y sont associés. Dans ce domaine, les viticulteurs se réfèrent beaucoup à l'expérience de leurs pairs, et l'observation de la nature et de ses phénomènes. Les aspects scientifiques et législatifs sont aussi très présents et doivent être pris en compte dans la prise de décisions. Du fait de la complexité du sujet exploré, il a été nécessaire de développer plusieurs grilles d'entretien permettant de prendre en considération les différents paramètres venant impacter le cycle de la vigne et la vinification. Ainsi, six grilles d'entretiens semi-directifs explorant différentes thématiques ont été développées, à savoir sur le cycle de la vigne, les maladies, les ravageurs, les risques climatiques, les risques physiologiques et la vinification. Un panel de 20 viticulteurs et de 5 professionnels du vin a été sélectionné pour passer les entretiens. Après restitution et analyse des savoirs collectés, nous avons proposé une structuration et une modélisation semi-formelle de ces savoirs sous forme de cartes mentales et de diagrammes UML, suivi d'une formalisation ontologique permettant de construire des solutions numériques capables de prédire l'état des vignes et de fournir aux viticulteurs des recommandations pour leur production (Belkaroui *et al.*, 2018). L'avenir dira si ce système permettra aux experts du domaine, comme ils le souhaitent à l'origine du projet, d'anticiper les risques liés aux ravageurs et

4. <https://www.europe-bfc.eu/beneficiaire/fui-23-wine-cloud/>

maladies, de réduire les traitements chimiques dangereux pour l'environnement, et de maîtriser la modification des caractéristiques organoleptiques de leurs vins suite à l'influence du changement climatique.

Nous appliquons notre approche dans plusieurs projets de collaboration, y compris le projet ANR DALHAI⁵, qui exige la collecte et la formalisation de connaissances expertes pour construire des solutions d'intelligence artificielle innovantes.

9. Conclusion

Cet article propose une approche méthodologique pour la construction d'une ontologie formelle basée sur le savoir-faire des experts et les connaissances du domaine. Nous avons défini une méthodologie d'acquisition des savoir-faire inspirée de la didactique professionnelle, et basée sur les apports de la psychologie du travail et du développement. Cette méthodologie est une combinaison des techniques analytiques, conversationnelles et observationnelles en élicitation de connaissances. Elle permet de collecter, qualifier et analyser les savoir-faire explicites et tacites, d'intégrer les connaissances génériques du domaine, et de faire face aux problématiques liées aux facteurs humains et à l'hétérogénéité sémantique des connaissances. Son principal objectif est de mettre l'humain qui détient la connaissance au centre des réflexions dans le développement des solutions numériques, et d'accompagner au mieux les analystes dans le processus de préparation, de collecte, d'analyse et de restitution de ces connaissances.

Les connaissances collectées sont par la suite organisées et structurées dans des modèles semi-formels, qui seront validés par les experts métier avant d'être formalisés dans une ontologie du domaine, consensuelle et partagée. Cette démarche permet de se doter d'une base de connaissance digitalisée et d'un ensemble d'outils numériques proches des experts et construits à base de leurs savoir-faire, leurs expériences et leurs attentes. Nous avons expérimenté notre approche dans le cadre de deux projets de collaboration industrielle. Le premier pour la digitalisation des savoir-faire d'une multinationale dans le domaine du packaging et le deuxième dans le domaine de la viticulture pour la prédiction du cycle de vie de la vigne. Ces deux expériences ont montré l'apport et la faisabilité industrielle de notre approche et remporté l'adhésion des experts du domaine comme celui du personnel IT.

Remerciements

Nous remercions le projet DALHAI pour le soutien apporté à ce travail scientifique.

Bibliographie

Aussenac-Gilles N., Charlet J., Reynaud C. (2012). Les enjeux de l'ingénierie des connaissances. *Information, Interaction, Intelligence-Le point sur le i* (3), p. 244–266.

5. <https://anr.fr/Projet-ANR-20-CE24-0001>

- Belkaroui R., Bertaux A., Labbani O., Hugol-Gential C., Nicolle C. (2018). Towards events ontology based on data sensors network for viticulture domain. In *Proceedings of the 8th international conference on the internet of things*, p. 1–7.
- Byrd T. A., Cossick K. L., Zmud R. W. (1992). A synthesis of research on requirements analysis and knowledge acquisition techniques. *MIS quarterly*, p. 117–138.
- Chen Y. (2012). *Requirement elicitation techniques*. https://www.umsl.edu/~sauterv/analysis/6840papers_f12/Chen/.
- Chevalier F., Meyer V. (2018). Chapitre 6. les entretiens. *Les méthodes de recherche du DBA*.
- Christel M. G., Kang K. C. (1992). *Issues in requirements elicitation*. Rapport technique. Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst.
- Dany L. (2016). Analyse qualitative du contenu des représentations sociales. *Les représentations sociales*, p. 85–102.
- De Nicola A., Missikoff M., Navigli R. (2005). A proposal for a unified process for ontology building: Upon. In *Database and expert systems applications: 16th international conference, dexa 2005, copenhagen, denmark, august 22-26, 2005. proceedings 16*, p. 655–664.
- Droesbeke J.-J., Fine J., Saporta G. (2002). *Méthodes bayésiennes en statistique*. Editions Technip.
- Euzenat J. (1995). *Building consensual knowledge bases: context and architecture*. IOS press.
- Euzenat J. (1996). Corporate memory through cooperative creation of knowledge bases and hyper-documents. In *10th workshop on knowledge acquisition (kaw)*, p. 36–1.
- Fernández-López M. (1999). Overview of methodologies for building ontologies.
- Gavrilova T., Andreeva T. (2012). Knowledge elicitation techniques in a knowledge management context. *Journal of Knowledge Management*.
- Gómez-Pérez A., Fernández-López M., Corcho O. (2004). Methodologies and methods for building ontologies. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*, p. 107–197.
- Gómez-Pérez A., Juristo N., Pazos J. (1995). Evaluation and assessment of knowledge sharing technology. *Towards very large knowledge bases*, p. 289–296.
- Gruber T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, vol. 43, n° 5-6, p. 907–928.
- Gruninger M., Fox M. S. (1996). The logic of enterprise modelling. In *Modelling and methodologies for enterprise integration: Proceedings of the ifip tc5 working conference on models and methodologies for enterprise integration, queensland, australia, november 1995*, p. 140–157.
- Hayes P., Eskridge T. C., Saavedra R., Reichherzer T., Mehrotra M., Bobrovnikoff D. (2005). Collaborative knowledge capture in ontologies. In *Proceedings of the 3rd international conference on knowledge capture*, p. 99–106.
- Holsapple C. W., Joshi K. D. (2002). A collaborative approach to ontology design. *Communications of the ACM*, vol. 45, n° 2, p. 42–47.

- Hudlicka E. (1996). Requirements elicitation with indirect knowledge elicitation techniques: comparison of three methods. In *Proceedings of the second international conference on requirements engineering*, p. 4–11.
- Iqbal T., Suaib M. (2014). Requirement elicitation technique:-a review paper. *Int. J. Comput. Math. Sci.*, vol. 3, n° 9, p. 1–6.
- Jones D., Bench-Capon T., Visser P. (1998). Methodologies for ontology development.
- Kanza S., Stolz A., Hepp M., Simperl E. (2018). What does an ontology engineering community look like? a systematic analysis of the schema. org community. In *European semantic web conference*, p. 335–350.
- Khan¹ S., Dulloo A. B., Verma M. (2014). Systematic review of requirement elicitation techniques. *India*.
- Maedche A., Schnurr H.-P., Staab S., Studer R. (2000). Representation language-neutral modeling of ontologies. In *Proceedings of the german workshop "modellierung-2000". koblenz, germany*, p. 129–142.
- Missikoff M., Schiappelli F. (2005). A method for ontology modeling in the business domain. In *Emoi-interop*.
- Mkhinini M. M., Labbani O., Nicolle C. (2019). Enterprise knowledge modeling, uml vs ontology: Formal evaluation. In *2019 IEEE 15th international conference on intelligent computer communication and processing (iccp)*, p. 127–134.
- Mkhinini M. M., Labbani-Narsis O., Nicolle C. (2020). Combining uml and ontology: An exploratory survey. *Computer Science Review*, vol. 35, p. 100–223.
- Noy N. F., McGuinness D. L. et al. (2004). *Ontology development 101: a guide to creating your first ontology. 2001*. See <http://protege.stanford.edu/publications>.
- OMG. (2017). *Omg© unified modeling language (omg uml©)*. In Version 2.5.1, © 2017 object management group, inc. Consulté sur <https://www.omg.org/spec/UML/>
- Pastré P. (2002). *L'analyse du travail en didactique professionnelle*. *Revue française de pédagogie*, p. 9–17.
- Pinto H. S., Martins J. P. (2004). *Ontologies: How can they be built?* *Knowledge and information systems*, vol. 6, p. 441–464.
- Rioufreyt T. (2016). *La transcription d'entretiens en sciences sociales*.
- Sharma S., Pandey S. K. (2014). *Requirements elicitation: Issues and challenges*. In 2014 international conference on computing for sustainable global development (indiacom), p. 151–155.
- Simperl E., Luczak-Rösch M. (2014). *Collaborative ontology engineering: a survey*. *The Knowledge Engineering Review*, vol. 29, n° 1, p. 101–131.
- Studer R., Benjamins V. R., Fensel D. (1998a). *Knowledge engineering: Principles and methods*. *Data & knowledge engineering*, vol. 25, n° 1-2, p. 161–197.
- Studer R., Benjamins V. R., Fensel D. (1998b). *Knowledge engineering: Principles and methods*. *Data & knowledge engineering*, vol. 25, n° 1-2, p. 161–197.
- Uschold M. (1996). *Building ontologies: Towards a uni ed methodology*. In *Proceedings of 16th annual conference of the british computer society specialists group on expert systems*.

Exploitation de scénarios prospectifs pour l'analyse des exigences

Application à la conception des avions futurs

Brieuc Danet¹

1. Département Traitement de l'Information et Systèmes
ONERA - The French Aerospace Lab - Centre de Toulouse
2, avenue Edouard Belin - BP 74025 - 31055 TOULOUSE CEDEX
brieuc.danet@onera.fr

RESUME. Cet article présente une démarche d'analyse systémique mise en place dans un objectif d'élicitation des exigences. Il s'agit d'exploiter des scénarios prospectifs par une méthode d'analyse stratégique tournée vers les acteurs, en adoptant le formalisme proposé par le langage iStar. La démarche est appliquée aux scénarios de transport aérien futur décrits dans le rapport d'une étude menée en 2021 par des organismes de recherche aéronautique européens. Les exigences obtenues doivent permettre d'alimenter le processus de conception des avions futurs en définissant des critères adaptés aux technologies à venir, notamment en termes d'impacts écologiques, sociétaux et éthiques.

ABSTRACT. This article presents a systemic analysis approach implemented with the aim of requirements elicitation. It is a question of exploiting prospective scenarios through a strategic "actor-oriented" analysis method by adopting the formalism proposed in the iStar language. The approach is applied to future air transport scenarios described in the report of a study conducted by European aeronautical research organizations in 2021. The requirements obtained must make it possible to feed the future aircraft design process by defining criteria adapted to future technologies, especially concerning ecological, societal and ethical impacts.

Mots-clés : Analyse systémique ; élicitation des exigences ; conception avion ; scénarios prospectifs ; critères d'évaluation ; impact écologique ; analyse stratégique ; analyse lexicologique.

KEYWORDS: Systemic analysis; requirements elicitation; aircraft design; prospective scenarios; assessment criteria; ecological impact ; strategic analysis; lexicological analysis.

1. Introduction

La conception d'un nouvel avion s'appuie sur un besoin exprimé ou anticipé d'un client. Ce besoin se traduit généralement sous forme d'une mission, exprimée en nombre de passagers à transporter et d'une distance à parcourir, en un temps donné. Par ailleurs, de nombreuses contraintes doivent être prises en compte dès la conception. Elles peuvent toucher aux opérations aériennes et émaner de la réglementation du trafic aérien (certification) ou de contextes aéroportuaires. Le travail de design répond enfin à des objectifs de performances techniques et économiques, qui doivent permettre in fine à l'entreprise de produire et vendre des appareils de façon rentable.

Des outils dits de MDAO (Multi-Disciplinary Analysis and Optimization) permettent aujourd'hui de prendre en compte tous ces aspects de la conception avion, en gérant la complexité de la modélisation autant que la problématique de réponse aux besoins (Defoort *et al.*, 2022). Ils se basent sur des algorithmes très performants pour pouvoir analyser des systèmes de grande dimension aux variables couplées, et mettent en place des stratégies d'optimisation permettant de gérer des objectifs et contraintes multiples (Bartoli *et al.*, 2019).

Ces outils, et les méthodologies sur lesquelles ils reposent, sont en perpétuelle évolution. Ils sont conçus pour pouvoir être utilisés par les industriels pour optimiser l'avion à construire, mais également par les chercheurs pour évaluer le potentiel de concepts novateurs, aux technologies encore peu matures. Il est alors possible de comparer des configurations géométriques, des technologies de propulsion, ou des sources d'énergie différentes, et d'en déduire lesquelles paraissent les plus prometteuses.

Pourtant, ces évaluations de concepts exploratoires posent problème. D'une part, on utilise des critères actuels pour évaluer des concepts et technologies qui ne seront disponibles qu'à un horizon parfois lointain, selon leur niveau de TRL (Technical Readiness Level). D'autre part, depuis la prise de conscience du dérèglement climatique et la crise sanitaire, la vision d'une industrie qui devrait avant tout viser la performance est devenue obsolète. Bien d'autres types de critères doivent et devront être pris en compte dans les années à venir (Henderson *et al.*, 2012) : écologiques, sociétaux et éthiques en premier lieu.

Il paraît donc indispensable d'envisager l'avenir pour pouvoir déterminer quel type d'appareil sera alors le plus adapté. Cela étant dit, il est évidemment illusoire de penser qu'il serait possible de prévoir, ou de prédire, les tenants et aboutissants du transport aérien futur. L'alternative consiste à envisager différents types de scénarios constituant un panorama dont la diversité devrait permettre d'englober l'ensemble des avènements possibles, quel que soit leur "réalisme". C'est l'objet des études de prospective (Godet, 2007), qui visent à anticiper des situations vraisemblables à venir, pour s'y préparer, et non de les prédire en tentant d'évaluer leur probabilité d'occurrence.

Par ailleurs, compte tenu de la complexité du transport aérien et de son évolution, il est nécessaire de prendre en compte la globalité du problème par une

analyse systémique. En effet, définir les besoins, exigences, contraintes et objectifs à venir revient à considérer les relations d'interdépendance de l'ensemble des acteurs du secteur, qu'ils soient privés ou publics, nationaux ou internationaux.

A titre d'exemple, l'EREA¹, l'organe européen regroupant l'ensemble des établissements de recherche aéronautique, a débuté en 2010 une étude prospective dans le but de construire une vision commune du transport aérien futur. La dernière mise à jour du rapport correspondant (EREA, 2021), présente quatre scénarios qui sont supposés couvrir l'ensemble des situations à envisager pour appréhender les problématiques à venir.

Ce rapport, et les scénarios qu'il détaille, constituent une source très riche d'informations collectées, mises en forme et validées par les principaux acteurs scientifiques du domaine aéronautique. C'est donc le point de départ idéal pour une analyse qui doit mener à la définition des exigences à prendre en compte pour les concepts d'avion futurs.

Cet article présente la démarche qui a été mise en œuvre pour exploiter ces scénarios. Une fois le cas d'étude présenté en détail, une analyse stratégique des acteurs exposera dans un premier temps les difficultés liées à l'exploitation d'un tel rapport. Ensuite, un travail de collecte et de formalisation des informations issues des scénarios grâce au langage iStar permettra de montrer l'intérêt et les limites d'une telle démarche pour l'analyse des exigences. Enfin, un cas d'application lié à la modularité des aéronefs futurs présentera l'intérêt d'envisager l'étude sous un angle transverse.

2. Cas d'étude

Dans son rapport, l'EREA avait pour objectifs d'aider à renforcer la coopération au niveau européen, soutenir les choix des décideurs politiques et nourrir le dialogue avec le grand public. La démarche a donc consisté à situer le transport aérien futur dans un contexte international, tenant compte d'une évolution des situations politiques, économiques, sociologiques et écologiques des différents pays concernés. Le rapport concrétise la longue réflexion collaborative d'un groupe d'experts à travers quatre scénarios « type ».

Le premier scénario, dénommé « Mad Max », anticipe un monde fragmenté, démondialisé, caractérisé par le nationalisme extrême et le populisme. Il y règne une grande instabilité économique, en raison d'économies protectionnistes et de niveaux élevés d'inégalité. L'aviation y est un produit coûteux, luxueux et hautement désirable, mais peu fiable en raison de la crise climatique, des faibles niveaux d'innovation et des différentes sources d'instabilité qui pèsent sur la disponibilité des ressources.

Le deuxième scénario, « Tech For You », décrit un monde caractérisé par la multipolarité et la libre concurrence, où le marché est guidé par le choix et les désirs des consommateurs. Une production à coût élevé et à faible économie d'échelle est

¹ Association of European Research Establishments in Aeronautics

compensée par un haut niveau d'automatisation. Les investissements en R&D sont élevés et se font au niveau national et à court terme. Le transport aérien y fait partie d'un système de transport flexible, intermodal et global.

Dans le troisième scénario, « Stripping Down », le monde est globalisé, organisé autour de quelques blocs démocratiques de taille continentale, au sein desquels règne une grande stabilité politique. La croissance économique est planifiée, lente mais stable, et les marchés sont organisés autour d'une régulation incitative avec priorité à la durabilité, au détriment du choix du consommateur. La mobilité est limitée et hautement contrôlée. Le transport aérien n'est toléré qu'en tant que faisant partie de solutions génériques intermodales durables adaptées pour réduire l'impact environnemental.

Le quatrième et dernier scénario, « Optimising Together », considère un monde où règnent la coopération et la collaboration mondiales, à travers des cadres juridiques et institutionnels unifiés et harmonisés. Cette stabilité globale et de hauts niveaux de standardisation et de confiance mutuelle permettent à la fois une croissance élevée et une durabilité forte. La mobilité se développe et est pleinement durable. Différentes solutions aéronautiques sont disponibles pour tous les segments de voyage, de la mobilité aérienne urbaine aux vols en formation en passant par les vols suborbitaux.

3. Analyse stratégique des scénarios

L'exploitation d'un tel rapport, synthèse d'une étude au long cours, n'est pas simple. D'une part, il s'agit d'une étude qualitative, dans laquelle chaque scénario est supposé bien distinct des autres, et décrit avec un réel effort de clarté et de justesse, mais sans données chiffrées. En effet, il s'agit de projections dans l'avenir, élaborées conjointement par plusieurs experts du domaine, mais non générées par un outil probabiliste dans lequel on aurait modélisé l'évolution des différents paramètres en jeu, même si certaines méthodes d'analyse structurelle (Godet et Durance, 2011) sont souvent utilisées au cours de la réflexion.

D'autre part, le document a été rédigé pour être lu par des publics divers (chercheurs, industriels, décideurs, grand public, etc.), pas nécessairement experts eux-mêmes. La présentation du document et le langage utilisé peuvent donc paraître manquer de précision et de systématisme.

Pour que ces scénarios puissent servir de base à la définition d'un système de transport futur, il est nécessaire d'en effectuer une analyse exhaustive, en démêlant la complexité du système étudié, pour comprendre les origines et interdépendances des exigences imposées aux différents acteurs, et aboutir in fine à un panel de critères d'évaluation représentatifs pour les concepts à venir.

Le travail d'analyse du rapport de l'EREA consiste tout d'abord à lister l'ensemble des acteurs mis en jeu, avant de détailler, pour chacun d'entre eux et pour chaque scénario, les objectifs, les tâches à accomplir, les ressources nécessaires et/ou disponibles, et les qualités à mettre en avant.

Cette démarche d'analyse stratégique rejoint un principe proposé par Crozier et Friedberg (1977), sociologues des organisations et pionniers de la Prospective en France, selon lesquels l'analyse de l'évolution d'un système complexe, tel que le système de transport aérien, débute par l'analyse pratique des relations de pouvoir entre acteurs du système et des règles, implicites ou explicites, qui gouvernent leurs interactions.

De nombreux acteurs du transport aérien sont présents dans les quatre scénarios, parfois nominativement, parfois à travers le rôle particulier qu'ils y jouent ou la fonction qu'ils y occupent. La multiplicité de ces acteurs n'est pas surprenante, et la prise en compte de leur place dans le système de transport aérien futur est a priori nécessaire pour l'étude. En effet, même si les exigences qui entrent en jeu dans la conception d'un avion s'appliquent en premier lieu aux industriels en charge de leur fabrication, ainsi qu'à leurs sous-traitants (motoristes par exemple), en aval elles proviennent de très nombreux acteurs : besoins des usagers, demandes des compagnies aériennes, contraintes des organismes de réglementation, ressources gouvernementales, etc. Un travail préalable à toute analyse systémique consiste donc à les lister, pour les trier et comprendre leurs interdépendances éventuelles.

Même nombreux, ces acteurs du transport aérien futurs restent globalement similaires aux acteurs actuels. Pour clarifier l'exercice et tenter une première analyse, on choisit de les classer selon leur appartenance à l'un des groupes suivants :

- les acteurs sociaux (population civile, passagers ou non, associations et collectifs),
- les acteurs de la recherche,
- les acteurs internationaux (organismes décisionnels ou de réglementation),
- les acteurs nationaux (gouvernements),
- les acteurs industriels (constructeurs, motoristes, vendeurs d'énergie),
- les acteurs opérationnels (compagnies aériennes).

Ainsi, on peut tenter de caractériser chacun des scénarios en notant le niveau d'occurrence des acteurs de chacun de ces groupes dans le texte qui le décrit.

On observe sur la Figure 1 que les acteurs nationaux et les acteurs sociaux sont les plus cités dans l'ensemble des scénarios, ce qui n'a rien de surprenant, étant donné que l'industrie aéronautique et le transport sont deux activités économiques majeures pour de nombreux gouvernements occidentaux, que ce soit en termes de finances ou d'emploi. Les citoyens, eux, sont évidemment concernés à plusieurs titres, directement en tant que passagers ou travailleurs du secteur, mais aussi indirectement, en particulier à travers l'impact du transport aérien sur leurs vies (bruit, pollution, dérèglement climatique, etc.) mais aussi dans leur éventuelle implication dans des groupes politiques ou militants.

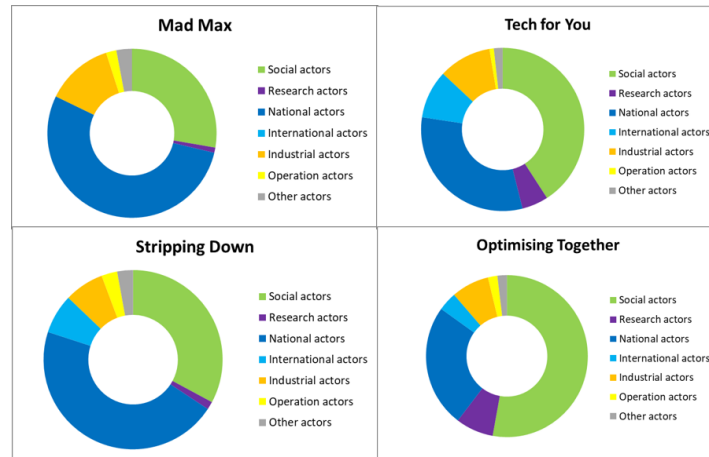


Figure 1. Diagramme d'occurrence des groupes d'acteurs dans les scénarios

Par ailleurs, on remarque que la part de ces acteurs sociétaux dans le nombre total d'occurrences varie considérablement d'un scénario à l'autre : de 28% pour « Mad Max » à 54% pour « Optimising Together ». L'étude EREA étant notablement tournée vers la question de la responsabilité sociétale du transport aérien, on a choisi de détailler davantage l'étude sur ces acteurs sociétaux pour tenter d'en comprendre un peu plus.

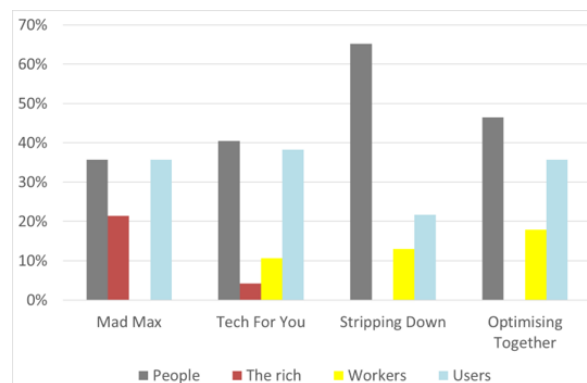


Figure 2. Pourcentage d'occurrence des acteurs sociétaux

La Figure 2 laisse apparaître que la composition même du groupe des acteurs sociétaux change fortement d'un scénario à l'autre. On distingue les scénarios davantage tournés vers les utilisateurs (passagers), en particulier les plus riches (« Mad Max » et « Tech For You »), et ceux où la population civile et les

travailleurs du secteur ont une influence plus importante (« Stripping Down » et « Optimising Together »). Ces tendances sont révélatrices de sociétés aux fonctionnements politiques radicalement différents, qui doivent fortement influencer sur les objectifs comme sur les moyens mis en œuvre pour concevoir les aéronefs futurs.

Le niveau d'implication des acteurs ou groupes d'acteurs permet donc de différencier, voire de caractériser les scénarios. Envisager un processus d'analyse par les acteurs pour extraire de ces scénarios l'information utile in fine à l'étude des exigences paraît donc pertinent, à condition d'introduire un formalisme adapté à une telle analyse systémique.

4. Formalisation du système d'acteurs

Plusieurs langages d'ingénierie des exigences ont déjà été utilisés pour représenter un tel système d'acteurs (pour KAOS cf. van Lamsweerde, 2009 ; pour URN/GRL cf. Amyot *et al.* 2009). Nous avons choisi iStar (Yu et Mylopoulos, 1997 ; Dalpiaz *et al.*, 2016) pour exploiter la description de chacun des scénarios selon la démarche stratégique et systémique envisagée précédemment.

En effet, iStar paraît très indiqué pour expliciter un système complexe tel que celui du transport aérien futur, où la notion d'acteur est centrale (Carvallo et Franch, 2009). Par ailleurs, iStar a déjà été utilisé avec succès dans le monde aéronautique pour représenter les objectifs liés à la conception d'avions (Chan *et al.*, 2022 ; 2022).

D'une part, le langage permet d'explorer le système d'acteurs à travers leurs buts et leurs relations d'interdépendance, à l'aide des diagrammes SD (« Strategic Dependency »). D'autre part, les éléments intentionnels, qui régissent le comportement des acteurs, permettent de mieux déterminer les exigences qu'ils s'imposent, individuellement ou mutuellement. Effet, ces exigences peuvent, par exemple, être liées à la disponibilité d'une ressource, à la sous-traitance de tâches à accomplir, à des fonctionnalités liées à des besoins client, ou à des critères de performance. Expliciter ainsi la nature des exigences à l'aide de diagrammes SR (« Strategic Rationale ») permet de mieux les prendre en compte dans le processus de conception.

La première étape de notre travail consiste donc à modéliser le système d'acteurs à partir des informations préalablement collectées. Le langage iStar permet de distinguer l'agent (organisation identifiée et nommée par exemple) du rôle qu'il joue dans la société. Par ailleurs, un rôle peut être défini comme la spécialisation d'un autre rôle (par exemple, un pilote est un industriel de l'aéronautique spécialisé) ; on utilisera alors le lien « is a ». Si un agent (ou un rôle) peut contribuer individuellement à un rôle (un citoyen qui ferait partie d'une association de défense de l'environnement par exemple), on utilisera le lien « participates in ».

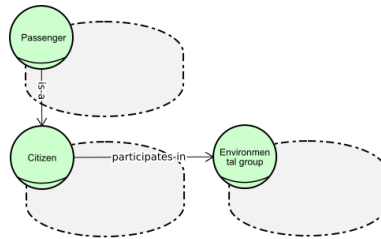


Figure 3. Exemples de liens de dépendance entre rôles

Une fois la liste des acteurs bien établie, à travers une mutualisation de leurs dénominations, il est possible d'extraire du texte de présentation des scénarios tous les contenus qui vont introduire une notion d'exigence. Ce travail doit une fois de plus s'effectuer de façon exhaustive, pour permettre in fine de tracer ces exigences.

On utilise pour cela le formalisme induit par le langage iStar, qui distingue plusieurs types d'*éléments intentionnels* pour un acteur :

- les **but**s à atteindre, dont l'accomplissement est établi à travers un critère précis, exprimé par un groupe nominal. Par exemple : pour un passager, "speed, comfortable, convenient trip", pour un gouvernement, "own energy needs ensured".
- les **qualités**, qui vont intervenir comme des critères de performance, exprimés par des noms (ou des adjectifs). Par exemple : "wealth", "sovereignty"
- les **tâches**, qui représentent les actions à exécuter, notamment pour atteindre un but. Elles sont exprimées par un groupe verbal. Par exemple : "exploit own fossil fuel resources", "control airspace"
- les **ressources**, qu'elles soient matérielles ou virtuelles (information), requises pour une tâche. Elles sont exprimées par un groupe nominal. Par exemple : "fossil fuels", "accessible information".

Cette étape de formalisation est facilitée par le développement d'une routine qui permet de générer automatiquement le fichier iStar à partir des données stockées pour chaque acteur (par exemple dans un fichier Excel).

La Figure 4 présente l'exemple d'un acteur "bloc" (qui représente un bloc de pays ayant signé des accords diplomatiques de coopération, comme la Communauté Européenne par exemple), et de l'ensemble de ses éléments intentionnels, buts, tâches, qualités et ressources.

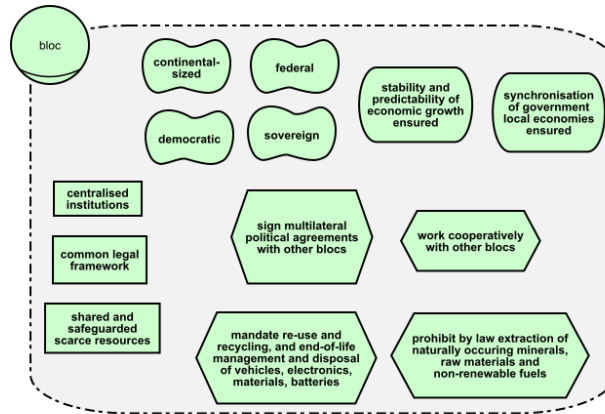


Figure 4. Détail de l'acteur « bloc » : éléments intentionnels

Une fois ces éléments mis en place pour chacun des acteurs, vient l'étape de définition des relations de dépendance qui les lient. Ces relations peuvent être internes à l'acteur, mais peuvent également relier plusieurs acteurs. En effet, aucun acteur n'est indépendant dans le système, et c'est justement dans l'analyse de ces interdépendances que vont s'expliciter les exigences recherchées.

En ce qui concerne les liens internes, on en distingue quatre types :

– Le **raffinement** s'entend entre un but (ou une tâche) et un(e) ou plusieurs but(s) (ou tâche(s)) et comprend un opérateur logique ET ou OU (inclusif)

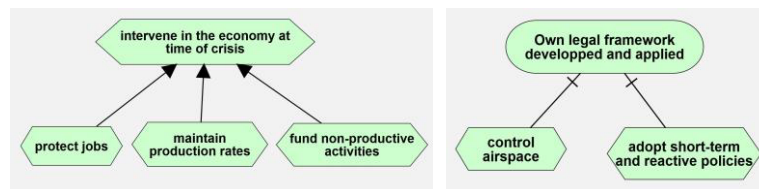


Figure 5. Interdépendance interne : raffinement

- **Requis par** désigne le besoin d'une ressource pour accomplir une tâche
- La **qualification** exprime un niveau de satisfaction attendu pour l'exécution d'une tâche, l'accomplissement d'un but, ou l'approvisionnement d'une ressource
- Une **contribution** participe positivement (Make ++, Help +) ou négativement (Break --, Hurt -) à satisfaire une qualité

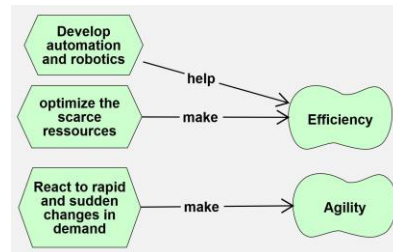


Figure 6. Interdépendance interne : contribution

Les liens entre acteurs se caractérisent à travers leurs éléments intentionnels. Par exemple, un acteur peut avoir besoin d'une ressource d'un autre acteur pour exécuter une de ses propres tâches, ou compter sur l'exécution des tâches d'un autre pour atteindre un de ses buts. Pour matérialiser ce lien le langage iStar définit la notion de *dependum*, qui doit appartenir à l'une des catégories des éléments intentionnels, et qui va cristalliser la notion d'exigence entre les acteurs.

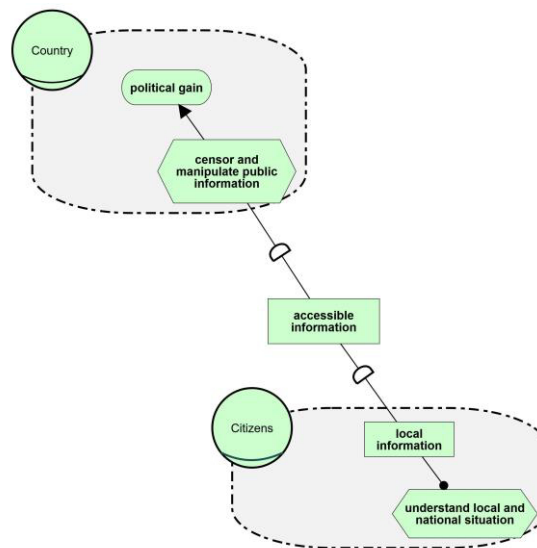


Figure 7. Interdépendance externe entre acteurs

La Figure 7 illustre une relation de dépendance entre un pays et ses citoyens dans le contexte du scénario « Mad Max » : les informations nécessaires aux citoyens pour comprendre leur situation locale et globale dépendent d'un éventuel niveau de censure (voire de manipulation) du gouvernement de leur pays. Ici, c'est l'information disponible qui est le *dependum*, en tant que ressource.

Grâce au formalisme du langage iStar, chacun des scénarios peut ainsi être exploité de façon exhaustive pour aboutir à une représentation du système d'acteurs, permettant de faire le lien entre la dynamique interne de chacun, et l'ensemble des relations entre acteurs.

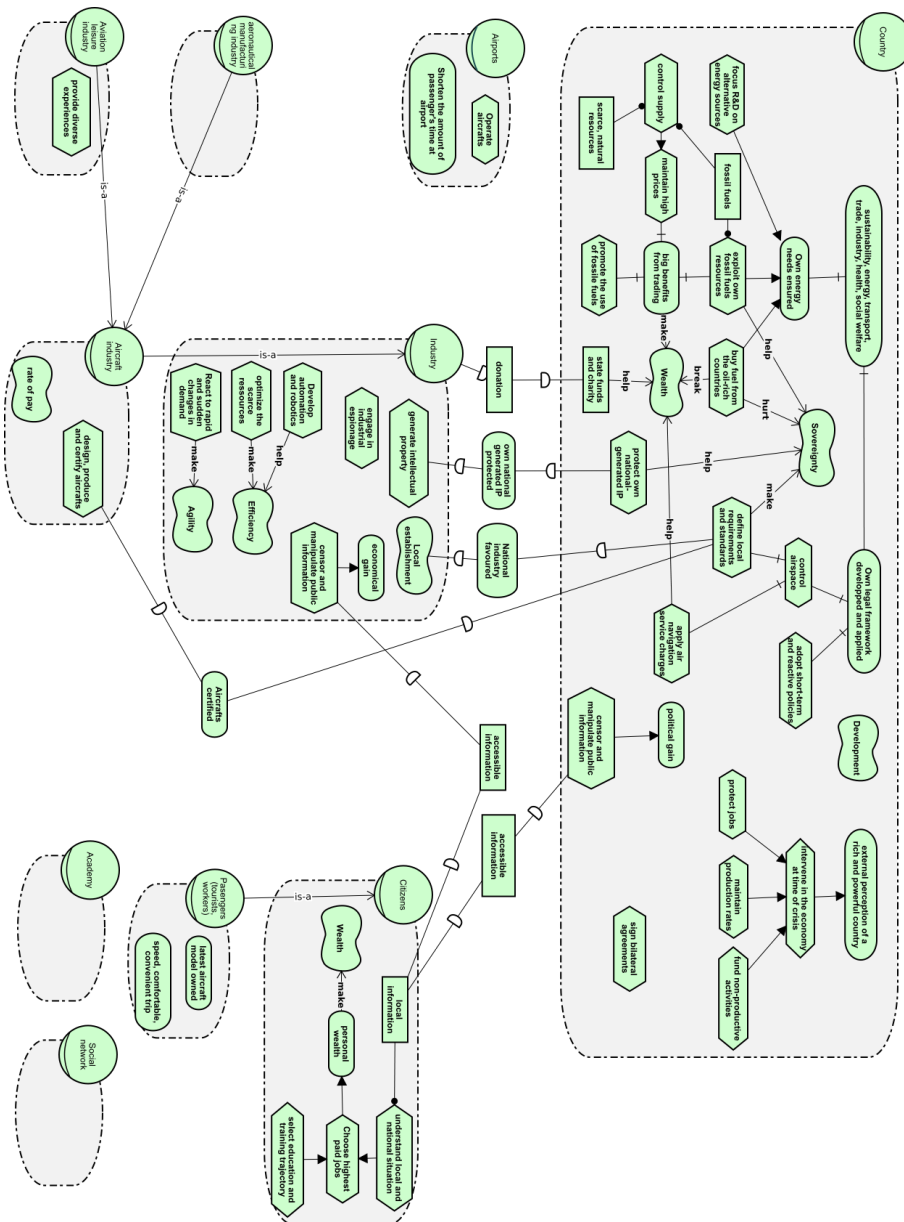


Figure 8. Vue globale et détaillée de l'exploitation du scénario « Mad Max »

La Figure 8 illustre la représentation à laquelle l'analyse aboutit pour le système de transport aérien futur décrit dans le cadre du scénario « Mad Max ».

Le diagramme produit met en valeur le rôle majeur de l'Etat dans ce scénario. Son appui financier est vital pour les industriels nationaux confrontés à un marché très fermé où seuls les citoyens les plus riches peuvent encore s'offrir l'accès au transport aérien. La maîtrise de l'information est d'ailleurs un enjeu de taille vis-à-vis de la population, pour permettre aux industriels et à l'Etat d'agir sans s'inquiéter de l'opinion publique. La stabilité du système global dépend alors majoritairement de la capacité de l'Etat à conserver sa souveraineté financière, l'accès aux ressources critiques, le contrôle de son territoire (y compris l'espace aérien) et son image dans le monde.

Le diagramme a bien permis, dans ce cas, de formaliser les buts des acteurs majeurs, et de préciser les contraintes imposées par leur environnement, pour aboutir aux exigences qui doivent régir leurs décisions.

Toutefois, cette représentation fait apparaître un problème de complétude dans la description des scénarios. En effet, de grandes différences de développement peuvent être notées entre les différents acteurs supposés jouer un rôle dans ce scénario. Pour certains, il n'a même pas été possible de trouver d'éléments permettant de cibler leurs objectifs, d'explicitier leurs éléments intentionnels, ou leurs relations avec les autres acteurs (au-delà des éventuels liens hiérarchiques). Ce manque d'informations invite à questionner les auteurs, pour déterminer si l'influence de l'acteur est effectivement mineure dans ce scénario, ou s'il s'agit en fait de mieux expliciter son rôle pour qu'il puisse intégrer le système et donner lieu à l'élicitation de nouvelles exigences.

La représentation formelle des jeux d'acteurs pourrait alors devenir une forme d'outil de validation de la complétude des scénarios (Alrajeh *et al.*, 2013), à inscrire dans une démarche itérative pour converger vers une vision commune, exhaustive et cohérente du scénario envisagé. Le formalisme proposé par l'outil pourrait même être complètement intégré à la démarche prospective, à l'image de la méthode MACTOR (Peerboccus, 2022) surtout utilisée en sciences sociales pour expliciter les jeux d'acteurs à travers leur niveau d'influence dans le système, mais sans volonté d'explicitier les relations de dépendance, ni la nature des exigences qu'elles génèrent.

5. Perspective

A l'avenir, nous pourrions proposer une autre forme d'analyse, qui s'opère de manière plus transverse aux scénarios en focalisant la réflexion sur une problématique spécifique, susceptible d'impacter plusieurs acteurs du système. Cela peut permettre à n'importe quel utilisateur de mieux se situer vis-à-vis de cette problématique, et d'identifier les liens à tisser avec les autres acteurs du domaine pour répondre à de nouvelles exigences.

Ce cas de figure peut être illustré en se penchant sur la notion de modularité des avions futurs. Ce terme désigne l'idée de faire varier la configuration de l'aéronef pour adapter son niveau de performance en fonction des exigences et du besoin. Il recouvre un grand nombre d'innovations autour de concepts technologiques qui ont en commun une organisation structurelle adaptable. Il peut s'agir, par exemple, d'une cabine amovible pour réduire le temps de rotation des avions à l'aéroport, ou d'un appareil dont l'ensemble des éléments (voilure, fuselage, système propulsif, etc.) sont interchangeable, pour produire une large gamme d'appareils à disposition des besoins du client tout en réduisant les coûts de conception, de sous-traitance et de fabrication.

Cette modularité pourrait ainsi être une innovation majeure, impactant un grand nombre d'acteurs du domaine, constructeurs et équipementiers, compagnies aériennes, aéroports, et bien sûr, in fine, passagers.

Nous nous sommes donc appuyés sur les scénarios EREA pour mieux saisir la complexité du problème, et anticiper les exigences à venir. Pour cela, on a recensé, dans l'ensemble des scénarios, les occurrences du terme de modularité (et des termes du même champ lexical). On a ensuite procédé comme précédemment, en identifiant quels acteurs étaient impactés ou engagés, et en formalisant l'action à travers un ou plusieurs éléments intentionnels, ainsi que les liens associés si nécessaire.

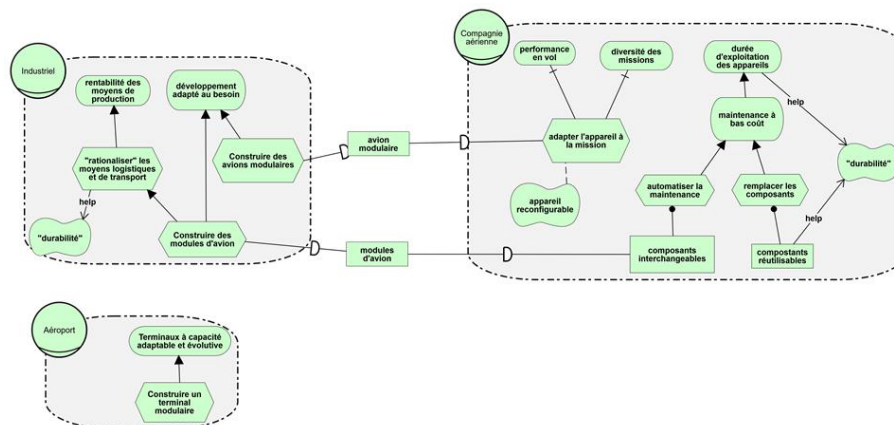


Figure 9. Analyse transverse de la modularité des avions futurs

Il en résulte la représentation schématique de la Figure 9, où peuvent être retrouvées les diverses notions de modularité : des besoins en missions, des appareils, des moyens de production et de transport, ou des terminaux d'aéroport. Les exigences qui apparaissent chez les acteurs considérés pourront ensuite être directement reprises par les utilisateurs intéressés par la problématique, pour être éventuellement qualifiées, voire quantifiées.

6. Conclusions

La démarche proposée pour l'exploitation des scénarios EREA permet d'aboutir à une première vision globale de la cartographie des exigences pour le système de transport aérien futur.

Une analyse systémique par les acteurs est d'abord menée, héritée de principes de sociologie des organisations. Elle consiste en une analyse statistique lexicale du texte descriptif des scénarios, qui conduit à dresser un panorama exhaustif des parties prenantes du système et permet d'estimer le niveau d'implication de chacune.

Le langage iStar est alors adopté pour formaliser cet inventaire des acteurs. D'une part, il s'agit de préciser leur nature (agent ou rôle) et leur logique de fonctionnement propre, en particulier les objectifs qui guident leurs actes, qu'ils soient l'expression directe de buts ou la volonté implicite de satisfaire des critères de qualité. Le formalisme d'iStar permet ensuite d'explicitier la nature des relations entre acteurs, qui peuvent être basées sur la dépendance à une ressource détenue par un autre, sur l'accomplissement de tâches pour lesquelles une compétence spécifique est nécessaire, ou même sur des objectifs à atteindre dont les niveaux sont définis par ailleurs.

Ce travail de formalisation aboutit à une cartographie des liens entre les objectifs des acteurs, les tâches à accomplir pour les atteindre, les ressources nécessaires et les critères de performance à surveiller.

Cette nouvelle forme de ressource, directement issue des informations contenues dans les scénarios élaborés et rédigés par l'EREA, peut ensuite être exploitée de différentes façons. Il peut s'agir, pour un acteur du système, de se projeter dans l'avenir en extrayant les exigences futures associées à son activité, pour mieux anticiper les décisions à prendre et envisager les actions à mener : un industriel pourrait revoir la chaîne d'assemblage de son futur appareil pour prendre en compte l'évolution du coût des ressources et des compétences disponibles, un organisme de réglementation pourrait préparer des mesures incitatives pour encourager les compagnies à satisfaire les demandes d'associations écologistes de plus en plus influentes, etc.

Ce type de cartographie multi-acteurs et multi-domaines est également essentielle pour pouvoir définir les critères d'évaluation des concepts futurs et amorcer les projets de recherche associés. Grâce aux récents travaux sur les algorithmes et les stratégies d'optimisation, il sera en effet envisageable d'analyser le système de transport aérien dans sa globalité, quels que soient la dimension du vecteur des données d'entrée ou le nombre d'objectifs en sortie.

Dans un contexte d'inquiétude pour l'environnement et le climat, et d'instabilité économique et politique, de nombreux scénarios prospectifs sont élaborés qui doivent permettre d'envisager des situations très diverses et se préparer au mieux. L'exploitation systémique de ces scénarios par tout utilisateur concerné pourrait être un préalable à une meilleure sensibilisation, mais surtout un outil d'aide à la décision robuste et transparent.

Une telle démarche de formalisation permettrait de faciliter l'émergence de visions communes d'avenirs, en encourageant un processus itératif permettant de s'assurer de la complétude et de la cohérence globale des informations fournies. Elle pourrait alors être avantageusement couplée à certaines méthodes de génération automatisée de scénarios en cours de développement (Blanchard *et al.*, 2021).

Bibliographie

- Alrajeh D., Russo A., Lockerbie J., Maiden N., Mavin A., Novak M. (2013). *Computational Alignment of Goals and Scenarios for Complex Systems*. Proceedings - International Conference on Software Engineering.
- Amyot D., Horkoff J., Gross D., Mussbacher G. (2009) *A Lightweight GRL Profile for i* Modeling*. ER Workshops 2009: 254-264
- Bartoli N., Lefebvre T., Dubreuil S., Olivanti R., Priem R. (2019). *Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design* - Aerospace Science and technology
- Blanchard C., Saurel C., Tessier C. (2021). *Futurs possibles d'un système d'acteurs : formalisation et génération automatique de scénarios*. In PFIA 2022 - Rencontres des jeunes chercheurs en intelligence artificielle, pages 115–122
- Carvalho, J., Franch, X. (2009). *On the Use of i* for Architecting Hybrid Systems: A Method and an Evaluation Report*. 38-53.
- Chan A., Fernandes Pires A., Polacsek T. (2022). *Trying to Elicit and Assign Goals to the Right Actors*. ER 2022: 413-422
- Chan A., Fernandes Pires A., Polacsek T., Roussel S. (2022) *The Aircraft and Its Manufacturing System: From Early Requirements to Global Design*. CAiSE: 164-179
- Crozier M., Friedberg E. (1977) *L'acteur et le système* ISBN 978-2020046770
- Dalpiaz F., Franch X., Horkoff J. (2016). *iStar 2.0 Language Guide*. arXiv:1605.07767
- Defoort S. *et al.* (2022). *Leveraging on OpenMDAO to enhance MDO capability @ONERA*, OpenMDAO Workshop, GRC, Cleveland, OH – October 24 th, 2022
- EREA (2021) *EREA Vision Study – The Future of Aviation in 2050*, <https://erea.org/>
- Godet M. (2007). *Manuel de prospective stratégique. L'art et la méthode. Vol. 2*. Dunod
- Godet M., Durance P. (2011). *La Prospective stratégique. Pour les entreprises et les territoires* UNESCO / Dunod, p. 62-68
- Henderson R., Martins J., Perez R. (2012). *Aircraft Conceptual Design for Optimal Environmental Performance*. The Aeronautical Journal. 116. 1-22.
- Peerbooccus R., Godet M., Bourse F. (2022) *L'analyse du jeu des acteurs en prospective - La méthode MACTOR*. Futuribles International
- van Lamsweerde A. (2009). *Requirements Engineering - From System Goals to UML Models to Software Specifications*. Wiley 2009, ISBN 978-0-470-01270-3
- Yu E., Mylopoulos J. (1997). *Enterprise modelling for business redesign: The i* framework*. SIGGROUP Bull., 18(1):59–63

L'avion et sa chaîne d'assemblage : des premières exigences à une conception globale

Anouck Chan¹, Anthony Fernandes Pires¹, Thomas Polacsek¹,
Stéphanie Roussel¹

1. ONERA

Toulouse France

prenom.nom@onera.fr

REFERENCE DE L'ARTICLE INTERNATIONAL. Cet article est un résumé de l'article :
Anouck Chan, Anthony Fernandes Pires, Thomas Polacsek, Stéphanie Roussel:
The Aircraft and Its Manufacturing System: From Early Requirements to Global Design. In
Advanced Information Systems Engineering: 34th International Conference, CAiSE 2022:
164-179.

MOTS-CLES : Exigences, Optimisation, Co-conception, Aéronautique, Industrie 4.0

KEYWORDS: Requirements, Optimization, Co-design, Aeronautics, Industry 4.0

Dans le cadre de la conception d'un nouveau modèle d'avion, le système industriel permettant de le construire doit être conçu ad hoc. En effet, de par la spécificité des techniques d'assemblage, chaque famille d'avion dispose d'un système industriel dédié. Cependant, la satisfaction d'objectifs de performance d'un système industriel, comme la cadence de production, dépend fortement des choix de conception de l'objet à fabriquer, ici l'avion. Afin d'optimiser une conception conjointe, nous proposons une méthode d'aide à la décision que nous avons utilisée dans un cas d'étude industriel pour concevoir le design du fuselage d'un avion et de sa ligne d'assemblage. La méthode a été développée au cours de six mois de sessions de travail régulières avec des experts métier. Dans un premier temps, nous avons construit un modèle de buts de type *GORE* (*Goals-Oriented Requirements Engineering*) pour la conception du fuselage d'un avion et de sa ligne d'assemblage, voir Figure 1. Ce modèle nous a permis d'identifier les dépendances entre le design d'un avion et des choix de conception de sa ligne d'assemblage, chaque choix contribuant positivement ou négativement aux objectifs, i.e. buts, de l'usine. Par

exemple, joindre les différentes parties de l'avion par perçage manuel contribuera positivement à un but de minimisation des investissements, mais négativement au but de minimisation du temps d'assemblage, le perçage manuel étant plus long que celui à l'aide d'un robot. À l'inverse, utiliser un robot de perçage contribuera négativement à minimiser les investissements, mais positivement à minimiser les temps d'assemblage. Chacun des choix implique un grand nombre d'alternatives dont les combinaisons ne peuvent pas être toutes explorées et évaluées à la main. En d'autres termes, ces alternatives ne peuvent donc pas être représentées dans un modèle de buts.

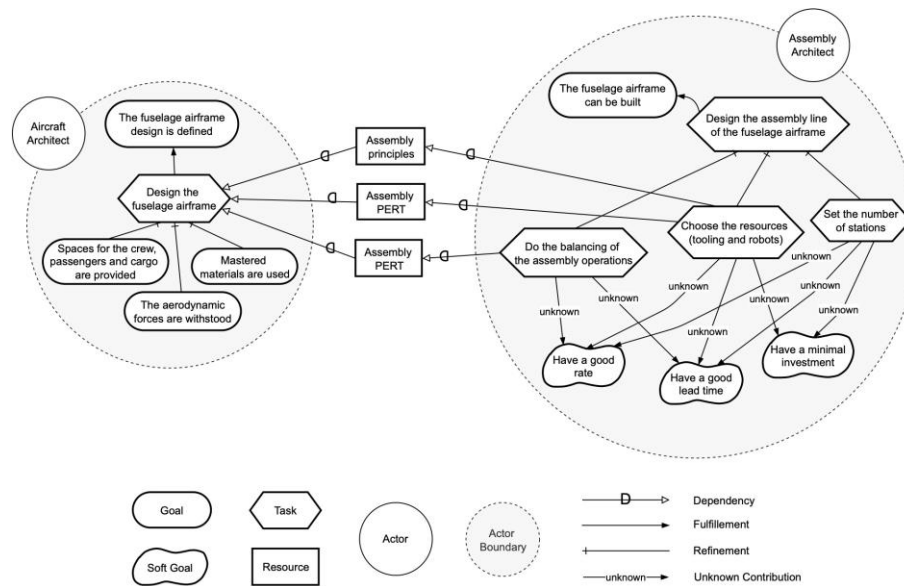


Figure 1 Modèle de but i^*

Dans un second temps, nous avons construit un modèle conceptuel s'appuyant sur le formalisme SysML pour décrire les principaux éléments composant une section d'avion ainsi que la ligne d'assemblage associée. Ce modèle nous permet de structurer l'information qui sera utilisée par l'outil de recherche opérationnelle.

Dans un troisième temps, nous avons conçu un outil d'aide à la décision pour concevoir une chaîne d'assemblage optimal. Cet outil consiste en un programme d'optimisation. Il prend en entrée la description d'un design d'avion ainsi que certaines caractéristiques de la ligne d'assemblage et renvoie des lignes d'assemblages optimales suivant les buts qui ont été définis. En d'autres termes, l'outil renvoie des choix de conception pour la ligne d'assemblage permettant de maximiser la satisfaction des objectifs précédemment élicités dans le modèle de but. Notre outil a été utilisé sur deux designs d'avions fournis par les experts industriels afin d'obtenir un ensemble des combinaisons optimales de design d'avion et design de ligne d'assemblage.

Éliciter, raffiner et attribuer des buts aux bons acteurs à partir d'objectifs de haut niveau

Anouck Chan¹, Anthony Fernandes Pires¹, Thomas Polacsek¹

1. ONERA

Toulouse France

prenom.nom@onera.fr

REFERENCE DE L'ARTICLE INTERNATIONAL. Cet article est un résumé de l'article :

Anouck Chan, Anthony Fernandes Pires, Thomas Polacsek: Trying to Elicit and Assign Goals to the Right Actors. In *Conceptual Modeling: 41st International Conference, ER 2022*: 413-422.

MOTS-CLES : Modélisation orientée buts, Exigences

KEYWORDS: Goal modeling, Requirements

De plus en plus d'organisations tendent à des objectifs très abstraits, de haut niveau, qui décrivent les caractéristiques que tout projet de l'organisation doit satisfaire. En raison de la nature générique, voire abstraite, de ces objectifs, il est parfois difficile de les décomposer en objectifs plus concrets et de décider qui, dans l'organisation, est responsable de quoi. Dans cet article, nous nous intéressons à une approche consistant à raffiner les buts de haut niveau (HLG pour High-Level Goal) d'une organisation en buts réalisables et attribuables aux acteurs de cette même organisation. Un HLG est un objectif de haut niveau, très abstrait, pour lequel il n'est pas toujours évident de savoir comment le satisfaire. Des exemples de HLGs sont : « être toujours dans les temps » ou « être à la pointe de l'innovation ». Une manière de procéder est de les raffiner en sous-buts plus concrets que des acteurs de l'organisation sachent satisfaire.

Depuis de nombreuses années, les approches de modélisation orientées buts proposent des cadres permettant d'obtenir et de définir les objectifs des différentes parties prenantes d'une organisation. Ainsi, plusieurs approches de la littérature permettent d'éliciter des ensembles de sous-buts satisfiables et de les assigner aux bons acteurs Letier *et al.* (2002), Bryl *et al.* (2006), Elrakaiby *et al.* (2018) et Mougouei *et al.* (2018). Cependant, la forte abstraction des HLGs, à laquelle vient s'ajouter le fait que les compétences des acteurs d'une organisation reposent en

grande partie sur des savoir-faire difficiles à modéliser formellement, ne permet pas d'appliquer directement la plupart de ces travaux. C'est pourquoi nous proposons une méthode pour obtenir à partir d'un HLG un ensemble de sous-buts concrets tels que la satisfaction de ces sous-buts induise la satisfaction du HLG et chacun des sous-buts soit satisfiable par un acteur identifié. Il s'agit d'une adaptation des travaux de Bryl *et al.* (2006).

Notre méthode est principalement composée de deux actions : la *délégation* d'un but et le *raffinement* en sous-buts. La délégation consiste à donner un but à un acteur. Ce dernier est désormais responsable de la satisfaction de ce but et doit par conséquent être en mesure, avec ses compétences, de le satisfaire. Le raffinement, dans notre approche, consiste à partitionner un but en ce que l'acteur sait satisfaire et en ce qu'il ne sait pas satisfaire. Ainsi, dans notre algorithme, un acteur peut raffiner un but g en deux sous-buts : le sous-but g_a comportant tous les éléments du but g que l'acteur est capable de satisfaire et le sous-but g_b , tous les autres éléments de g tels que la satisfaction de g_a et g_b implique la satisfaction de g . Le but g_b sera ensuite délégué à un autre acteur.

Nous avons appliqué notre méthode sur des HLGs d'une entreprise aéronautique. Pour cela, nous avons organisé six sessions de travail avec des experts métier. Cette application nous a amenés à enrichir notre méthode, notamment avec l'ajout d'un acteur nommé *Unknown*. Ce dernier récupère les buts qui ne peuvent pas être associés à un acteur existant. L'ajout d'*Unknown* permet, entre autres, d'identifier des acteurs (ou rôles) absents de l'entreprise et pourrait aider celle-ci à revoir son organisation pour satisfaire aux mieux ses objectifs.

Bibliographie

- (Bryl *et al.*, 2006) Bryl V., Giorgini P., Mylopoulos J. Designing cooperative IS: exploring and evaluating alternatives, Actes OTM Conferences, volume 4275 of Lecture Notes in Computer Science. Springer, p. 533-550. (2006)
- (Elrakaiby *et al.*, 2018) Elrakaiby, Y., Ferrari, A., Mylopoulos, J.: Care: A refinement calculus for requirements engineering based on argumentation semantics. Dans : IEEE 26th International Requirements Engineering Conference (RE). pp. 364–369 (2018)
- (Letier *et al.*, 2002) Letier, E., van Lamsweerde, A.: Agent-based tactics for goal-oriented requirements elaboration. Dans: Proceedings of the 24th International Conference on Software Engineering (ICSE), pp. 83–93 (2002)
- (Mougouei *et al.*, 2018) Mougouei, D., Perera, H., Hussain, W., Shams, R.A., Whittle, J.: Operationalizing human values in software: a research roadmap. Dans : Proceedings of the ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ESEC/SIGSOFT FSE. pp. 780–784. ACM (2018)

Outils pour calculer l'empreinte TIC des organisations : adaptation d'une étude européenne

**Guillaume Bourgeois ¹, Kassandra Bigot ¹, Vincent Courboulay ¹,
et Benjamin Duthil ²**

¹ *Laboratoire Informatique, Image, Interaction (L3I), La Rochelle Université, 17000 La Rochelle, France*

² *EIGSI, 17041 La Rochelle, France*

Cet article est un résumé de l'article publié dans EMCIS 2022.

https://www.researchgate.net/publication/367058442_Tools_for_calculating_the ICT_footprint_of_organisations_adaptation_of_a_European_study

eBook ISBN :
978-3-031-30694-5
Sortie: 31 Mai 2023

MOTS-CLES : Systèmes d'information durables ; calculateur d'empreinte carbone ; actions individuelles ; informatique verte ; éco-responsable ; impact écologique ; développement durable ; consommation d'énergie.

KEYWORDS : Sustainable Information Systems; carbon footprint calculator; individual actions; Green Computing; eco-responsible; ecological impact; sustainable development; energy consumption.

Le changement climatique est un problème mondial qui affecte notre planète et notre société. Plusieurs initiatives internationales visent à réduire les émissions de gaz à effet de serre et à atteindre la neutralité climatique d'ici 2050. Les organisations doivent mesurer et réduire leur empreinte carbone, en particulier dans le secteur informatique. Elles doivent aussi utiliser les technologies numériques pour répondre aux urgences climatiques et sociales.

Pour répondre à cette problématique, l'ISIT (Institut du Numérique Responsable) en France, en Suisse et en Belgique a mené la seule étude gratuite visant à mesurer l'empreinte quantitative et qualitative des systèmes d'information, WeNR 2021. Cette recherche est un héritage direct des travaux de WeGreenIT publiés par le WWF en 2018.

Grâce à un processus simplifié, la méthodologie WeNR permet à chacun d'évaluer facilement l'empreinte carbone de son équipement informatique. Une équipe d'ingénieurs WeNR est chargée de collecter, compiler et vérifier la fiabilité des données afin de proposer un référentiel commun à la communauté. Ainsi, WeNR représente un grand pas en avant dans le domaine.

Cependant, la collecte de données peut être compliquée pour ceux qui souhaitent participer au processus. Par exemple, la consommation énergétique totale d'une salle de serveurs peut être difficile à obtenir. Cependant, cela ne devrait pas être un obstacle à la participation à WeNR.

Pour résoudre ce problème, WeNR Light a été créé. Cette version allégée remplace le long questionnaire par un questionnaire beaucoup plus court, d'environ dix questions à remplir directement sur le site, permettant d'obtenir rapidement une estimation du nombre de kilogrammes de CO₂ équivalent produits par an par un employé pour une organisation. Toutefois, il y a une marge d'incertitude d'environ 10 à 15%.



Figure 1 Page d'accueil du WeNR Light

La principale contribution scientifique de cette initiative est de fournir aux organisations un outil simple et pratique pour mesurer et réduire leur empreinte carbone dans le domaine informatique. En particulier, la méthodologie WeNR et son application WeNR Light permettent de calculer l'empreinte carbone de manière rapide et efficace, en fournissant un référentiel commun à la communauté. Cela permet aux organisations de prendre des décisions éclairées pour réduire leur impact environnemental et de participer à l'effort collectif pour atteindre la neutralité climatique d'ici 2050.

De plus, en utilisant des pratiques recommandées par la DINUM, les organisations peuvent maximiser leur impact environnemental et social positif.

En résumé, l'ISIT et sa méthodologie WeNR sont une contribution importante à la lutte contre le changement climatique et la promotion de la durabilité dans le secteur informatique.

Un simulateur pour une adoption de Cloud plus éco-consciente

Evgeny Vasyuk¹, Arthur Vervaeet^{1,2}, Mar Callau-Zori¹, Raja Chiky¹

1. Outscale, France {prénom.nom}@outscale.com

2. LISITE, ISEP, Ecole d'Ingénieur du Numérique, Paris, France

RÉSUMÉ. Cet article présente un prototype de simulateur pour l'estimation de la consommation énergétique d'un système d'information déployé dans le Cloud. L'objectif est de proposer des méthodes et des outils d'estimation comparative de différentes configurations d'un tel système. Le simulateur prend en compte la consommation électrique au repos des serveurs ainsi que la consommation électrique additionnelle générée par l'activité du système et des applications.

ABSTRACT. This paper presents a prototype simulator for power consumption estimation of an Information system deployed in the Cloud. The objective is to propose methods and tools for the comparative estimation of different configurations of such a system. The simulator takes into account the idle power consumption of the servers and the additional power consumption generated by the system activity and its jobs.

MOTS-CLÉS : simulateur; cloud; consommation d'énergie

KEYWORDS: simulator; cloud; energy consumption

1. Introduction

Au cours des dernières années, l'informatique verte (*Green computing*) est devenue un sujet d'intérêt émergent, aussi bien dans l'académie que dans l'industrie. Plusieurs études discutent de l'impact du secteur tel que (Murugesan, Gangadharan, 2012) qui propose une analyse des bonnes pratiques et les stratégies pour répondre aux enjeux inhérents. Une étude récente (Verdecchia, Lago, Ebert, Vries, 2021) souligne entre autres les aspects suivants pour l'accélération vers une infrastructure numérique économe en énergie : 1/ **La migration vers le Cloud**, avec une réduction potentielle d'entre 80 et 90% (Verdecchia, Lago, D, 2021 ; Desjardins *et al.*, 2014); 2/ **Associer le logiciel à son contexte d'utilisation** pour une bonne prise de décision en matière d'éco-conception. Dans la pratique, exploiter pleinement le potentiel du Cloud pour

la conception d'infrastructures numériques économe en énergie reste compliqué. Un des principaux freins étant la difficulté d'évaluer l'impact de choix de conception et d'architecture des applications qui opèrent sur un Cloud. Par exemple, une question pourrait concerner le choix entre : a/ démarrer 1 machine virtuelle (VM) avec 4 coeurs et 2 MB de RAM ou b/ 2 VMs avec 2 coeurs et 1MB chacune ?

Les simulateurs ont déjà montré leur potentiel pour l'évaluation des comportements et des choix dans de larges infrastructures distribuées (Kumar, Kumar, 2019). CloudSim (Calheiros *et al.*, 2011 ; Lefebvre *et al.*, 2014) fournit un environnement de simulation d'une infrastructure Cloud utilisée par les fournisseurs pour évaluer l'impact des algorithmes de load-balancing ou de scheduling. Pour l'efficacité énergétique, GreenCloud (Kliazovich *et al.*, 2012) ajoute des éléments de consommation énergétique de hardware (serveurs, switches, etc.) à CloudSim. Bien que la simulation soit principalement adressée aux fournisseurs pour l'aide à la prise de décision à grande échelle, le simulateur CloudAnalyst (Wickremasinghe *et al.*, 2010) étend les fonctionnalités de CloudSim pour tester des applicatifs déployés.

Aucun de ces travaux précités n'adresse tous les facteurs qui impactent la consommation énergétique des applications qui tournent sur un Cloud mais aussi le choix de la configuration des infrastructures déployées. Ce papier présente un simulateur de consommation énergétique qui permet aux clients de se projeter sur l'impact énergétique de différentes configurations. alors, tel simulateur contribué à une adoption de Cloud écoresponsable en évaluant de façon simplifiée plusieurs configurations.

La première version du simulateur, accessible en ligne, a été proposé lors d'un hackathon organisé conjointement avec l'ISEP et les entreprise OUTSCALE et Sagemcom (github.com/outscale/hackathon202210). Dans la suite de ce papier, nous introduisons le simulateur (Sec. 2) ainsi que des conclusions et perspectives (Sec. 3).

2. Mesurer l'impact des choix infrastructures déployées dans le Cloud

La mutualisation des ressources est un des piliers du Cloud, elle permet aux opérateurs de maîtriser leurs coûts tout en fournissant aux clients des ressources virtuelles de façon transparente. Par exemple, une machine virtuelle (VM) est un ensemble de processus qui tournent sur un serveur physique hôte; ainsi, plusieurs VMs peuvent être allouées à un même serveur physique et gérées de façon indépendante. Les ressources virtuelles ne consomment pas d'énergie en elles-mêmes, mais elles contribuent à augmenter la consommation de l'équipement physique qui les héberge.

Un des grands défis pour mesurer l'empreinte énergétique d'une infrastructure, consiste à isoler la consommation énergétique des ressources virtuelles pour fournir une vision d'ensemble au client. Cependant, pour contribuer à une conception des infrastructures Cloud plus économes en termes d'énergie, des mesures comparatives des différentes configurations sont suffisantes.

Alors, notre simulateur vise à fournir une comparaison entre différents choix de déploiements. À cette fin, il reçoit : la configuration du déploiement et la configuration

des applications (*jobs*) qui vont générer la charge de travail. D’abord, le simulateur déploie une infrastructure de type infrastructure avec l’outillage pour estimer la consommation énergétique; ensuite, le *job* s’exécute en temps réel en collectant des métriques qui serviront, enfin, pour fournir un résultat de consommation d’énergie (Fig. 1) qui sert à comparer deux déploiements différents. À titre d’exemple, le simulateur compare deux configuration de calcul de fractale, l’une qui consiste à concentrer dans une seule VM tout le traitement de calcul, et l’autre qui se fait sur 3 VMs. Le simulateur propose 8 *jobs* pour effectuer des tests, mais d’autres jobs peuvent être ajoutés.

```

Total traffic: 0.0826(Gb), Traffic power consumption: 8.26(Wh)
Total consumption: 44.22(Wh)
=====
Execution time: 0:03:24
Correctness: fact.txt - OK
prime.txt - OK
sink_aggregation.txt - OK
transport_stream.txt - OK
template.txt - OK
decode_frame.txt - OK
get_x_max.txt - OK
store_prices.txt - OK
VM consumption: 30.29 (Wh)
Idle compute consumption: 5.67 (Wh)
Total traffic: 0.0826(Gb), Traffic power consumption: 8.26(Wh)
TOTAL: 44.22(Wh)
=====

```

FIGURE 1.

Le modèle énergétique du simulateur. Dans le Cloud, les trois principaux types de ressources sont : le compute, le stockage et le réseau. Dans (Zhu *et al.*, 2004) estime le stockage à 27% de consommation électrique tandis que (Heddeghem *et al.*, 2014) estime le réseau à 37%. Cette variabilité pourrait s’expliquer par les infrastructures mises en place ou la charge de travail générée par les clients. Nous avons conduit une étude en OUTSCALE qui estime la distribution d’énergie à : 90% au compute, 7% au stockage et moins de 3% au réseau.

Au regard de cette distribution, le compute apparaît comme la ressource plus énergivore. Dans les modèles proposés dans l’état de l’art (Ahvar *et al.*, 2019), cette consommation est imputée à deux facteurs : *énergie statique*, consommation de base indépendante de la configuration, du déploiement ou de la charge de travail et une *énergie dynamique*, consommation induite par les choix de l’infrastructure et les comportements des utilisateurs finaux. L’*énergie liée au réseau interne* dans l’infrastructure IaaS peut se calculer selon le trafic généré entre les VMs. Notre simulateur calcule une estimation de ces trois sources d’énergie grâce aux informations collectées pendant le temps de fonctionnement des *jobs* basées sur la configuration du déploiement IaaS; le temps et d’autres mesures d’exécution; et les informations de consommation énergétique fournies par PowerTOP (github.com/fenrus75/powertop).

Premiers résultats comparatifs. Les deux gagnants du Hackathon ont réussi à économiser 50% de l’énergie en utilisant des approches orthogonales : une équipe s’est concentrée sur l’optimisation de l’infrastructure en réduisant autant que possible les nombres des VMs, et l’autre équipe a opté pour une optimisation des *jobs* sans toucher l’infrastructure. Bien que ces deux approches sont intuitives pour un design moins énergivore, il n’est pas toujours possible de les appliquer de façon aussi radicale. Des approches hybrides s’imposent avec des optimisations partielles dans l’infrastructure et dans les *jobs* à faire tourner. Notre simulateur permet de donner une

perspective pour des optimisations hybrides de façon simplifiée et permet de guider les utilisateurs dans leur choix pour une éco-responsabilité numérique.

3. Conclusion et perspectives

Les clients du Cloud ont besoin d'outils qui leurs permettent de prendre des décisions éco-responsables de leur utilisation. Pour y contribuer, nous proposons un simulateur déployable dans le Cloud, adaptable à l'architecture des systèmes, des composants utilisés, des langages de programmation et des frameworks. Notre prototype a démontré son potentiel en mettant en évidence l'effet, ou le manque d'effet, de différents types d'optimisations. Plusieurs pistes pourront être explorées pour perfectionner le prototype : améliorer l'estimation en prenant en compte d'autres facteurs, tels que le GPU, le stockage, etc. ; ajouter des scénarios qui poussent à la recherche d'approches d'optimisation hybrides entre l'infrastructure et les jobs.

Bibliographie

- Ahvar, Orgerie, Lebre. (2019). Estimating energy consumption of cloud, fog and edge computing infrastructures. *IEEE Transactions on Sustainable Computing*.
- Calheiros, Ranjan, Beloglazov, AF D. R., Buyya. (2011). Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software: Practice and experience. Wiley Online Library*.
- Desjardins, Ang, Lu. (2014). *Cloud computing growth-visual capitalist*.
- Heddeghem V., Lambert, Lannoo, Colle, Pickavet, Demeester. (2014). Trends in worldwide ICT electricity consumption from 2007 to 2012. *Computer Communications*.
- Kliazovich, Bouvry, Khan. (2012). Greencloud: a packet-level simulator of energy-aware cloud computing data centers. *The Journal of Supercomputing*.
- Kumar, Kumar. (2019). Issues and challenges of load balancing techniques in cloud computing: A survey. *ACM Computing Surveys*.
- Lefebvre, Kumar, Chiky. (2014). Simizer: Evaluating consistency trade offs through simulation. In *PaPEC'14*.
- Murugesan, Gangadharan. (2012). Green IT: an overview. *Harnessing green IT: Principles and practices*.
- Verdecchia, Lago, D V. (2021). LEAP technology landscape: Trends and scenarios. *Persistent URL: <https://tinyurl.com/leapLandscape>. Accessed.*
- Verdecchia, Lago, Ebert, Vries D. (2021). Green it and green software. *IEEE Software*.
- Wickremasinghe, Calheiros, Buyya. (2010). Cloudanalyst: A cloudsim-based visual modeller for analysing cloud computing environments and applications. In *AINA'10*.
- Zhu, M D., Devaraj, Li, Zhou, Cao. (2004). Reducing energy consumption of disk storage using power-aware cache management. In *HPCA'04*.

RADIANCE: A tool for software behavior design and energy consumption categorization

Jorge Andrés Larracoechea¹, Philippe Roose², Sergio Ilarri³, Sébastien Laborie⁴, Yudith Cardinale⁵

1. LIUPPA, Université de Pau et des Pays de l'Adour
2 Allée du Parc Montaury, 64600 Anglet, France
jorge-andres.larracoechea@etud.univ-pau.fr
2. LIUPPA, Université de Pau et des Pays de l'Adour
2 Allée du Parc Montaury, 64600 Anglet, France
Philippe.Roose@univ-pau.fr
3. I3A, Universidad de Zaragoza
María de Luna 3, 50018 Zaragoza, Spain
silarri@unizar.es
4. LIUPPA, Université de Pau et des Pays de l'Adour
2 Allée du Parc Montaury, 64600 Anglet, France
Sebastien.Laborie@univ-pau.fr
5. Universidad Internacional de Valencia
C. del Pintor Sorolla, 21, 46002 Valencia, Spain
ycardinale@usb.vc

ABSTRACT. With the recent increase of the research community's attention, green software, also called sustainable software, has gained the spotlight as an opportunity for cutting back on the electrical energy expense world-wide. As a result, new approaches that cover the development, deployment, and execution of software have been garnered with the promise of a more efficient use of computational resources. Nevertheless, their adoption has been hindered by technicalities, and a lack of consistent guidelines from the analysis and design stages of the development.. In this paper, we present RADIANCE: a proof of concept web app for designing greener software with a model-driven approach based on the Behavior-Based Consumption Profiles (BBCP) Domain-Specific Modeling Language. RADIANCE, in contrast to other tools, embraces users with different levels of knowledge on green software and the BBCP approach, simplifying green software design. Moreover, RADIANCE assesses and rates the energy consumption of the profiles generated by the user, assisting the user to identify possible design changes that promote the creation of greener software.

KEYWORDS: Green Software Design, Green Software, Sustainable Software, Energy consumption, Software development

1. Introduction

With the recent increase of awareness on sustainability as a strategy for addressing the on-going energy and environmental crisis, green software has regained the spotlight as an opportunity to reduce the electrical energy expense. In our previous research (Larracochea et al., 2022), we noticed a lack of green software approaches that target the analysis and design phases of the Software Development Life-Cycle (SDLC), with a majority of them targeting the development, testing and the deployment and maintenance phases, as seen in Table 1.

Table 1 Distribution of green within software tools throughout the SDLC

Contribution	Development	Testing	Deployment & Maintenance
(Aggarwal et al., 2015)	✓	✓	✓
(Hao et al., 2013)	✓	✓	✓
(Manotas et al., 2014)	✓	×	×
(Pathak et al., 2012)	✓	✓	✓
(Peltonen et al., 2015)	×	×	✓
(Oliner et al., 2013)	×	✓	×

Furthermore, as previous studies (Pang et al., 2015) conclude, industry programmers and university students are aware of the electrical energy expenditure associated to software execution, but, they are hindered by the lack of guidelines and tools to implement existing techniques with ease. The same can be said about the software development industry, where a previous study (Jagroep et al., 2017) identified that engineers consider software's energy efficiency as a parameter for successful software, but stakeholders neglect the expense of the overhead time that this goal demands.

In the following sections we present our tool, RADIANCE (softwaRe behAvior DesIgn And eNergy Consumption assEssment). RADIANCE is a web-based software behavior design and analysis tool based on the Behavior-Based Consumption Profiles (BBCP) Domain-Specific Modeling Language (DSML). RADIANCE's goal is twofold: (1) generate a consumption profile of the software designed on it so that energy consumption can be estimated from the analysis and

design phases of the SDLC and (2) blend the topic of green software with software design in order to lower the barrier of entry to green software development for students and professionals. The latter goal has a high priority to us, as we developed RADIANCE with an emphasis on usability and customizable levels of expertise that maximize its future adoption.

2. The core features of RADIANCE

RADIANCE is a tool tailored for the BBCP DSML. The BBCP, in difference to other approaches, provides users with a series of properties that allow them to express the intensity of software consumption over time: how much software usage changes throughout a given set of hours, days or specific time-based events. RADIANCE supports the BBCP with 3 types of artifacts: schemas, profiles and collections. Schemas define the properties that constitute a profile. A profile is an instance of a schema filled with values that constitute the behavior of a software unit. Collections are agglomerated profiles that can represent a nuclear unit of software, such as a microservice, up to complete applications.

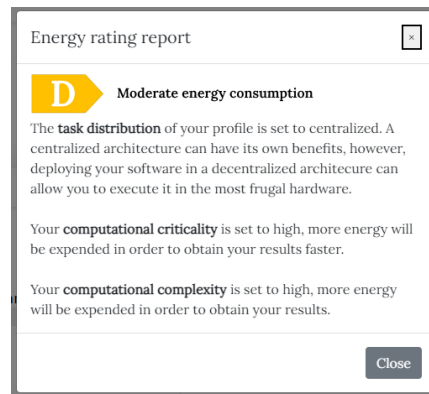
Figure 1 The UI of RADIANCE's builder



The UI of RADIANCE is divided into 2 main components: the manager and the builder. The manager is a library where the user can add, delete, duplicate and instance artifacts. The builder, visible in Figure 1 The UI of RADIANCE's builderFigure 1, is the most important component as it is responsible for assisting

the user throughout the process of software profiling with several mechanisms: a customizable UI, a dynamic sequence diagram of the designed operations, a dynamic step diagram that labels the energy consumption of parallel operations with shades of color and, finally, an algorithm that delivers a report with important points concern from an energy consumption stand point as seen in Figure 2 .

Figure 2 A sample energy rating report generated with RADIANCE



4. Demonstration

The demonstration will be divided in 2 parts: (1) an in deep explanation of the features available in RADIANCE and (2) a guided example of the creation of a profile that resembles GeForce Now, a cloud gaming platform. We selected GeForce Now due to its modern architecture that utilizes a cloud cluster in the network to render the video and sound stream, which is then sent back to the user by establishing an active exchange of data that negatively impacts the energy consumption of this software type. In addition, the profile will reflect the impact of the business model's constraints that affect the energy consumption, utilizing properties unique to the BBCP that constrain the usage of the profile through time.

5. Future work

There are still important validations that we need to perform before asserting that RADIANCE will in fact promote the design of greener software and introduce new users to it. To begin, we plan a usability test in the near future with university students in the IT field. The objective of the test will be to profile a specific example of a software unit and assess how many students can successfully achieve an energy rating report after a quota of time. Furthermore, we have a second usability test planned with several experts in software engineering and software design, so that we can insure that RADIANCE will be able to satisfy the needs of users with diverse levels of expertise.

Acknowledgements:

This publication belongs to the project PID2020-113037RB-I00, funded by MCIN/AEI/ 10.13039/501100011033. We also thank the support of the Government of Aragon (COSMOS research group; last group reference: T64_23R).

Bibliography

- Aggarwal, K., Hindle, A., Stroulia, E., 2015. GreenAdvisor: A tool for analyzing the impact of software evolution on energy consumption, in: 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME). Presented at the 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME), pp. 311–320. <https://doi.org/10.1109/ICSM.2015.7332477>
- Hao, S., Li, D., Halfond, W.G.J., Govindan, R., 2013. Estimating mobile application energy consumption using program analysis, in: 2013 35th International Conference on Software Engineering (ICSE). Presented at the 2013 35th International Conference on Software Engineering (ICSE), pp. 92–101. <https://doi.org/10.1109/ICSE.2013.6606555>
- Jagroep, E., Broekman, J., Werf, J.M.E.M. van der, Lago, P., Brinkkemper, S., Blom, L., Vliet, R. van, 2017. Awakening Awareness on Energy Consumption in Software Engineering, in: 2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering in Society Track (ICSE-SEIS). Presented at the 2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering in Society Track (ICSE-SEIS), pp. 76–85. <https://doi.org/10.1109/ICSE-SEIS.2017.10>
- Larracochea, J., Roose, P., Ilarri, S., Cardinale, Y., Laborie, S., González, M., 2022. Towards Services Profiling for Energy Management in Service-oriented Architectures. Presented at the 17th International Conference on Web Information Systems and Technologies, pp. 209–216.
- Manotas, I., Pollock, L., Clause, J., 2014. SEEDS: a software engineer’s energy-optimization decision support framework, in: Proceedings of the 36th International Conference on Software Engineering, ICSE 2014. Association for Computing Machinery, New York, NY, USA, pp. 503–514. <https://doi.org/10.1145/2568225.2568297>
- Oliner, A.J., Iyer, A.P., Stoica, I., Lagerspetz, E., Tarkoma, S., 2013. Carat: collaborative energy diagnosis for mobile devices, in: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, SenSys ’13. Association for Computing Machinery, New York, NY, USA, pp. 1–14. <https://doi.org/10.1145/2517351.2517354>
- Pang, C., Hindle, A., Adams, B., Hassan, A.E., 2015. What Do Programmers Know about Software Energy Consumption? IEEE Software 33, 1–1. <https://doi.org/10.1109/MS.2015.83>
- Pathak, A., Hu, Y.C., Zhang, M., 2012. Where is the energy spent inside my app? fine grained energy accounting on smartphones with Eprof, in: Proceedings of the 7th ACM European Conference on Computer Systems, EuroSys ’12. Association for Computing Machinery, New York, NY, USA, pp. 29–42. <https://doi.org/10.1145/2168836.2168841>
- Peltonen, E., Lagerspetz, E., Nurmi, P., Tarkoma, S., 2015. Energy modeling of system settings: A crowdsourced approach, in: 2015 IEEE International Conference on Pervasive Computing and Communications (PerCom). Presented at the 2015 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 37–45. <https://doi.org/10.1109/PERCOM.2015.7146507>

CircuSChain : une méthode pour améliorer la circularité dans les chaînes logistiques

Une adaptation du framework As-Is/As-If

Asiye Kurt^{1,2}, Mario Cortes-Cornax¹, Agnès Front¹, Van-Dat Cung², Fabien Mangione²

1. Univ. Grenoble Alpes, CNRS, Grenoble INP* , LIG
prenom.nom@univ-grenoble-alpes.fr

2. Univ. Grenoble Alpes, CNRS, Grenoble INP*, G-SCOP
prenom.nom@grenoble-inp.fr

RESUME L'économie circulaire représente une solution potentielle aux challenges imposés par la diminution des ressources naturelles et la pollution de l'environnement. Dans le domaine de la logistique, les entreprises cherchent à revoir leurs pratiques et à adopter les principes de l'économie circulaire. En pratique, les entreprises essayent de rendre leurs chaînes logistiques davantage circulaires. Cet article présente une méthode d'évolution continue appelée CircuSChain dont l'objectif est de guider les entreprises pour améliorer la circularité dans leurs chaînes logistiques.

ABSTRACT. The Circular Economy has emerged as a potential solution to the challenges due to natural resource depletion and environmental pollution. Pressure from laws, stakeholders and customers leads companies to review their practices and adopt the Circular Economy to improve the circularity of their supply chains. In practice, companies look for a way to make their supply chains more circular. A successful transition towards circular supply chains requires continual measurement of progress towards circularity. The main contribution of this paper is a method called CircuSChain that aims at guiding organizations in order to improve the circularity of their supply chains.

Mots-clés : Méthode d'évolution continue, Chaîne logistique, Économie circulaire.

KEYWORDS: Continual Evolution Method, Supply Chain, Circular Economy

* Institute of Engineering, Univ. Grenoble Alpes
30000 Grenoble, France

1. Introduction

L'économie circulaire, par opposition à l'approche linéaire (« take-make-dispose »), vise à minimiser les apports de ressources, les déchets et la pollution en (ré)utilisant le plus longtemps possible les produits, composants et matériaux (Geissdoerfer et al., 2017). De fait, la circularité est fortement liée aux Cycles de Vie des produits, de la conception jusqu'à la gestion de fin de vie, en passant par l'approvisionnement, la production, la distribution, l'usage, la collecte, etc. Des stratégies de conception de produits et des activités « E-o-L » (End of Life), également appelées activités d'économie circulaire, doivent donc être plus spécifiquement réalisées pour cela, telles que la réutilisation, la remise à neuf, le recyclage, la refabrication, etc. L'économie circulaire est guidée par les principes suivants : (1) appliquer plusieurs activités d'économie circulaires en parallèle (Blomsma et Brennan, 2017), (2) utiliser et réutiliser des matériaux (Genovese et al., 2017), et (3) favoriser des boucles ouvertes entre des secteurs distincts par le biais de la réaffectation, qui consiste à transférer des produits inutiles d'une chaîne logistique dans un domaine d'activité vers un autre domaine d'activité, pour un produit différent (Farooque et al., 2019). D'un point de vue systémique, l'ensemble de ces activités tout au long du Cycle de Vie d'un produit, forme une chaîne logistique constituant une brique fondamentale dans la transition vers l'économie circulaire car elle est au cœur du système de conception et fabrication du produit. Les chaînes logistiques sont décrites comme des structures « gérant les apports de biens ou de services comprenant une gamme d'activités pour les utilisateurs finaux depuis l'approvisionnement en matières premières jusqu'à la fin de vie des produits » (Teck-Yong, 2005 ; Ahi et Searcy, 2013). Diverses notions dans la littérature, telles que Logistique inverse (*Reverse logistics*) ou encore Chaîne logistique verte (*Green Supply Chain*), traitent partiellement de ces principes d'économie circulaire (Jain et al., 2018). Une chaîne logistique circulaire est donc considérée comme une convergence de toutes les notions mentionnées soutenant les principes d'économie circulaire.

La figure 1 montre un exemple informel d'une chaîne logistique circulaire. Dans cet exemple, la chaîne logistique circulaire est composée de 2 chaînes logistiques destinée à la fabrication de produits différents (*Product 1*, par exemple une voiture électrique, et *Product 2*, par exemple des lampadaires). Chacune de ces chaînes logistiques comporte des activités linéaires (e.g. *Linear Supply Chain 1*) telles que l'extraction de matières premières, la fabrication des composants (les roues de la voiture, le capot, le moteur, etc.), l'assemblage du produit, la distribution au client, puis en fin de chaîne, la collecte du produit, voire sa destruction. Suite à la collecte du produit usagé, une partie du produit collecté (par ex, certains composants de la voiture) peuvent être réutilisés et réinsérés dans la même chaîne logistique pour fabriquer d'autres voitures : la chaîne logistique du produit 1 est alors considérée comme une *chaîne logistique circulaire en boucle fermée* car elle comporte des activités de réutilisation dans la même chaîne logistique. D'autres parties du produit collecté peuvent également être réutilisées pour la fabrication d'autres types de produits et intégrées ainsi dans d'autres chaînes logistiques. Par exemple, la batterie d'une voiture électrique usagée pourrait être réutilisée comme élément de batterie

d'un lampadaire (e.g. *Linear Supply Chain 2*). On parle alors d'une *chaîne logistique circulaire en boucle ouverte* permettant de réutiliser des composants ou produits pour une autre activité que celle pour laquelle ils avaient été conçus au départ.

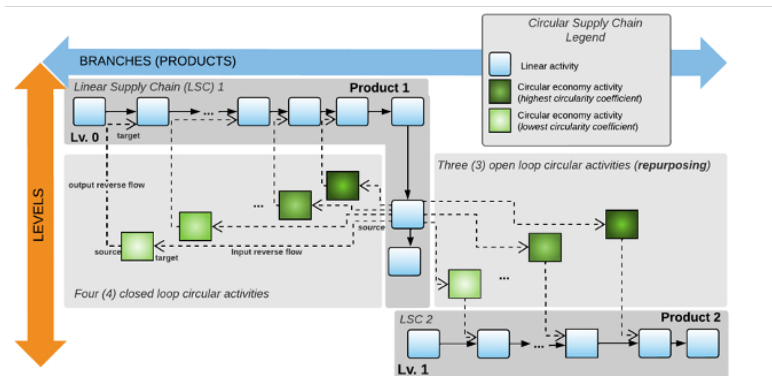


Figure 1. Chaîne logistique circulaire : un exemple informel

La pression des lois, des parties prenantes et des clients conduit les entreprises à revoir leurs pratiques pour adopter les principes de l'économie circulaire et améliorer la circularité de leurs chaînes logistiques. Cependant, une transition réussie vers une chaîne logistique circulaire nécessite une mesure continue de la progression vers la circularité (Jain et al., 2018). Le manque de connaissances et de sensibilisation sur les chaînes logistiques circulaires constitue un obstacle difficile pour les gestionnaires de chaînes logistiques et les entreprises désireuses d'améliorer leur circularité. De nouveaux outils et méthodes sont nécessaires pour promouvoir les chaînes logistiques circulaires et accompagner leur conception et leur évolution.

Cet article propose une méthode appelée CircuSChain pour améliorer la circularité dans les chaînes logistiques. La méthode proposée s'appuie sur des travaux antérieurs :

- le framework As-Is/As-If (Cela et al., 2019) destiné aux ingénieurs de méthodes pour les aider à développer des méthodes d'évolution continue. Le framework As-Is/As-If propose un modèle de processus et un méta-modèle de produit qui peuvent être considérés comme des modèles dans la construction de méthodes d'évolution continue ;

- plusieurs outils tels qu'un indicateur de circularité, un outil de classification des indicateurs et un jeu sérieux sur le thème des chaînes logistiques circulaires (Kurt, 2021). La méthode CircuSChain « orchestre » les différents outils proposés afin d'obtenir une amélioration continue de la circularité des chaînes logistiques.

Dans la suite de cet article, la section 2 présente des travaux connexes de l'état de l'art en se concentrant sur les méthodes qui pourraient être appliquées pour améliorer la circularité des chaînes logistiques. La section 3 fournit une définition

détaillée de la notion de chaîne logistique circulaire à l'aide d'un modèle générique de chaînes logistiques décrit en UML. L'aperçu de la méthode est présenté dans la section 4, suivi dans la section 5 d'un cas d'utilisation illustratif dans le domaine du textile. La section 6 discute de l'apport de la construction de la méthode CircuSChain dans l'amélioration du framework As-Is/As-If. Enfin, la section 7 conclut l'article et présente quelques perspectives de recherche.

2. Travaux connexes

Les méthodes d'évolution continue répondent à divers besoins d'amélioration allant d'améliorations mineures à des changements radicaux (Cela, 2021). Cette vision holistique de l'évolution continue, que nous proposons dans des travaux antérieurs (Cela et al., 2019), a un impact sur les méthodes à utiliser pour mettre en œuvre cette évolution. Notre point de vue est que ces méthodes doivent d'une part s'appuyer sur des cycles d'évolution continue (par opposition aux approches par projet qui ont un budget et des dates délimités) et d'autre part favoriser au maximum la participation active et collective des acteurs de l'organisation. En ce sens, les approches ludiques et participatives sont particulièrement adaptées afin de réduire la barrière à l'entrée pour les participants. Nous prenons également en compte le niveau de formalisation de la méthode, par exemple par la proposition d'un méta-modèle de produit pour définir les différents composants de la méthode et leurs relations, peut aider la conception de l'outil support à la méthode. De plus, la spécificité du domaine est un avantage important d'une méthode afin d'obtenir une orientation plus fine si les utilisateurs finaux ne sont pas des experts du domaine. L'ensemble de ces critères nous permet de comparer les travaux connexes dans le tableau 1 selon les considérations mentionnées ainsi que de positionner notre approche.

Différents exemples de méthodes d'évolution ou d'amélioration continue peuvent être trouvés dans la littérature. *PDCA (Plan, Do, Check, Act)* (Deming, 2018), *DMAIC (Define, Measure, Analyze, Improve, Control)* (Deep et al., 2018), *8D (8 Disciplines)* (Zarghami et Benbow, 2017) et *A3* (Shook, 2008) sont des méthodes génériques utilisant principalement une approche projet. Bien que la plupart d'entre elles aient été appliquées pour améliorer les applications ou les chaînes logistiques vertes/durables, elles n'ont pas été particulièrement appliquées pour améliorer la circularité des chaînes logistiques. De plus, aucune de ces méthodes ne favorise une approche participative. À l'exception de DMAIC (Deep et al., 2018), les méthodes susmentionnées n'ont pas été formalisées, ce qui rend difficile la construction d'outils support à ces méthodes. Ces méthodes adoptent et promeuvent une interprétation et une application ad hoc. Par ailleurs, *UN Global Compact Business for Social Responsibility* (UN, 2015) propose un rapport intitulé « *Supply Chain Sustainability : A Practical Guide for Continuous Improvement* » contenant six étapes : *Commit, Assess, Define, Implement, Measure, et Communicate*. Ce guide vise à améliorer la durabilité d'une chaîne logistique et propose une approche participative. Malgré la spécificité de son domaine et son approche participative, il n'est cependant pas spécifiquement dédié aux chaînes logistiques circulaires ni formalisé.

Table 1. Comparaison de méthodes et positionnement de la méthode CircuSChain

Critères	Méthodes d'évolution continue		Techniques de résolution de problème		Méthodes dédiées	
	PDCA	DMAIC	8D	A3	UN Supply Chain Sustainability	CircuSChain
Appliqué aux chaînes logistiques	✓	✓	✓		✓	✓
Appliqué aux applications «green» / responsables	✓	✓		✓	✓	✓
Appliqué aux chaînes logistiques circulaires						✓
Formalisation		Méta-modèle, outil support				Méta-modèle, outil support en construction
Approche participative					✓	✓

Le framework As-Is/As-If a été intrinsèquement construit et formalisé pour soutenir les cycles d'évolution continue et promouvoir une approche participative afin d'instiller dans les entreprises une culture de l'évolution continue. La méthode CircuSChain, qui est une adaptation de ce framework, hérite de ces caractéristiques. De plus, elle est dédiée aux chaînes logistiques circulaires.

3. Chaînes logistiques circulaires : définition et formalisation

Les principales caractéristiques d'une chaîne logistique circulaire s'appuient sur les principes d'économie circulaire suivants :

- **C0- Au moins une activité d'économie circulaire.** Une structure de chaîne logistique circulaire doit contenir au moins une activité d'économie circulaire (réutilisation, remise à neuf, recyclage, etc.) en plus des activités linéaires de la chaîne logistique (extraction de matériaux, production, distribution, utilisation, collecte et destruction).

- **C1- Boucles consécutives d'utilisation du matériel.** Cette caractéristique est basée sur le principe que les cycles consécutifs d'utilisation des produits, composants ou matériaux usagés prolongent leur durée de vie (« *power of circling longer* ») (Ellen Mac Arthur Fondation, 2013, 2014). Par exemple, un téléphone

portable peut être reconditionné et réutilisé en seconde main (2^{ème} boucle), puis ses composants seront eux-mêmes réutilisés pour fabriquer d'autres produits (non nécessairement des téléphones) (3^{ème} boucle), et enfin les composants seront collectés pour être détruits (4^{ème} boucle).

- **C2- Plusieurs options d'activités d'économie circulaire simultanées :** « *power of inner circle* » (Ellen Mac Arthur Fondation, 2013 ; 2014) et « *inertia principle* » (Stahel, 1982 ; 2010) Ainsi, plus la boucle est serrée, moins un produit doit être changé lors du retraitement. Par conséquent, avoir des boucles plus courtes implique une circularité plus élevée. Avoir plusieurs activités d'économie circulaire dans une chaîne logistique implique que l'activité la plus appropriée peut être choisie pour traiter un produit usagé en fonction de sa qualité. Cela évite une perte de valeur en termes d'énergie, de matériaux, de main-d'œuvre et de pollution.

- **C3- Boucles ouvertes (« *open-loops* ») et intégration de chaînes logistiques distinctes.** Cette caractéristique repose sur l'utilisation en cascade d'un produit dans différentes chaînes logistiques : « *power of cascaded use* » (Ellen Mac Arthur Fondation, 2013 ; 2014), qui permet à un produit d'être réutilisé pour une autre activité que celle pour laquelle il avait été conçu au départ. Par exemple, les batteries de véhicules électriques peuvent être réutilisées pour être utilisées comme batteries alimentant des lampadaires.

La figure 2 présente sous forme d'un diagramme de classes UML un modèle générique permettant de caractériser les différents composants d'une chaîne logistique circulaire selon ces différentes caractéristiques. Un coefficient de circularité (*circularityCoefficient*) est défini pour chaque activité circulaire afin de déterminer son degré de circularité.

L'exemple informel présenté dans l'introduction sert d'illustration en partie basse de la figure. Les activités circulaires sont représentées par des carrés verts. Les carrés bleus représentent les activités linéaires telles que l'extraction, la production, la distribution et l'utilisation. Les activités circulaires telles que la réutilisation, la refabrication, le recyclage, etc., ont des dégradés de couleurs différents en fonction de leur coefficient de circularité. Plus les activités sont sombres, plus leur circularité est grande.

Une chaîne logistique a deux attributs dérivés selon le type de produit (branche) et ses exigences (niveau) :

- une branche fait référence au type de produit fabriqué tout au long de la chaîne logistique.

- un niveau fait référence aux exigences du produit fabriqué le long de la chaîne logistique linéaire associée : la remise à neuf étant la réutilisation de produits usagés pour des applications moins exigeantes que les applications d'origine, les produits des chaînes logistiques linéaires intégrées ont des exigences moindres : par exemple, les exigences sont moindres pour des téléphones reconditionnés que pour des téléphones neufs.

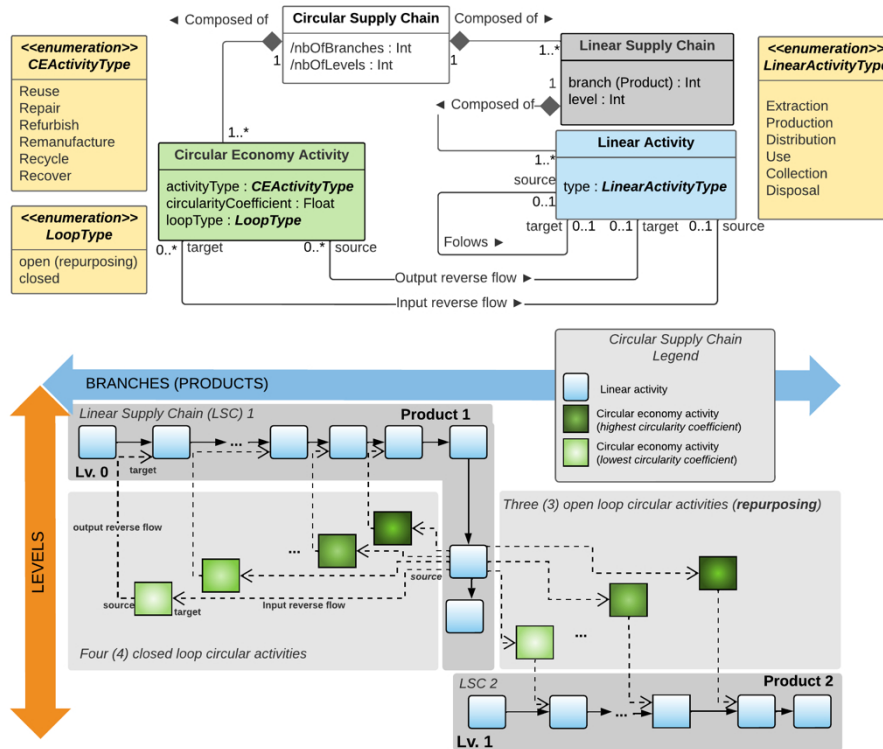


Figure 2. Modèle générique de chaîne logistique circulaire

Ce modèle est au cœur de la méthode CircuSChain car il définit les différents composants dans lesquelles l'amélioration de la circularité peut être appliquée. Comme nous le verrons dans la section suivante, ce modèle sera intégré dans un modèle plus large qui formalise les différents aspects de l'évolution continue, comme les points bloquants, les indicateurs ou les changements.

4. Présentation générale de la méthode CircuSChain

Comme mentionné précédemment, le framework As-Is/As-If (Cela et al., 2019) est composé d'un modèle de processus exprimé avec le formalisme intentionnel MAP (Rolland et al., 1999) et d'un méta-modèle de produit formalisé en UML. Les modèles MAP sont des graphes orientés avec des nœuds représentant des intentions et des arcs capturant des stratégies pour atteindre ces buts. La figure 3 présente la map de 1^{er} niveau représentant la vue générale du modèle de processus et les méta-classes de base du méta-modèle de produit du framework As-Is/As-If. En tant qu'adaptation du framework As-Is/As-If, la méthode CircuSChain est également composée d'un modèle de processus et d'un méta-modèle de produit. Comme

mentionné précédemment, CircuSChain est une méthode participative impliquant les acteurs de la chaîne logistique afin d'améliorer la circularité d'une chaîne logistique. Les stratégies de la méthode sont décrites à travers des protocoles et mises en œuvre à travers des séances animées par un animateur avec la participation des acteurs. Un exemple de protocole est donné au paragraphe 5.2. (tableau 2). Les autres protocoles sont accessibles en ligne[†].

La méthode CircuSChain vise à faire évoluer une chaîne logistique appelée *As-Is Supply Chain*, en une ou plusieurs chaînes logistiques plus circulaires, appelées *As-If Supply Chains*. La chaîne logistique « *As-Is* » peut être soit une chaîne logistique linéaire dans laquelle l'objectif est d'intégrer la circularité, soit une chaîne logistique circulaire dans laquelle l'objectif est d'améliorer la circularité. Par conséquent, le modèle de processus est composé de deux intentions « *Characterize As-Is Supply Chain* » et « *Imagine As-If Supply Chain* ». La stratégie d'analyse vise à identifier les composants de la chaîne logistique « *As-Is Supply Chain* » et à mesurer sa circularité. La stratégie de diagnostic vise à diagnostiquer la chaîne logistique afin de définir des objectifs et des contraintes pour améliorer sa circularité. La stratégie d'évolution vise à faire évoluer la chaîne logistique en une ou plusieurs autres chaînes logistiques plus circulaires appelées « *As-If Supply Chains* ».

Concernant le méta-modèle de produit, le package principal contient la méta-classe *Supply Chain (SC)* et ses deux sous-classes : *As-Is Supply Chain* et *As-If Supply Chain*. Le package d'analyse est composé de deux méta-classes : *Component* et *Circularity Indicator*. La méta-classe *Component* est ensuite spécialisée selon le modèle générique présenté en section 3. Dans le package diagnostic, la notion de « *Best Available Supply Chain* » est introduite afin de diagnostiquer une chaîne logistique par rapport au cas idéal de chaîne logistique circulaire dans le même domaine. Ainsi lors du diagnostic, des écarts entre la chaîne logistique « *As-Is* » et le cas idéal peuvent être détectés ; améliorer la circularité d'une chaîne logistique consistera ainsi à essayer de réduire ces écarts par rapport au cas idéal dans le même domaine. Enfin, dans le package d'évolution, une chaîne logistique « *As-Is* » pourrait évoluer vers une chaîne logistique « *As-If* ». La méta-classe *Change* représente les changements nécessaires pour atteindre une chaîne logistique plus circulaire. Un changement est modélisé industriellement par des changements opérationnels et est évalué afin de déterminer si le problème associé est résolu.

5. Illustration de 3 stratégies par l'exemple

Dans cette section, nous présentons une illustration de la méthode CircuSChain sur une chaîne logistique dans le domaine du textile. L'évolution du modèle de la chaîne logistique et de l'indicateur de circularité est présentée pour mieux expliquer les étapes de la méthode.

[†]https://drive.google.com/file/d/1hlWQckgH_UweepI0qQEe_rZy8_fVwqRV/view?usp=sharing

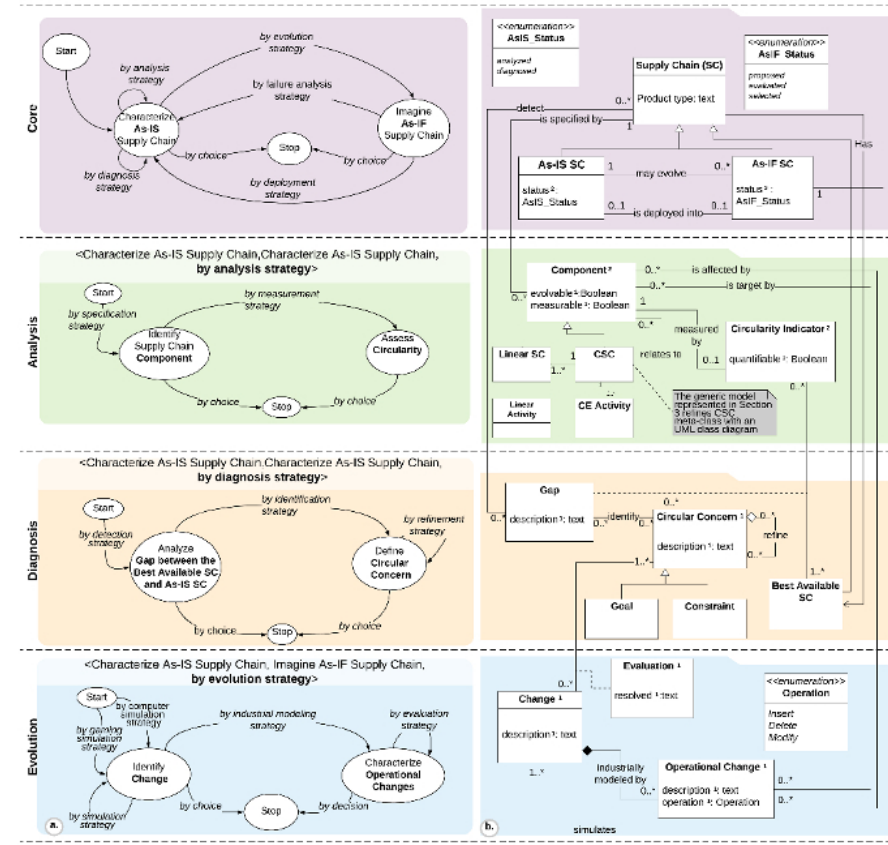


Fig. 3. Modèle de processus (a) et méta-modèle de produit (b) de la méthode CircuSChain

5.1. Stratégie d'analyse

La stratégie d'analyse « *by analysis strategy* » vise à caractériser la chaîne logistique actuelle « *As-Is* » en identifiant d'abord ses composants. Dans un premier temps, l'animateur décrit la chaîne logistique avec des informations recueillies lors d'entretiens téléphoniques avec les acteurs, telles que le type de produit, les acteurs, les principales activités, etc. Ensuite, la chaîne logistique est modélisée par des experts de la chaîne logistique à travers un groupe de discussion en utilisant le modèle générique présenté dans la section 3. Un exemple de modèle obtenu est illustré en figure 4 dans le domaine du textile.

Ensuite, la circularité de la chaîne logistique « *As-Is* » doit être mesurée. La mesure commence par le choix des dimensions de circularité à mesurer et des

indicateurs de circularité. Les participants sélectionnent les indicateurs par le biais d'un groupe de discussion en utilisant un outil de classification proposé dans des travaux précédents (Kurt et al., 2021). Ensuite, la circularité de la chaîne logistique est calculée en fonction des indicateurs choisis. L'indicateur de circularité global issu de (Kurt, 2021) peut être utilisé à cette fin.

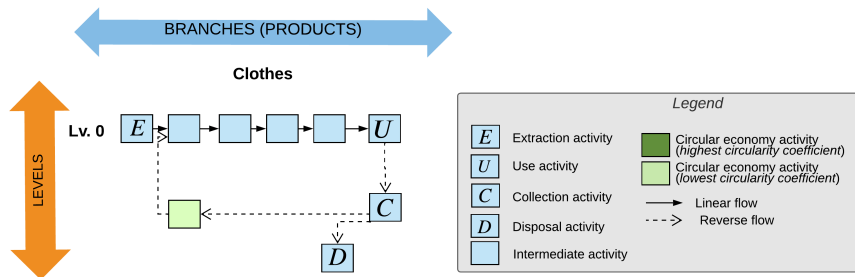


Figure 4. Un exemple de chaîne logistique « As-Is » pour le domaine Textile, après exécution de la stratégie d'analyse

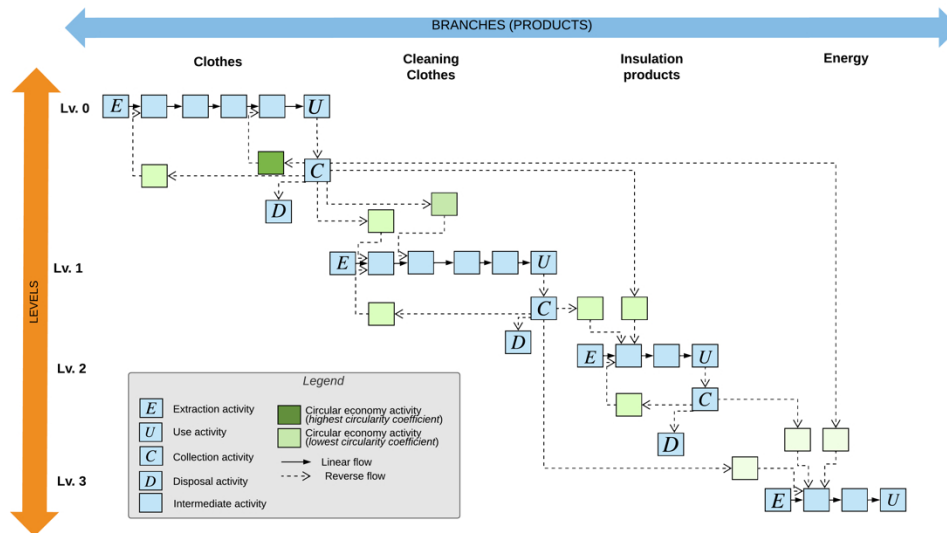
Dans l'exemple du domaine Textile, nous supposons que les participants choisissent comme indicateur de circularité la mesure de la pollution. L'animateur calcule alors la circularité de la chaîne logistique « As-Is » selon cet indicateur et constate que cette circularité est actuellement calculée à 0,03.

5.2. Stratégie de diagnostic et « Cas idéal de Chaîne logistique »

La stratégie de diagnostic « *by diagnosis strategy* » vise à diagnostiquer la chaîne logistique actuelle en tenant compte de son écart en terme de mesure de circularité, avec un « cas idéal de chaîne logistique » selon le domaine et le type de produit concerné.

La structure d'une chaîne logistique et sa circularité peuvent varier considérablement selon le type de produit. En effet, certaines activités circulaires ne sont pas applicables à tous les types de produits (par exemple, certains vêtements ne peuvent pas être reconditionnés). Le cas idéal de chaîne logistique contient toutes les activités circulaires possibles pour le type de produit donné. Il est identifié grâce à la littérature et les caractéristiques de la chaîne logistique (type de produit et choix des indicateurs de circularité à améliorer). La figure 5 montre un exemple de modèle du cas idéal de chaîne logistique dans le domaine Textile. Dans ce cas idéal, différentes activités circulaires sont possibles pour les vêtements : réutilisation en boucle fermée (redistribution des vêtements via des friperies), réutilisation en boucle ouverte (réutilisation pour la fabrication d'autres types de produits tels que des matériaux d'isolation), réutilisation en boucle fermée et en boucle ouverte... La chaîne logistique linéaire pour les matériaux d'isolation est basée sur les travaux de

(Pavel et Blagneva, 2022). Les produits qui ne peuvent pas être réutilisés ou recyclés peuvent être utilisés pour produire de l'énergie.



L'analyse de l'écart entre la chaîne logistique idéale et la chaîne logistique « *As-Is* » peut également conduire au calcul des indicateurs retenus pour le cas idéal en fonction de la pollution émise pour chaque activité, ainsi que la quantité optimale de produits collectés et leur répartition entre les activités circulaires. Dans l'exemple, la circularité de la chaîne logistique idéale est calculée à 0,11, cette valeur optimale étant extraite à partir de la littérature ou d'entretiens avec des acteurs industriels. Ensuite, l'écart entre les deux chaînes logistiques est discuté avec les membres participants, et la sous-stratégie « *by identification strategy* » doit permettre de définir les objectifs et les contraintes en terme de circularité. Les participants sont invités à discuter de leurs objectifs et de leurs contraintes dans le cadre d'une séance de focus-group. Les étapes de ce focus-group sont résumées avec le protocole défini dans le tableau 2.

Dans notre exemple, les nouvelles activités en boucle ouverte qui existent dans le cas idéal de chaîne logistique nécessitent une collaboration avec de nouveaux acteurs. Par conséquent, le manque d'acteurs pourrait être une contrainte. L'objectif de circularité à atteindre est défini en tenant compte de ces contraintes et des écarts avec le cas idéal, il peut donc être inférieur au cas idéal selon les contraintes et les indicateurs de circularité choisis. Les contraintes et objectifs identifiés pour le cas Textile sont présentés dans la Fig. 6.

Table 2. Protocole de la stratégie « by identification strategy »

Protocole		<Analyze Gap between Best Available SC and As-IS SC, Define Circular Goal and Constraint, by identification strategy>		
Type de Session		Focus Group		
Participants		Gestionnaire et acteurs de la chaîne logistique		
N°	Etape	Description	Rôle des participants	Supports
1	Rappel des écarts avec le cas idéal (10 min)	L'animateur rappelle les principaux écarts entre la chaîne logistique idéale et la chaîne logistique « <i>As-Is</i> »	Passif	Modèle et indicateurs de circularité de la chaîne logistique idéale et de la chaîne logistique « <i>As-Is</i> »
2	Identification des contraintes (30 min)	L'animateur initie une discussion sur les contraintes liées à l'amélioration de la chaîne logistique : <ul style="list-style-type: none"> • Contraintes économiques ? • Besoin d'un nouvel acteur ? Si oui, quelles sont les contraintes liées à cet acteur ? • Capacités liées au flux de matières ? • Contraintes du marché ? 	Actif/ Collectif	Post-its avec les différents types de contraintes
3	Détermination des objectifs (10 min)	L'animateur initie une discussion sur les objectifs de circularité à atteindre, en considérant les contraintes et le cas idéal de chaîne logistique : <ul style="list-style-type: none"> • Quels objectifs souhaitez-vous atteindre pour vous rapprocher du cas idéal ? 	Actif/ Collectif	Post-its
4	Synthèse (10 min)	L'animateur fait une synthèse des objectifs et des contraintes et établit les priorités avec les participants.	Actif/ Collectif	-

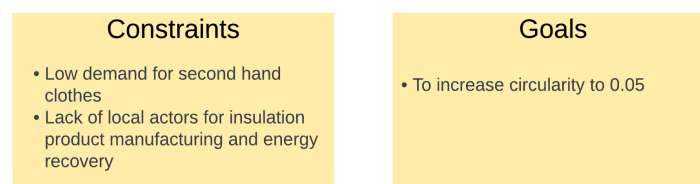


Fig. 6. Exemples de buts et contraintes identifiés pour le cas Textile

5.3. Stratégie d'évolution

La stratégie d'évolution « *by evolution strategy* » vise à imaginer une ou plusieurs chaînes logistiques « *As-If* », c'est-à-dire des chaînes logistiques plus circulaires. Elle se compose de deux intentions : « *Identifier les changements* » et « *Caractériser les changements opérationnels* ». La stratégie d'évolution commence par des stratégies de simulation de jeu ou de simulation informatique pour identifier les changements qui aideraient à atteindre les objectifs de circularité précédemment déterminés. Les changements proposés sont simulés à l'aide du jeu sérieux CircuSChain Game proposé dans des travaux antérieurs (Kurt et al., 2022) ou d'un outil de simulation industriel pour prédire l'évolution des indicateurs sélectionnés. Dans l'exemple du domaine Textile, nous supposons que les participants identifient les changements en tenant compte du cas idéal de chaîne logistique et des objectifs et des contraintes identifiés à la figure 6, et proposent la chaîne logistique « *As-If* » telle que représentée en figure 7. Dans ce modèle, à cause des contraintes spécifiées, l'activité de réutilisation en boucle fermée et l'intégration de matériaux d'isolation n'ont pas pu être envisagées. Malgré tout, la circularité est calculée à 0,053, ce qui est mieux que la circularité de la chaîne logistique « *As-Is* » calculée à 0,03 lors de la stratégie d'analyse.

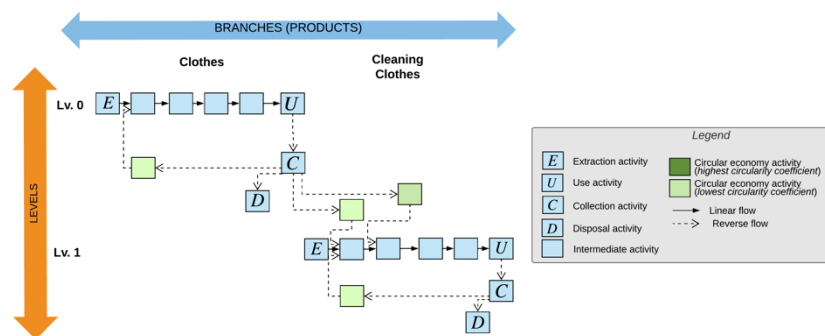


Fig. 7. Un exemple de chaîne logistique « *As-If* » pour le domaine Textile

6. Discussion

La méthode CircuSChain est une méthode participative qui vise à aider les responsables Supply Chain à transformer une chaîne logistique linéaire en chaîne logistique circulaire, ou à améliorer la circularité d'une chaîne logistique de manière continue. Elle orchestre différents outils proposés dans des travaux précédents : 1) dans la stratégie d'analyse, un modèle générique est utilisé pour analyser la circularité de la chaîne logistique, 2) dans la stratégie de diagnostic, des indicateurs existants sont réutilisés et de nouvelles dimensions de circularité et leurs indicateurs correspondants sont introduits pour diagnostiquer la chaîne logistique par rapport au cas idéal de chaîne logistique ; 3) dans la stratégie d'évolution, le jeu sérieux CircuSChain Game aide les acteurs à imaginer des scénarios possibles de chaînes

logistiques plus circulaires. Tous ces outils ont été validés indépendamment par des publications ou une validation utilisateur. Par exemple, le jeu sérieux CircuSChain Game (Kurt et al., 2022) a été évalué lors d'expérimentations pour observer la valeur éducative, le rythme, la valeur ludique et la simplicité du jeu. De plus, la méthode CircuSChain est une adaptation du framework As-Is/As-If (Cela et al. 2019) qui avait déjà été utilisé pour proposer des méthodes d'évolution continue dans deux types de systèmes différents : les écosystèmes (méthode ADInnov (Cortes-Cornax et al., 2016)) et les processus métiers (méthode CEFOP (Cela et al., 2019)). Ces différents cas d'utilisation du framework As-Is/As-If nous permettent de légitimer le fait que la méthode CircuSChain propose des techniques participatives et ludiques pour impliquer les acteurs dans l'évolution continue d'un système, ici une chaîne logistique circulaire. Néanmoins, la méthode CircuSChain en est à sa première version et nécessite une utilisation dans des cas industriels réels de chaînes logistiques pour être améliorée et validée.

Du point de vue de l'ingénierie des méthodes, l'objectif de l'adaptation du framework As-Is/As-If à la méthode CircuSChain présentée dans cet article était d'évaluer davantage la généricité et l'extensibilité du framework et de le renforcer en le confrontant à un nouveau type de système évolutif : les chaînes logistiques circulaires. En effet, l'adaptation du framework au domaine des chaînes logistiques circulaires nous amène à une évolution importante du framework pour prendre en compte la notion de « cas idéal » dans la stratégie de diagnostic, notion apparue comme très importante pour diagnostiquer la circularité de la chaîne logistique étudiée. Le concept de cas idéal de chaîne logistique a été ajouté au méta-modèle de produit de la méthode CircuSChain, et les intentions correspondantes ont été ajoutées au modèle de processus. Le framework As-Is-As-If a ainsi prouvé son extensibilité, nous permettant d'ajouter de nouvelles classes et associations dans le méta-modèle de produit et de nouvelles stratégies dans le modèle de processus. Cependant, du point de vue de l'Ingénierie des méthodes, nous devons encore approfondir pour déterminer si cette notion de cas idéal est pertinente pour d'autres domaines et doit être inclus dans le framework As-Is/As-If lui-même ou s'il ne correspond qu'au domaine des chaînes logistiques circulaires.

7. Conclusion

L'économie circulaire apparaît aujourd'hui comme une solution potentielle aux problèmes environnementaux, et les chaînes logistiques jouent un rôle essentiel dans son adoption auprès des industriels. Une transition réussie vers des chaînes logistiques circulaires nécessite une mesure continue de l'amélioration de leur circularité. Dans cet article, nous proposons la méthode participative CircuSChain permettant l'évolution continue des chaînes logistiques pour les rendre plus circulaires. Cette méthode est une adaptation du framework As-Is/As-If, qui facilite la construction d'une méthode d'évolution continue, jouant le rôle d'un template sur lequel une méthode cible peut s'appuyer. La méthode CircuSChain intègre des outils précédemment proposés, tels qu'un modèle générique, un indicateur de circularité, un outil de classification des indicateurs et un jeu sérieux pour promouvoir les

chaînes logistiques circulaires. Elle aide les responsables des chaînes logistiques à transformer leurs chaînes logistiques linéaires en chaînes logistiques circulaires, puis à les faire évoluer vers davantage de circularité de manière continue.

Nous avons illustré l'utilisation de la méthode par un exemple du domaine Textile basé sur la littérature. Cependant, des exemples issus de différents secteurs et des cas industriels réels sont encore nécessaires pour évaluer et valider la méthode CircuSChain. De plus, du point de vue de l'ingénierie des méthodes, le framework As-Is/As-If a prouvé son adaptabilité pour un nouveau type de système évolutif : les chaînes logistiques. Néanmoins, cette adaptation nous conduit à nous interroger sur l'introduction de la nouvelle notion de « Cas idéal » dans le méta-modèle produit de la méthode. Des travaux futurs sont nécessaires pour explorer l'utilité de cette notion dans d'autres domaines.

Remerciements

Cette recherche est soutenue par l'Agence Nationale de la Recherche dans le cadre du programme « Investissements d'avenir » (ANR-15-IDEX-02) à travers le Programme Transversal CIRCULAR et par la Région Auvergne Rhône Alpes dans le cadre du programme « R&D BOOSTER 2019 – Projet INROBOT ».

Bibliographie

- Ahi, P., Searcy, C.: A comparative literature analysis of definitions for green and sustainable supply chain management. *Journal of Cleaner Production*. 52, 329–341 (2013). <https://doi.org/10.1016/j.jclepro.2013.02.018>
- Blomsma, F., Brennan, G.: The Emergence of Circular Economy: A New Framing Around Prolonging Resource Productivity. *Journal of Industrial Ecology*. 21, 603–614 (2017). <https://doi.org/10.1111/jiec.12603>
- Cela, O.: A General Framework for the Continual Evolution Methods ; Adaptation to the Continual Evolution of Organization' s Business Processes PhD Thesis. HAL Id : tel-03346021, (2021)
- Cela, O., Front, A., Rieu, D.: CEFOP: A method for the Continual Evolution of Organizational Processes. *Proceedings - International Conference on Research Challenges in Information Science*. 33–43 (2017). <https://doi.org/10.1109/RCIS.2017.7956515>
- Cela, O., Cortes-Cornax, M., Front, A., Rieu, D.: Methodological Framework to Guide the Development of Continual Evolution Methods. In: *International Conference on Advanced Information Systems Engineering*. pp. 48–63 (2019)
- Cortes-Cornax, M., Front, A., Rieu, D., Verdier, C.: ADInnov: an Intentional Method to Instil Innovation in Socio-Technical Ecosystems. In: *International Conference on Advanced Information Systems Engineering* (2016)
- Deeb, S., Haouzi, H.B. El, Aubry, A., Dassisti, M.: A generic framework to support the implementation of six sigma approach in SMEs. *IFAC-PapersOnLine*. 51, 921–926 (2018). <https://doi.org/10.1016/j.ifacol.2018.08.490>
- Deming, W.E.: *Out of the Crisis*, reissue. MIT press (2018)

- Ellen MacArthur Foundation: Towards the Circular Economy. (2013)
- Ellen MacArthur Foundation: Towards the Circular Economy Vol.3: Accelerating the scale-up across global supply chains. (2014)
- Farooque, M., Zhang, A., Thürer, M., Qu, T., Huisingh, D.: Circular supply chain management: A definition and structured literature review. *Journal of Cleaner Production*. 228, 882–900 (2019). <https://doi.org/10.1016/j.jclepro.2019.04.303>
- Geissdoerfer, M., Savaget, P., Bocken, N.M.P., Hultink, E.J.: The Circular Economy – A new sustainability paradigm? *Journal of Cleaner Production*. 143, 757–768 (2017). <https://doi.org/https://doi.org/10.1016/j.jclepro.2016.12.048>
- Genovese, A., Acquaye, A.A., Figueroa, A., Koh, S.C.L.: Sustainable supply chain management and the transition towards a circular economy: Evidence and some applications. *Omega*. 66, 344–357 (2017). <https://doi.org/https://doi.org/10.1016/j.omega.2015.05.015>
- Jain, S., Jain, N.K., Metri, B.: Strategic framework towards measuring a circular supply chain management Sourabh. *Benchmarking: An International Journal*. 25, 3238–3252 (2018). <https://doi.org/DOI 10.1108/BIJ-11-2017-0304>
- Kurt, A.: Models and Tools for the Design , Assessment , and Evolution of Circular Supply Chains, Thèse de doctorat en Génie Industriel de l'Université Grenoble Alpes (2021) (NNT : 2021GRALI112). (tel-03615489)
- Kurt, A., Cortes-Cornax, M., Cung, V.-D., Front, A., Mangione, F.: A Classification Tool for Circular Supply Chain Indicators. In: *Advances in Production Management Systems*. pp. 644–653. Springer (2021)
- Kurt, A., Cortes-Cornax, M., Cung, V.-D., Mangione, F., Kaddouri, S.: CircuSChain Game: A Serious Game to Explore Circular Supply Chains. In: *Handbook of Research on Promoting Economic and Social Development Through Serious Games*. pp. 395–419. IGI Global (2022)
- Pavel, C.C., Blagoeva, D.T.: Competitive landscape of the EU's insulation materials industry for energy-efficient buildings. (2018)
- Rolland, C., Prakash, N., Benjamen, A.: A multi-model view of process modelling. *Requirements engineering*. 4, 169–187 (1999)
- Shook, J.: *Managing to learn: using the A3 management process to solve problems, gain agreement, mentor and lead*. Lean Enterprise Institute (2008)
- Stahel, W.R.: The Product-Life Factor. In: Orr, S.G. (ed.) *An Inquiry Into the Nature of Sustainable Societies: The Role of the Private Sector*. NARC (1982)
- Stahel, W.R.: *The Preformance Economy*. Springer (2010)
- Teck-Yong Eng: The Influence of a Firm's Cross-Functional Orientation on Supply Chain Performance. *Journal of Supply Chain Management*. 41, 4–16 (2005). <https://doi.org/10.1111/j.1745-493X.2005.04104002.x>
- UN Global Compact Business for Social Responsibility: Supply Chain Sustainability: A Practical Guide for Continuous Improvement (Second Edition). (2015)
- Zarghami, A., Benbow, D.: Introduction to 8D problem solving. *The Journal for Quality and Participation*. 40, 23–28 (2017)

Vers un Profil Environnemental des Processus Métier

Mario Cortes-Cornax, Claudia Roncancio, and Raquel Oliveira

Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France
{prenom.nom}@univ-grenoble-alpes.fr

Abstract. Face aux enjeux environnementaux, la maîtrise des impacts à l'échelle des organisations est indispensable. Pour cela, l'analyse des activités et des ressources mobilisées au regard d'indicateurs environnementaux est clé. L'objectif de nos travaux est de contribuer à enrichir la gestion de processus avec des profils environnementaux. Cet article porte sur une phase d'analyse où nous identifions les composants des modèles de processus qu'il convient d'enrichir pour clarifier leurs "coûts" environnementaux et ainsi contribuer à une vision systémique. Ces informations permettront de soutenir des démarches d'éco-conception ou d'évolution de processus pour ajouter des alternatives plus éco-responsables.

Keywords: Processus métier · Impact environnemental · Analyse de cycle de vie · Systèmes d'information

1 Introduction et Objectifs

Les enjeux environnementaux, notamment l'urgence de contrôler le réchauffement planétaire inférieur à 1.5°C, nous amène à regarder nos modes de vie sous un angle plus responsable. Le plan d'action pour le climat de la Commission Européenne¹, montre le besoin de réduire les émissions de gaz à effet de serre d'au moins 55% d'ici à 2030. Tenant compte du caractère systémique et de l'ampleur des problèmes, il est important de maîtriser les impacts environnementaux à l'échelle individuelle, des organisations et globalement.

La gestion de processus métier (*Business Process Management*, BPM) traitant la modélisation, l'analyse et l'amélioration des processus, est principalement motivée par les performances économiques et temporelles. Cependant, les aspects environnementaux tels que l'empreinte CO2 ou la consommation d'eau, sont peu, ou pas pris en compte. Notre travail s'inscrit dans la thématique *Green BPM* [11] qui vise à fournir des concepts, des méthodes et des outils pour soutenir la modélisation, la mise en œuvre, l'exécution, la surveillance et le changement continu des processus, en tenant compte de leurs impacts environnementaux. Certains travaux précurseurs [5, 10, 12] prennent en compte principalement l'impact environnemental de la phase d'usage sans inclure le cycle de vie complet des ressources mobilisées. D'autres aspects comme le flux des données parcourant ou générés par le processus (ainsi que l'infrastructure

¹ <https://climate.ec.europa.eu/eu-action/european-green-deal>

nécessaire pour leur stockage et leur traitement), ou les possibles effets rebonds de la mise en place de certaines activités ne sont souvent pas considérés. Les impacts peuvent ainsi être sous-estimés.

Cet article introduit des recherches en cours sur l'amélioration des analyses des processus métier visant à clarifier leur impact environnemental. Cela cherche à aider les responsables à prendre des décisions éclairées dans la mise en place des processus, l'introduction d'un budget CO2 ou d'une limite de consommation électrique, ou bien la proposition d'alternatives en fonction d'un coût environnemental. En s'appuyant sur le langage BPMN (*Business Process Model and Notation*) [9], nous présentons une analyse des points à développer pour construire un profil environnemental des processus métier. Dans nos travaux futurs, sans chercher à étendre BPMN, nous souhaitons formaliser une approche qui facilite l'analyse des modèles selon des aspects environnementaux.

Dans la suite, notre analyse est présentée en section 2, les travaux connexes en section 3 et nos conclusions et perspectives en section 4.

2 Vers un Profil Environnemental des Processus Métiers

Nous nous intéressons à la construction d'un profil environnemental des processus qui réunira autant que possible les informations liées aux ressources et aux activités. Il s'agit de considérer l'impact de l'ensemble des phases du cycle de vie et non seulement la phase d'usage ou d'exécution des processus comme c'est souvent le cas. Nous cherchons à introduire des vues permettant de clarifier les impacts par activité, processus, inter-processus et contribuer à une vision systémique. Les processus, de par leur nature, constituent une base d'analyse intéressante et permettent des raisonnements au niveau des modèles et des instances. Il faut noter que renseigner de manière complète et fiable le profil environnemental peut être difficile, voire impossible par manque d'information. L'idée est d'introduire cette notion de manière incrémentale et dans une démarche d'amélioration continue. Nous proposons une gestion du profil en associant un contexte et des indicateurs de qualité tels que la précision, la complétude, la validité et la confiance. Les indicateurs de qualité concernent en premier lieu les mesures des impacts et devront être agrégées pour les différents niveaux (par exemple, tâche, puis sous-processus, puis processus).

2.1 Expliciter l'Implicite lors de la Modélisation du Processus

BPMN v2.0 [9] fait figure de standard pour la modélisation de processus métier et est implémenté par la plupart de systèmes de gestion de processus. Lorsqu'on modélise un processus, notamment pour communiquer et s'abstraire de la complexité, certaines informations restent implicites. Même si peu importantes au niveau organisationnel, elles peuvent "cacher" des actions qui ont un impact environnemental qui mérite d'être pris en compte (c.f. fig. 1). Dans la suite, nous analysons des composants des modèles de processus qui peuvent cacher des impacts environnementaux, dans le but de construire le profil environnemental.

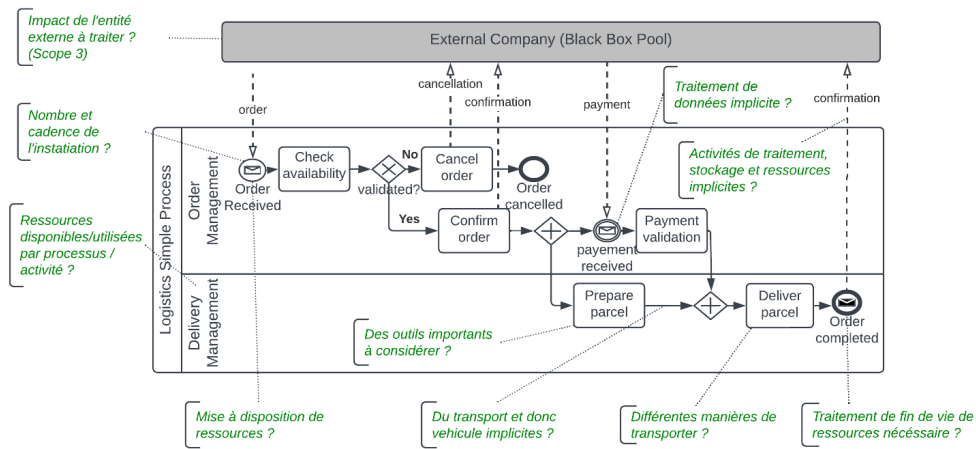


Fig. 1. Exemple de modèle BPMN et analyse de l'impact environnemental

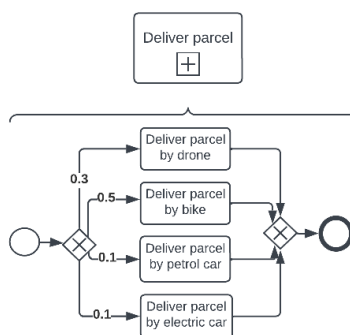


Fig. 2. Sous-processus avec probabilités de branchement

Les Activités peuvent être atomiques (tâches) ou composites (sous-processus). Les tâches atomiques, selon la logique métier, peuvent avoir une implémentation complexe. Le profil environnemental d'une activité est composé d'autant d'entrées que de possibilités d'exécution et de choix de ressources. Par exemple², en fig. 1, la tâche *Deliver parcel* correspond à la livraison d'un colis déjà payé. Cette tâche au niveau métier pourrait être transformée en sous-processus (voir fig. 2) pour indiquer le type de ressources utilisées ainsi que la probabilité moyenne d'exécuter un chemin lors de l'instanciation. Une entrée du profil se réfère à l'impact environnemental d'une exécution dans un contexte

Les flux de séquence (*Sequence Flows*) définissent les enchainements d'activités et en général n'ont pas de signification métier. Cependant, ils peuvent avoir un impact environnemental sous-jacent. Cela devra être explicité le cas échéant. La fig. 1 illustre ce point sur le couloir (*Lane*) de *Delivery Management* avec un possible transport implicite.

Un modèle de processus peut capturer plusieurs formes selon notamment les choix dans les branchements (*gateways*) et l'instanciation des activités. Des informations sur l'impact environnemental de chaque alternative feront partie du profil du processus. Des méta-données sur la qualité et le contexte de mesure

² Cet exemple, figures 1 et 2, sont inspirés de [3].

seront associées. Tenant compte qu'une vision globale est importante, une estimation de la répartition des alternatives selon les choix de branchement peut être utilisée pour estimer l'impact des *instances*. La fig. 2 montre les probabilités pour chaque branchement ce qui est important pour prédire le nombre de ressources nécessaires. Des informations sur l'ordonnancement des instances (e.g. parallèle ou séquentielle) contribuent aussi à l'estimation des ressources mobilisées.

Multi-processus: l'analyse globale des impacts amène à considérer les processus reliés. Cela introduit des points d'analyse supplémentaires : i) les flux de messages (*Message Flows*) représentent la communication entre processus et peuvent correspondre à des phases ayant un impact environnemental à inclure dans le profil ; ii) les processus externes en boîtes noires (*black box Pool*) représentent généralement des activités ou processus délégués en sous-traitance.

2.2 L'Impact des Ressources dans leur Cycle de Vie

Afin de rendre compte des impacts, il est nécessaire de disposer d'un modèle de ressources enrichi. L'étude des ressources (notamment pas humaines) est clé, car leur mise à disposition, utilisation, maintenance et fin de vie ont un impact qui peut être mis en relation avec les processus. L'inventaire des ressources doit être explicite ainsi que les informations de partage et d'allocation pour chaque activité et processus. Nous identifions les aspects suivants à considérer : i) introduire dans le profil environnemental des indicateurs de toutes les phases du cycle de vie des ressources ; ii) prendre en compte et expliciter la durée de vie utile estimée des ressources selon les patrons d'utilisation connus ; iii) étudier et expliciter l'impact des ressources lorsqu'elles sont inutilisées ou sous-utilisées ; iv) expliciter les ressources mobilisées pour des calculs et la gestion de données et leur impact environnemental ; v) d'introduire un *profil des données* et considérer les données comme des ressources à part entière. En effet, nous proposons d'explicitier leur cycle de vie de manière à clarifier les étapes, la volumétrie et les coûts associés. Le cycle de vie inclut notamment la collecte de données, la préparation, le stockage, stratégie de suppression ou archivage ; vi) renseigner au mieux les indicateurs environnementaux, gaz à effet de serre (en équivalent CO₂), mais aussi, la consommation d'eau, la pollution de l'eau, etc.

Des approches d'analyse de cycle de vie (telles que [4]) proposent des analyses multidimensionnelles des impacts des ressources. Diverses bases de données émergent sur des ressources de divers domaines dont l'informatique. Pour associer ces informations à la gestion des processus, nous travaillons sur une évolution dans des modèles de ressources et l'intégration de vues permettant de rendre compte des impacts, en incluant notamment ceux liés à la mise à disposition des ressources et aux traitements de fin de vie ou recyclage.

2.3 Contexte et Qualité

Les processus modélisés peuvent donner lieu à plusieurs instances qui peuvent potentiellement s'exécuter dans des contextes différents et utiliser des ressources

différentes. Leur impact environnemental peut ainsi différer. Une des ressources essentielles est l'électricité dont l'empreinte CO2 varie selon les endroits et les moments. Le profil environnemental d'un processus peut ainsi être très différent selon la période de l'année. Le maintien d'information spatio-temporelle est essentiel dans le cadre de l'analyse de l'impact environnemental.

Associer des informations de contexte aux indicateurs environnementaux aiderait à préciser leur pertinence et cohérence d'ensemble. En plus de cela, nous proposons d'associer des indicateurs de qualité tels que la précision, la complétude, la validité et la confiance. Ces indicateurs de qualité concernent en premier lieu les données des impacts par rapport à un certain contexte et plus largement, les profils environnementaux.

3 Travaux Connexes

Ce travail s'inscrit dans le courant *Green BPM* [11] où la durabilité est associée à la gestion de processus métier. En 2020, [2,6], revues systématiques du domaine, pointent un besoin de développement de la thématique, incitant la proposition d'indicateurs environnementaux pertinents pour atteindre un modèle de maturité environnemental des organisations. Certains travaux tels que [5, 10, 12] s'intéressent principalement aux émissions de CO2 dans les processus en suivant les *Scopes* du protocole GHG³: *Scope 1*, émissions des activités propres, *Scope 2*, celles provenant des sources d'énergie et *Scope 3*, émissions indirectes non contrôlées par l'organisation. Selon [8], les émissions du *Scope 3* représentent en moyenne 84 % des émissions totales des organisations. Associé à des méthodes *Activity Based Costing (ABC)* [1] et des approches top-down et bottom-up, il est possible de répartir les émissions dans les différentes tâches des processus [10,12]. Recket et al. [10] présente une extension de BPMN afin d'explicitier l'empreinte CO2 des processus, ainsi qu'une méthode d'analyse, appelé *Activity Based Emission (ABE)*, inspirée de la méthode ABC. Ils se focalisent sur les scopes 1 et 2 excluant les phases de fabrication, transport et fin de vie.

L'Analyse de Cycle de Vie (ACV, ISO 14040 [7]) est une approche normalisée pour quantifier les impacts environnementaux potentiels d'un produit ou d'un service depuis l'extraction des matières premières jusqu'à leur fin de vie. Il existe des bases de données d'ACV⁴ et des *Product Environmental Profile* concernant les appareils électroniques. À notre connaissance, le lien entre les systèmes de gestion de processus et les bases de données ACV, reste à développer.

Certains produits récents, tels que Climatiq et PlanA⁵, aident à une meilleure gestion des émissions carbone des entreprises. Ils ne proposent pas une approche intégrée avec des processus métiers. L'outil Voyance⁶ propose une analyse simplifiée des émissions de CO2 qui intègre l'approche processus avec BPMN et des efforts d'optimisation [3]. Cependant, les aspects ACV ne sont pas encore considérés et pour l'instant seulement la dimension CO2 est prise en compte.

³ <https://ghgprotocol.org/>

⁴ <https://eplca.jrc.ec.europa.eu/ResourceDirectory>

⁵ <https://www.climatiq.io/> et <https://plana.earth/>

⁶ <https://voyance.systems/>

Ce travail en cours cherche à enrichir l'existant et à contribuer à des démarches d'éco-conception et d'amélioration des processus par l'information des impacts environnementaux en considérant plusieurs critères et les cycles de vie complets.

4 Conclusion et Travaux Futurs

Par ce travail, qui est à ces débuts, nous cherchons à étendre la gestion de processus pour plus de conscience environnementale. Nous avons présenté une analyse où nous identifions les composants des modèles de processus qu'il convient d'enrichir, ainsi que les dimensions relatives aux impacts environnementaux à prendre en compte. L'explicitation des "coûts" environnementaux et des aspects les plus déterminants aide à aller vers plus d'éco-conception et peut motiver des évolutions vers des alternatives plus éco-responsables, jouer sur les stratégies d'exécution des processus et aboutir à l'association de notions telles qu'un budget environnemental. Cette démarche cherche également à contribuer à la traçabilité et l'explication des impacts environnementaux.

Il reste un travail conséquent à faire pour aboutir à une solution opérationnelle qui associe des profils environnementaux à la gestion de processus. Par la suite, nous travaillerons sur la formalisation de la proposition et explorerons des alternatives de mise en œuvre en considérant l'extension d'un outil existant, l'ajout d'un *Domain Specific Language* ou l'utilisation d'une approche *low code*.

References

1. Bromwich, M., Hong, C.: Activity-based costing systems and incremental costs. *Management accounting research* **10**(1), 39–60 (1999)
2. Couckuyt, D., Van Looy, A.: A systematic review of green business process management. *Business Process Management Journal* **26**(2), 421–446 (2020)
3. Durán, F., Rocha, C., Salaün, G.: Analysis of the runtime resource provisioning of BPMN processes using maude. In: *WRLA 2020*, 2020. pp. 38–56. Springer (2020)
4. Finnveden, G., Hauschild, M.Z., Ekvall, T., et al.: Recent developments in life cycle assessment. *Journal of environmental management* **91**(1), 1–21 (2009)
5. Ghose, A., Hoesch-Klohe, K., Hinsche, et al.: Green business process management: A research agenda. *Australasian Journal of Information Systems* **16**(2) (2010)
6. Gohar, S.R., Indulska, M., et al.: Environmental sustainability through green business process management. *Australasian Journal of Information Systems* **24** (2020)
7. ISO: Iso 14040:2006, environmental management — life cycle assessment — principles and framework (2006)
8. Makower, J., et al.: State of green business 2022. GreenBiz group (2012)
9. OMG: Business Process Model and Notation (BPMN), Version 2.0 (January 2011), <http://www.omg.org/spec/BPMN/2.0>
10. Recker, J., Rosemann, M., Hjalmarsson, A., Lind, M.: Modeling and analyzing the carbon footprint of business processes. In: *Green business process management: Towards the sustainable enterprise*, pp. 93–109. Springer (2012)
11. Seidel, S., Recker, J., vom Brocke, J.: Green business process management. *Green business process management: Towards the sustainable enterprise* pp. 3–13 (2012)
12. Wesumperuma, A., Ginige, A., Ginige, A., Hol, A.: Green activity based management (ABM) for organisations. In: *ACIS 2013 Proceedings*. 144. ACIS (2013)

Faciliter la migration vers une architecture microservices

Approche basée sur la rétroingénierie dirigée par les modèles et l'apprentissage par renforcement.

MohammadHadi Dehghani¹, Shekoufeh Kolahdouz-Rahimi¹, Massimo Tisi², Dalila Tamzalit³

1. *Isfahan University of Technology, Iran*
mohammadhadi.dehghani@jku.at, sh.rahimi@eng.ui.ac.ir,

2. *IMT-Atlantique, CNRS, LS2N, F-44000 Nantes, France*
massimo.tisi@imt-atlantique.fr,

3. *Nantes Université, CNRS, LS2N, F-44000 Nantes, France*
dalila.tamzalit@univ-nantes.fr

RESUME. Le présent document synthétise l'article accepté dans *Software and Systems Modeling*, Dehghani, M., Kolahdouz-Rahimi, S., Tisi, M., & Tamzalit, D. (2022). *Facilitating the migration to the microservice architecture via model-driven reverse engineering and reinforcement learning*, et présenté à *Models 2022* dans *First-Journal Track*.

Mots-clés : Architecture de microservices – Apprentissage par renforcement – Rétroingénierie dirigée par les modèles – Migration.

1. Introduction et problématique

Les microservices (MS) sont un style architectural de développement d'applications (Fowler, 2016). De par leurs caractères distribués et faiblement couplés, ils permettent une grande souplesse dans le développement logiciel, les évolutions et le déploiement. Les microservices permettent donc aux entreprises une grande réactivité face aux besoins changeants, augmentant leur compétitivité. Pour cette raison, plusieurs organisations entreprennent le coûteux processus de migration de leurs architectures logicielles traditionnelles vers des microservices (MS). Cependant, la migration est souvent faite de manière manuelle, notamment de par la complexité à déterminer la bonne granularité des MS et la difficulté à identifier l'affectation des fonctions métier au sein des MS (Gouigoux et Tamzalit. 2017).

2. Proposition

L'approche proposée offre une assistance semi-automatisée aux architectes logiciels pour faire face à la migration vers des MS. Elle suit trois étapes clés :

1. Extraire du code source existant son modèle orienté-objet grâce à MoDisco, un outil de rétroingénierie dirigé par les modèles (Bruneliere *et al.* 2014).
2. Extraire le modèle à base de MS en analysant les modèles entité-association et de cas d'utilisation du système existant avec Service Cutter (Gysel. *et al.* 2016).
3. **Contribution** : automatiser l'affectation des méthodes existantes de l'étape (1) aux MS identifiés à l'étape (2). Cette affectation est un problème complexe et multifactoriel. L'approche est basée sur le Deep Q-Learning par renforcement (Reinforcement Learning), une technique d'apprentissage d'IA. Elle génère des épisodes d'entraînement et d'apprentissage cherchant la meilleure correspondance possible entre les méthodes et les MS, basée sur des réseaux de neurones. L'applicabilité et la correction de l'approche développée ont été évaluées sur des cas d'utilisation et en exécutant des tests de performance.

3. Conclusion et perspectives

La solution développée est basée sur une technique d'IA pour approcher la meilleure correspondance possible entre les méthodes et les MS, afin de remodulariser les systèmes logiciels en vue d'une migration vers une architecture MS. Dans le futur, nous souhaiterions améliorer l'identification de la granularité des microservices en prenant mieux en compte la dimension métier, notamment avec la notion de contextes bornés (*bounded context*) du Domain-Driven Design (Evans 2003). Un contexte borné spécifie un périmètre métier cohérent et unifié pour chaque service. Cette notion est fondamentale pour identifier la bonne granularité d'un microservice car elle représente un sous-système aligné sur une partie du domaine métier.

Principales références

- Bruneliere H, Cabot, J., Dupé, G., Madiot (2014), F.: Modisco: A model driven reverse engineering framework. *Information and Software Technology* 56(8), 1012–1032.
- Evans (2003) Domain-Driven Design: Tackling Complexity in the Heart of Software. *Addison-Wesley*
- Fowler (2016), Production-ready microservices: Building standardized systems across an engineering organization. *O'Reilly Media, Inc.* "
- Gouigoux et Tamzalit (2017) From monolith to microservices: Lessons learned on an industrial migration to a web oriented architecture. *IEEE International Conference on Software Architecture Workshops*
- Gysel, Kölbener, Giersche, Zimmermann (2016): Service cutter: A systematic approach to service decomposition. *European Conference on Service-Oriented and Cloud Computing*.
- Sutton et Barto (2018) A.: Reinforcement learning: an introduction. *The MIT Press*.

Méthodologie de transition vers l'économie circulaire à l'aide d'un système d'information responsable

Louise Noël¹, Christophe Ponsard²

1. Soreal, Bruxelles, Belgique

louise.noel@soreal.be

2. CETIC - Centre de recherche, Gosselies, Belgique

christophe.ponsard@cetic.be

RÉSUMÉ. De nos jours, nous faisons face à la nécessité d'une meilleure gestion des ressources limitées de notre planète. Ceci exige la transition d'une économie linéaire à une économie circulaire à l'aide d'une utilisation responsable de systèmes d'information. Cet article propose un cadre d'analyse d'une chaîne de valeur circulaire et de l'apport net de solutions technologiques existantes ou émergentes. La démarche est illustrée et discutée sur un cas d'étude issu du secteur de la construction.

ABSTRACT. Nowadays, we are facing the need to improve the use of our limited earth resources. This requires the transition from a linear to a circular economy with the wise use of information systems. This paper proposes a framework for the analysis of a circular value chain in order to mobilise existing or emerging technological solutions. Our approach is illustrated and discussed on a case study from the construction sector.

MOTS-CLÉS: Économie circulaire, modèle de processus, transformation numérique, canevas

KEYWORDS: Circular Economy, Process modelling, Digital Transformation, canvas

1. Introduction

Depuis la révolution industrielle du XIX^{ème} siècle, le monde suit essentiellement une économie de type linéaire où en fin de vie, les produits sont immobilisés ou jetés. Ce modèle pose évidemment dans un problème de soutenabilité dans notre monde aux ressources finies et alimenté par la croissance.

Le terme « économie circulaire » (EC) émerge en 1990 avec l'idée de cycle de rétention maximale de la valeur d'un produit comme illustré à la figure 1. L'ADEME la définit comme: « un système économique d'échange et de production qui, à tous

les stades du cycle de vie des produits (biens et services), vise à augmenter l'efficacité de l'utilisation des ressources et à diminuer l'impact sur l'environnement tout en permettant le bien-être des individus ». Le développement de l'EC est progressif avec l'émergence de la notion de « Cradle-to-Cradle » (C2C) vers 2002 et les actions de la fondation Ellen MacArthur (Fondation Ellen MacArthur, 2009).

Actuellement, notre monde connaît une autre transformation liée au développement exponentiel des technologies de l'information et de la communication. Cette Transformation Numérique (TN) provoque un changement de paradigme des activités économiques et sociales (Mergel *et al.*, 2019). Dans le contexte de l'EC, la TN fait partie, d'une part *du problème*, car les systèmes d'information (SI) mobilisent de multiples ressources: production des équipements, conception des logiciels, consommation d'énergie avec émission de gaz à effet de serre, production de déchets électroniques,... D'autre, elle peut aussi contribuer à *sa solution*, par son soutien au changement vers l'EC en améliorant la gestion des ressources et en réduisant la production de déchets, tout en sachant que toute transformation mobilisera de l'énergie.

Notre objectif est de proposer une méthodologie soutenant la transition vers l'EC en s'appuyant de manière raisonnée sur les SI. Plus spécifiquement, nous tentons de répondre à la question de recherche « Comment les technologies numériques peuvent-elles soutenir le partage d'informations afin d'accélérer la transition circulaire ? ». A cette fin, nous proposons une méthodologie, elle-même circulaire, étendant des outils issus de l'analyse métier tels que le business canevas et les chaînes de valeur. Le travail est validé sur un cas d'étude du domaine de la construction.

La structure est la suivante. La section 2 synthétise un état de l'art des modèles d'EC. La section 3 décrit le domaine d'étude. La section 4 présente, illustre et discute notre méthodologie. Enfin la section 5 conclut en mettant en évidence quelques barrières et possibilités de généralisation pour alimenter nos travaux futurs.

2. État de l'art

Cette section identifie les éléments utiles des modèles d'EC tels que les types de cycles et diverses stratégies et outils.

Fondamentalement, on distingue d'une part les *cycles biologiques* avec les éléments biodégradables, tels que coton, des denrées alimentaires, bois, etc. La stratégie de ces cycles est de restaurer les nutriments au sein de la biosphère en reconstituant du capital naturel, que cela soit naturellement ou via une intervention humaine. D'autre part, les *cycles techniques* concernent l'ensemble des éléments non-biodégradables, tels que métaux et certains plastiques. Il importe d'évaluer à quel(s) cycle(s) sont mobilisés par une entreprise/un écosystème.

Le diagramme « Butterfly » (en papillon) développé par (Fondation Ellen MacArthur, 2009) est illustré à la figure 1. Les cycles biologiques et techniques sont représentés respectivement à gauche et droite avec des boucles liées à des stratégies décrites plus loin. Ainsi, au niveau technique, les boucles les plus courtes et efficaces sont du

partage puis de la réparation, les plus longues sont le recyclage des matières premières, qui implique des coûts plus importants de collecte, tri et traitement.

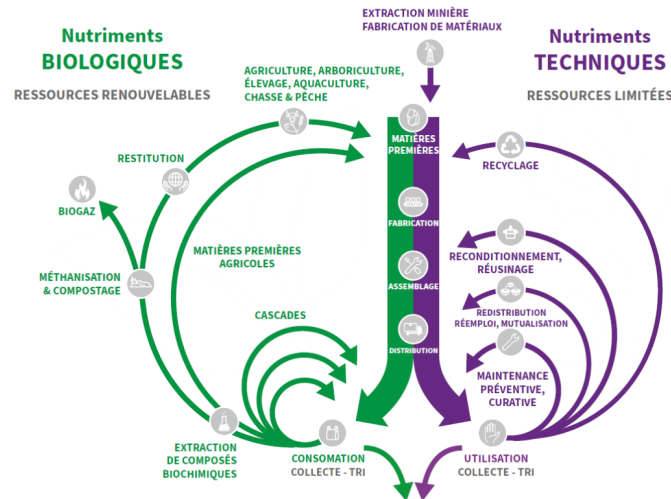


FIGURE 1. *Modèle « Butterfly »*

Diverses *stratégies de maintien/remise en circulation* alimentent l'EC, notamment les « stratégies en R » (cf. l'initiale « R »), avec notamment des approches de *Réduction* pour limiter la quantité de ressources nécessaires à la fabrication; de *Réutilisation* pendant la vie du produit pour la prolonger au maximum en réutilisant ou en réparant, ou encore de *Recyclage*, en fin de vie du produit. Ce concept peut s'étendre et se décliner sous d'autres formes notamment ReSOLVE (Régénérer, Partager, Optimiser, Boucler, Visualiser et Échanger) (MacArthur Foundation and McKinsey, 2015).

Pour mettre en place l'EC, des extensions à des outils d'analyse sont proposés, notamment un business model canevas et une chaîne de valeur circulaire développés par (Circulab, 2012). Ils permettent d'identifier des possibilités de prochain usage, les flux circulaires, ainsi que les impacts positifs et négatifs au niveau sociétal ou environnemental, y compris de possibles « effets de rebond ».

3. Cas d'étude: la construction

L'industrie de la construction serait responsable de plus de 30% de l'extraction des ressources naturelles et de 25% des déchets solides mondiaux. La plupart des bâtiments existants ne peuvent pas être déconstruits facilement afin d'y récupérer les matériaux qui deviennent inutilisables à la fin de vie de ce bâtiment. Au sein de l'Union Européenne, 5 à 12% des émissions de gaz à effet de serre proviennent de ce domaine, ce qui en fait un secteur à haut potentiel de réduction des émissions, estimé à 80% (ARUP, 2016)

Pour ce secteur, une prise de conscience des besoins de circularité émerge mais sa mise en oeuvre est lente, même pour de nouveaux projets car les parties prenantes ne maîtrisent pas encore les concepts et sont confrontés à divers obstacles liés à la complexité de l'écosystème ou de la réglementation. Il est cependant possible d'envisager des « stratégies en R » plus efficaces en termes de réemploi des éléments des bâtiments à différents niveaux de finition en considérant les bâtiments comme des banques de matériaux tel que défini par le projet BAMB, « Building As Material Banks » (Bamb, 2016). Ceci nécessite cependant la mise en place de SI spécifiques qui se heurtent au retard du secteur en matière de TN. Le défi est donc d'activer des leviers sur les deux dimensions simultanément afin de faire progresser tout un secteur. Pour notre étude de cas, nous considérerons un écosystème belge décrit et caractérisé au point suivant.

4. Méthodologie illustrée sur la cas d'étude

La méthodologie est structurée en 4 étapes d'un cycle de montée en maturité:

1. situer le niveau de maturité du domaine en question, à la fois en EC et TN;
2. construire la chaîne de valeur circulaire et/ou du business model circulaire;
3. analyser les besoins/barrières et les stratégies pour lever ces derniers;
4. identifier les technologies numériques utilisables pour la mise en oeuvre.

Étape 1 - Situer la maturité tant au niveau circulaire que numérique peut s'appuyer sur des indicateurs publiés. Dans notre cas, nous avons combiné la maturité numérique fournie par le baromètre digital wallon (Wallonia, 2022) et de maturité de l'économie circulaire selon l'index CEBIX (Circular Economy Business Index) (García-Sánchez *et al.*, 2021). Il combine 17 indicateurs couvrant des dimensions d'efficacité des processus, de réduction des émissions, de mise en oeuvre de boucles, de gestion des déchets, de recherche/innovation. Les deux indicateurs sélectionnés nous ont permis de situer la maturité du secteur de la construction à la figure 2.

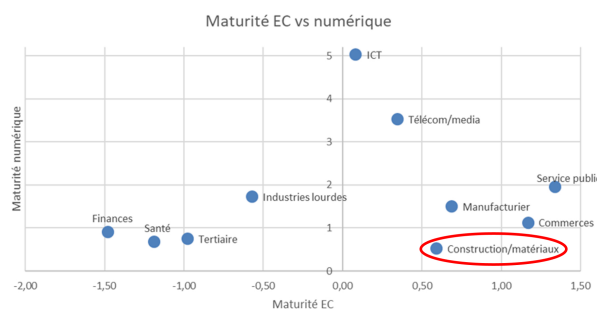


FIGURE 2. Maturité en EC vs TN

Étape 2 - L'élaboration de la chaîne de valeur circulaire permet d'identifier les acteurs actuels/émergents, ainsi que leurs manières actuelles de fonctionner en écosystème. Dans notre cas, ceci nous a permis de capturer la complexité des flux et d'identifier des acteurs sources d'innovation: Rotor DC, une société spécialisée dans

le démantèlement et de revente de matériaux, et Multipick réalisant le tri automatisé très rapide et fiable de déchets basé sur des techniques d'intelligence artificielle.

Etape 3 - La définition de la stratégie se base sur l'analyse des besoins et barrières issus de la chaîne de valeur établie au point précédent. Un point clef pour le succès de la transition circulaire est de gérer le nombre d'acteurs et d'interactions qui peut être important dans certains écosystèmes. Alors que dans le cadre d'une économie linéaire, les informations sont généralement peu partagées et suivent une chaîne logistique bien déterminée, dans un cadre circulaire, le partage doit être plus ouvert et dynamique. Ceci exige aussi un système de partage d'information plus adapté en termes d'ouverture, de décentralisation et de pérennité. La figure 3 illustre la stratégie envisagée qui simplifie les flux à travers tout le cycle de vie d'un bâtiment via un SI dédié à la conservation d'information sur le bâti.

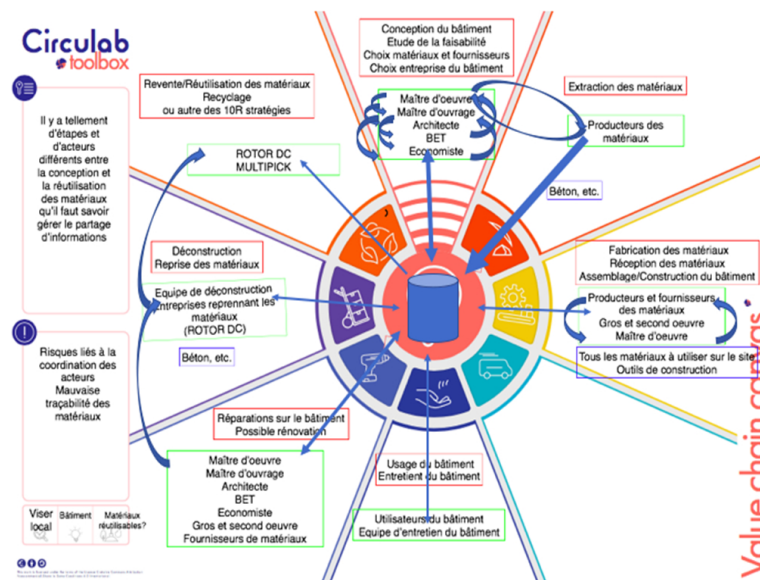


FIGURE 3. Flux structuré par un SI de partage d'information

Etape 4 - L'identification de l'apport des SI permet de proposer les technologies numériques les plus adéquates. Diverses technologies (communication, analytics, cyber physical systems, stockage de données, simulation,...) sont envisageables en fonction de leur capacité à soutenir certaines stratégies en R (Demestichas, Daskalakis, 2020). Dans notre cas, des solutions décentralisées de type Blockchain sont envisageables (Turk, Klinc, 2017) mais des obstacles de maturité et de coût énergétique font privilégier une solution plus centralisée gérée par un acteur neutre jouant le rôle de régulateur. Outre l'impact énergétique et environnemental, d'autres aspects sont aussi à étudier tels que l'adoption par les acteurs et la cybersécurité des SI déployés. A plus long terme, le changement de paradigme vers le « bâtiment comme service » est aussi à prendre en compte, avec comme facteur déclencheur l'adoption du télétravail et du flex office suite à la crise sanitaire du COVID.

5. Conclusion et perspectives

Cet article a esquissé une méthodologie itérative structurée permettant de soutenir la transition vers l'économie circulaire en s'appuyant sur une démarche de transformation numérique et basée sur l'utilisation d'outils d'évaluation de maturité ainsi que des canevas et chaînes de valeur adaptés pour un cadre circulaire. Nous avons illustré la démarche sur un cas issu de la construction, grand consommateur de ressources et en pleine phase de transformation numérique.

Bien que ciblée et limitée à une itération, notre démarche est généralisable pour un accompagnement plus long dans la transition et à travers d'autres domaines. Des travaux complémentaires seraient à envisager dans ce cadre. D'abord, il faudrait mettre en place une mesure de la montée en maturité ainsi qu'une feuille de route long terme pour la co-conduite de la transition circulaire et de la transformation numérique. Ensuite, les outils d'analyse, essentiellement papier, pourraient eux-mêmes être digitalisés et affinés, en considérant potentiellement des modèles plus élaborés, notamment d'architecture d'entreprises et d'accompagnement au changement. Enfin, des projets pilotes devraient être menés dans d'autres secteurs afin de collecter différents retours d'expérience pour alimenter des recommandations.

Bibliographie

- ARUP. (2016). *The circular economy in the build environment*. <https://www.arup.com/perspectives/publications/research/section/circular-economy-in-the-built-environment>.
- Bamb. (2016). *Materials passports*. <https://www.bamb2020.eu/topics/materials-passports>.
- Circulab. (2012). *La toolbox circulab - des outils pour rendre votre entreprise plus circulaire*. <https://circulab.com/fr/toolbox-circular-economy>.
- Demestichas K., Daskalakis E. (2020). Information and communication technology solutions for the circular economy. *Sustainability*, vol. 12, n° 18, p. 7272.
- Fondation Ellen MacArthur. (2009). *Circular economy introduction*. <https://ellenmacarthurfoundation.org/topics/circular-economy-introduction/overview>.
- García-Sánchez I.-M., Somohano-Rodríguez F.-M., Amor-Esteban V., Frías-Aceituno J.-V. (2021). Which region and which sector leads the circular economy? cebix, a multivariant index based on business actions. *Journal of Environ. Management*, vol. 297, p. 113299.
- MacArthur Foundation and McKinsey. (2015). *Growth within: a circular economy vision for a competitive europe*.
- Mergel I., Edelman N., Haug N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly*, vol. 36, n° 4.
- Turk Ž., Klinc R. (2017). Potentials of blockchain technology for construction management. *Procedia engineering*, vol. 196, p. 638–645.
- Wallonia D. (2022). *Baromètre 2022 de maturité numérique des entreprises wallonnes*. <https://www.digitalwallonia.be/fr/publications/entreprises2022/>.

Vers une optimisation de la diffusion de l'information dans une ville intelligente

Malika Grim-Yefsah¹, Mohamed Chachoua², Lea Jeantet¹

1. Laboratoire en Sciences et Technologies de l'Information Géographique, Université Gustave Eiffel, ENSG, IGN, LASTIG
6-8 avenue Blaise Pascal, F-77454 Marne-la-Vallée, France
malika.grim-yefsah@ensg.eu, lea.jeantet@ensg.eu

2. Laboratoire en Sciences et Technologies de l'Information Géographique, Université Gustave Eiffel, EIVP, LASTIG
80 rue Rébeval, F-75019 Paris, France
chachoua@eivp-paris.fr

RESUME. Les panneaux d'affichage numérique introduisent Internet dans les espaces publics et simplifient la diffusion de l'information dans les environnements urbains. Avec l'avènement de la ville intelligente, en plus de la diffusion publicitaire, ces panneaux offrent des opportunités pour la gestion de l'information en temps réel et pour développer des applications innovantes. Cet article explore la problématique du positionnement de ces panneaux dans l'espace urbain et propose une approche basée sur les représentations spatiales pour déterminer les points d'installation optimaux, leur permettant d'assurer de nouvelles fonctions, selon le lieu.

ABSTRACT. Digital billboards are bringing the internet experience into public spaces and stepping up the pace with which information penetrates urban environments. With the advent of the smart city, emerging opportunities in several domains have appeared, as such environmental monitoring and real-time information management. Thus, this article explores an optimisation for a strategic deployment of digital billboards. So an approach based on spatial representations to determine the optimal installation points, allowing them to perform new functions, depending on the location is proposed.

MOTS-CLES : Ville intelligente, Panneaux d'affichage numérique, Point d'intérêts, SIG.

KEYWORDS: Smart city, Digital out-of-Home media, POI, GIS.

1. Introduction

De plus en plus de villes utilisent l'affichage légal dématérialisé en installant des bornes interactives ou des écrans tactiles par application d'une nouvelle loi¹. Dans le

¹ Article L2131-1 du code général des collectivités territoriales modifié en août 2015 et du décret n° 2021-1311 du 7 octobre 2021.

cadre de cette loi, plusieurs villes souhaitent remplacer les panneaux d'affichage extérieurs "classiques" par des panneaux numériques à des fins économique, écologique, et pour la rapidité de diffusion de l'information.

Le contexte des JO 2024 vient précipiter ce déploiement. L'objectif spécifique est de répondre aux différents besoins tels que l'affichage des adresses des infrastructures sportives, des créneaux horaires des compétitions, des sites touristiques, des points de rassemblement en cas de mouvements de foule, etc. D'une manière générale, l'affichage numérique urbain offre une solution intéressante pour la diffusion des informations en temps réel, d'une manière dynamique et avec un service de gestion et d'assistance à distance. D'ailleurs, Vangelov (2022) nous indique que l'affichage extérieur numérique est la communication qui pourrait ajouter un nouveau sens aux zones urbaines tout en rapprochant les diffuseurs d'information et la population, en augmentant l'interaction. Cependant, le problème est de trouver les meilleurs emplacements pour ces panneaux. Huang et al (2021) pose le problème en terme spatio-temporel ; *les annonceurs doivent décider de l'espace et du temps de l'affichage.*

Nos recherches se concentrent sur une nouvelle approche pour le déploiement optimal des panneaux numériques, en répondant aux questions suivantes :

(1) *comment identifier les lieux d'installation des panneaux pour bien couvrir l'espace urbain ?* (2) *comment rendre les panneaux multifonctions ?*

Notre objectif est de proposer une stratégie optimale de déploiement de ces panneaux dans l'environnement urbain, en affectant à chaque panneau plusieurs fonctions selon le lieu et les besoins du moment.

Cet article est organisé comme suit. La section 2 offre un aperçu des méthodes utilisées dans l'affichage urbain. La section 3 présente notre approche du déploiement des panneaux numériques et son application à notre étude de cas, JO'2024. Nous concluons et présentons nos perspectives à ce travail en section 4.

2. État de l'art : Diffusion de l'information à travers l'affichage

Les panneaux d'affichage sont largement répandus dans l'espace urbain. Deman (2015) a identifié huit types de panneaux différents, (1) les *dispositifs muraux*, (2) les *dispositifs scellés au sol*, (3) les *enseignes temporaires*, (4) le *meublement d'affichage dit libre*, (5) les *vitres pour affichage mural*, (6) les *dispositifs éclairés*, (7) *meublement urbain (abris bus)* et (8) les *panneaux d'affichage numérique*. L'affichage des sept premiers types nécessite des déplacements sur les sites, alors que l'affichage numérique se fait à distance et en temps réel.

Par ailleurs, Roux et al. (2013) ont identifié trois formes principales de publicité, (1) *la publicité sur les moyens de transport en commun*, (2) *la publicité sur le meublement urbain* et (3) *les canaux médiatiques numériques*. Ils ont également noté que la publicité numérique est particulièrement efficace pour atteindre des publics spécifiques dans différents environnements, en dehors du domicile. Ces panneaux numériques sont aussi appelés « Digital Out of Home Media » (DOOH). Huang et al

(2021) ont proposé trois modèles pour le déploiement de DOOH. Le modèle «*Audience Matching*» qui utilise les données de téléphonie mobile et d'internet pour maximiser la couverture du public cible. Le modèle «*Environment Matching*» prend en compte l'environnement géographique (les Point of Interest -POI- dans une région), les conditions météorologiques, etc. Le modèle «*Integrated Matching*» combine les deux modèles précédents. Dans plusieurs pays, dont la France, l'usage des données des téléphones portables soulève des problèmes de confidentialité et souvent ces données sont inaccessibles, ainsi il y serait difficile d'implémenter les deux modèles «*Audience Matching*» et «*Integrated Matching*». Dans le cadre législatif actuel, seul le «*Environment Matching*» est possible en France.

3. Une nouvelle approche pour le déploiement des panneaux numériques

Les panneaux numériques répondent aux besoins de la ville intelligente et à la communication institutionnelle. Le principal défi est de trouver les meilleurs emplacements pour ces panneaux, ce qui nécessite de penser à la fois aux lieux et à leur fréquentation. Dans le cadre de notre étude de cas (JO'2024), nous proposons une approche pour un déploiement optimal des panneaux dans l'environnement urbain. Cette approche est basée sur le modèle "*Environment Matching*".

3.1. Installation selon l'importance des réseaux de transports

Les JO'2024 attireront des touristes et des habitants d'Île-de-France qui se déplaceront en grande majorité en transports en commun. Nous avons identifié les lignes desservant les lieux des épreuves, dans lesquelles des installations de panneaux sont envisageables. Nous avons utilisé le logiciel QGIS, un Système d'Information Géographique open source, pour sélectionner tous les arrêts de transport situés dans un rayon de 1 km des lieux des épreuves (Cf. Figure 1).

Nous avons analysé la manière dont les arrêts de transports desservent les sites olympiques recensés. Dans une première étape, nous avons simplement compté le nombre d'apparitions des lignes de transport sur l'ensemble de ces sites. Ensuite, nous avons associé des poids aux stations en fonction de leur importance dans un plan desservant les sites olympiques. Les poids de premier plan (P1) sont attribués aux lignes les plus apparentes, avec une valeur de 1 pour la ligne la plus importante et des valeurs inférieures pour les autres lignes en fonction de leur nombre d'apparitions. Nous avons appliqué le même principe pour les poids de second plan (P2) en prenant en compte les correspondances avec les lignes de bus et de métro de notre liste. Les poids finaux (P) ont été calculés pour chaque type de transport, et les lignes ont été hiérarchisées en fonction de leur poids.

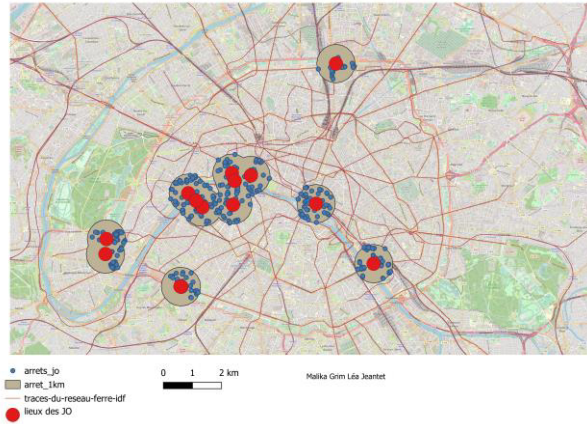


Figure 1. Résultat de l'analyse QGIS- Carte centralisant les lignes de transport en commun (hors bus) et stations de transports en commun au service des JO'2024

Les résultats sont présentés dans le Tableau 1 ci-après.

Tableau 1. Poids des lignes de transports

	Ligne	Poids	Type		Ligne	Poids	Type		Ligne	Poids	Type		Ligne	Poids	Type
2	C	1	RER	15	63	0.64	bus	28	75	0.43	bus	41	N	0.23	Train
3	72	1	bus	16	82	0.64	bus	29	11	0.42	metro	42	22	0.21	bus
4	6	0.97	metro	17	A	0.59	RER	30	13	0.41	metro	43	24	0.21	bus
5	69	0.93	bus	18	47	0.36	bus	31	7	0.4	metro	44	27	0.21	bus
6	1	0.90	metro	19	B	0.58	RER	32	10	0.4	metro	45	29	0.21	bus
7	9	0.85	metro	20	32	0.57	bus	33	21	0.36	bus	46	39	0.21	bus
8	42	0.97	bus	21	R	0.57	Train	34	30	0.36	bus	47	54	0.21	bus
9	8	0.78	metro	22	86	0.5	bus	35	92	0.36	bus	48	83	0.21	bus
10	12	0.78	metro	23	4	0.44	metro	36	93	0.36	bus	49	84	0.21	bus
11	14	0.73	metro	24	38	0.43	bus	37	73	0.36	bus	50	94	0.21	bus
12	D	0.73	RER	25	52	0.43	bus	38	77	0.29	bus	51	L	0.25	Train
13	80	0.71	bus	26	61	0.43	bus	39	89	0.29	bus				
14	28	0.64	bus	27	67	0.43	bus	40	2	0.26	metro				

Nous pouvons remarquer que presque tous les RER (A, B, C, D) sont représentés {(A, 0.59), (B, 0.58), (C, 1) et (D, 0.73)}, la ligne R (R, 0.57), ainsi que certaines lignes de métros (1, 6, 8, 9,12, 14). Il en résulte une première proposition : *Prévoir l'installation de panneaux numériques à proximité des lignes portant un haut score (supérieur à 0.50) ou à proximité d'un point regroupant un maximum de moyens de transport.* Notre proposition est en accord avec les travaux de Zhang et al. (2018).

3.2. Sélection selon les POI et axes de circulation

Pour cibler un public, il est important de zoner par quartier ou par voie, et de prendre en compte la présence de commerces. Nous avons donc calculé un poids pour chaque rue, voie ou boulevard de Paris, en considérant les poids de ses

commerces (cf. Formule 1). Nous avons ainsi postulé que les rues commerçantes attirent les touristes et les riverains. Ce calcul peut être amélioré en prenant en compte la fréquentation des enseignes, comme l'a suggéré (Roussel et al. 2013).

Nous avons utilisé les données ouvertes des commerces à Paris, fournies par l'APUR² via l'enquête "BDCOM 2020". Chaque arrondissement est présenté dans une fiche avec des chiffres-clés, des concentrations commerciales, les types de commerces et les évolutions depuis 2000. À partir de ces données, nous avons calculé pour chaque voie le nombre de commerces, noté NbT. Les scores par voie sont présentés dans le tableau 2.

$$Poids(enseigne) = \frac{\sum_{k=0}^n enseigne(k)}{Nbt} \quad (1)$$

Tableau 2 : Résultats de l'algorithme d'attribution des scores aux voies

VOIE	SCORE
BD MAGENTA	80.16
RUE BELLEVILLE	73.0
RUE La FAYETTE	72.4
RUE RIVOLI	69.0
RUE FAUBOURG SAINT DENIS	68.8
BD MONTPARTNASSE	66.96
RUE FAUBOURG SAINT MARTIN	66.4
RUE SAINT DENIS	66.4
BD VOLTAIRE	64.32
RUE SAINT HONORE	64.0
AV CLICHY	62.0

Nous avons remarqué que les voies identifiées par l'algorithme étaient des voies bien connues du grand public et souvent associées au tourisme, ce qui est cohérent avec notre objectif. Il en résulte une seconde proposition : *Prévoir l'installation de panneaux numériques à proximité des POI (commerces, sites touristiques, infrastructures sportives, ...) localisés sur des voies possédant un haut score.*

4. Conclusion et perspectives

Les panneaux numériques ont pris le dessus sur les panneaux classiques en raison de leur réactivité, de leur affichage rapide et de leur capacité à fournir des informations multiples en temps réel. Nous avons proposé une approche permettant d'optimiser l'emplacement de panneaux numériques pour bien couvrir l'espace urbain. Elle est construite sur le substrat des données ouvertes et de critères tels que la proximité des transports en commun et des points d'intérêt (POI). Elle a été étudiée dans le contexte des JO'2024. Les algorithmes que nous avons implémentés sont adaptables à d'autres contextes, en utilisant d'autres POI. Cependant, l'approche doit être améliorée en considérant la robustesse de l'attribution des poids. Il est

²Atelier Parisienne de l'Urbanisme : <https://www.apur.org/dataviz/commerces-paris>

également possible d'optimiser la diffusion d'information en fonction des heures d'ouverture des POI, il s'agit d'un travail en cours. Dans le contexte de notre étude de cas (JO'2024) sur les panneaux à proximité des infrastructures sportives, il est intéressant de paramétrer les horaires de diffusion de l'information en fonction de son importance, comme les finales d'athlétisme et de natation, qui se déroulent en soirée. Dans la continuité de ces travaux, nous envisageons d'inclure des fonctionnalités telles que la capture de bruit ou de pollution de l'air et leur affichage sur ces panneaux numériques.

Bibliographie

- Anagnostopoulos, A., Petroni, F., and Sorella, M., 2018. Targeted interest-driven advertising in cities using Twitter. *Data Mining and Knowledge Discovery*, 32 (3), 737–763.
- Deman L. (2015) Guide de l'affichage publicitaire. Rapport de l'Association agréée pour la protection de la Nature et de l'Environnement en Charente-Maritime
- Huang M., Fang Z., Weibel R., Zhang T., Huang H. (2021) Dynamic optimization models for displaying outdoor advertisement at the right time and place, *International Journal of Geographical Information Science*, 35:6, 1179-1204, DOI:10.1080/13658816.2020.1823396
- Kall F., Blankenbach, K. (2022), Flexible and Lightweight RGB LED Video Foil for Digital Signage. *Information Display*, 38: 23-27. <https://doi.org/10.1002/msid.1302>
- Lai J., Cheng, T., and Lansley, G., 2017. Improved targeted outdoor advertising based on geotagged social media data. *Annals of GIS*, 23 (4), 237–250.
- Rousell A., Zipf, A. (2017) Towards a Landmark-Based Pedestrian Navigation Service Using OSM Data. *ISPRS Int. J. Geo-Inf.* 2017, 6, 64. <https://doi.org/10.3390/ijgi6030064>
- Roux T., De la Rey van der Walddt, Ehlers L. (2013) A classification framework for out-of-home advertising media in South Africa, *Communicatio*, 39:3, 383-401, DOI: [10.1080/02500167.2013.837598](https://doi.org/10.1080/02500167.2013.837598)
- Stalder U. (2011). Digital Out-of-Home Media: Means and Effects of Digital in Public Space. In: Müller, J., Alt, F., Michelis, D. (eds) *Pervasive Advertising*. Human-Computer Interaction Series. Springer, London. https://doi.org/10.1007/978-0-85729-352-7_2
- Tucker C.E. (2014). Social networks, personalized advertising, and privacy controls. *Journal of Marketing Research*, 51 (5), 546–562.
- Vangelov N., “Digital Marketing and Outdoor Advertising in Smart Cities”, *SCRD*, vol. 6, no. 3, pp. 81–91, Jun. 2022. <https://doi.org/10.25019/scrd.v6i3.138>
- Wang, J. & Kwan, M.P. 2018, An analytical framework for integrating the spatiotemporal dynamics of environmental context. *International Journal of Environmental Research and Public Health*, 15, 9
- Wilson R.T. (2022) Out-of-Home Advertising: A Systematic Review and Research Agenda, *Journal of advertising*, DOI: [10.1080/00913367.2022.2064378](https://doi.org/10.1080/00913367.2022.2064378)
- Zhang, P., et al., 2018. Trajectory-driven influential billboard placement. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. London, United Kingdom: Association for Computing Machinery, 2748–2757.

Aide alimentaire : digital et dignité ?

Trois études de cas de systèmes d'information au service de la dignité des personnes

Pierre-Emmanuel Arduin ¹, Doudja Saïdi-Kabeche ²

1. Université Paris-Dauphine, PSL, DRM UMR CNRS 7088
Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France
pierre-emmanuel.arduin@dauphine.psl.eu
2. SADAPT, UMR 1048 Inrae AgroParisTech, Université Paris-Saclay
22 place de l'agronomie, CS 20040, 91123 Palaiseau Cedex, France
doudja.kabeche@agroparistech.fr

RÉSUMÉ. L'Organisation des Nations Unies pour l'alimentation et l'agriculture définit clairement la sécurité alimentaire comme une situation non seulement de disponibilité et d'accessibilité des denrées alimentaires, mais aussi d'acceptabilité sociale des conditions d'accès à l'alimentation. En France, la lutte contre la précarité alimentaire est l'objet de politiques publiques qui se traduisent majoritairement par la distribution d'aide alimentaire sous forme de denrées (colis alimentaires) par le biais d'associations caritatives. Cette aide alimentaire, telle qu'elle est organisée se heurte souvent à la question de la dignité des personnes. Dans cet article, nous proposons une étude monographique s'appuyant sur du matériel de première main (six années de terrain d'étudiants auprès d'associations) et de seconde main (analyse de rapports spécialisés, juridiques et d'activité). Nous décrivons des systèmes d'information inspirants au sein de trois associations françaises pour montrer concrètement comment ils peuvent renforcer la dignité dans l'aide alimentaire.

ABSTRACT. The Food and Agriculture Organization of the United Nations clearly defines food security as a situation not only of availability and accessibility of foodstuffs, but also of social acceptability of food access conditions. In France, the fight against food insecurity is the subject of public policies which mainly result in the distribution of food aid in the form of foodstuffs (food parcels) through charities. As currently organized, food aid often raises the question of the dignity of the recipients.

In this article, we present a monographic study based on first-hand material (six years of field-work by students with associations) and second-hand material (analysis of specialized, legal and activity reports). We describe inspiring information systems within three French associations to concretely show how they can enhance dignity in food aid.

MOTS-CLÉS : Aide alimentaire; dignité; logistique; système d'information.

KEYWORDS: Food aid; dignity; supply chain; information system.

1. Introduction

Alors que 607 millions de personnes étaient sous-alimentées en 2014, la pandémie mondiale a exacerbé la faim dans le monde jusqu'à 811 millions de personnes en 2020 (Nations-Unies, 2021b). Pourtant, le droit à l'alimentation est reconnu par le Haut-Commissariat des Nations Unies aux droits de l'Homme comme un droit fondamental car il est essentiel à la jouissance de tous les droits (Nations-Unies, 2021a). La sécurité alimentaire consiste à assurer les conditions pour que « toutes les personnes, à tout moment, aient un accès physique, social et économique à une alimentation suffisante, sûre et nutritive qui répond à leurs besoins et préférences alimentaires pour une vie active et saine » (Nations-Unies *et al.*, 2021). L'insécurité alimentaire survient lorsque certains individus ne sont pas en situation de sécurité alimentaire et, bien que contre-intuitif, cela est également vrai dans les pays développés, avec une acuité et des enjeux très différents.

Dans cet article, nous proposons d'étudier trois cas dans lesquels les systèmes d'information ont pu renforcer la dignité dans la logistique de l'aide alimentaire. Les systèmes ont permis aux organisations à but non lucratif de parvenir à une distribution de l'aide alimentaire plus acceptable socialement et plus digne.

Dans la deuxième section de cet article, nous présentons trois études de cas : (i) l'initiative de l'association ReVivre et son système d'aide à la décision AlimHotel, développé pour améliorer la situation alimentaire des personnes hébergées en hôtels, (ii) le futur Système d'Information Géographique (SIG) développé par la Fédération Française des Banques Alimentaires (FFBA) pour résoudre le problème des zones dites « blanches », c'est-à-dire non couvertes par un réseau d'aide alimentaire, et (iii) l'initiative de l'association HopHopFood qui a développé une application smartphone visant à lutter contre le gaspillage alimentaire des commerces au profit des populations défavorisées.

2. Trois études de cas d'associations d'aide alimentaire en France

Dans cette section, nous présentons trois études de cas d'associations françaises qui, au travers d'initiatives « digitales » c'est-à-dire du déploiement de systèmes d'information, ont développé des projets innovants de modernisation de l'organisation de l'aide alimentaire.

2.1. L'association ReVivre et un système d'aide à la décision

AlimHotel est un service développé par ReVivre, une organisation d'aide alimentaire ainsi qu'une entreprise sociale d'insertion professionnelle dans le domaine de la logistique, pour améliorer la situation alimentaire des personnes hébergées en hôtel par l'État en Île-de-France. Ce projet a été lancé par le SAMU Social et financé par la région Île-de-France pour améliorer la situation alimentaire de ces personnes.

Les hôtels sont souvent situés dans des banlieues isolées et n'offrent aucun accès à une cuisine pour préparer les repas. Le type de produits pouvant être proposés est par conséquent limité. Aussi, en plus de la proposition d'un service logistique pour les livraisons hebdomadaires de colis alimentaires, l'association a créé un système d'aide à la décision pour la préparation des colis et des menus hebdomadaires. Ce système a été créé par deux bénévoles, un nutritionniste et un informaticien (voir figure 1). Il aide à composer des colis-repas en prenant en compte :

- L'équilibre alimentaire et les recommandations du Programme National Nutrition Santé (PNNS).
- L'inventaire disponible (possibilité de substituer des produits pour respecter l'équilibre nutritionnel des colis proposés).
- La composition de la famille, l'aide alimentaire est ciblée par tranche d'âge (colis pour adultes, pour enfants et pour bébés).
- Les habitudes alimentaires, par exemple ne pas donner de porc aux familles musulmanes ni de viande aux végétariens.

Enfin, le système propose aux destinataires une rotation des denrées alimentaires : selon les familles, une rotation de quatre colis types est prévue pour éviter les repas récurrents plusieurs semaines de suite. Ce projet est considéré par les parties prenantes comme une véritable innovation sociale renforçant la dignité dans l'aide alimentaire.

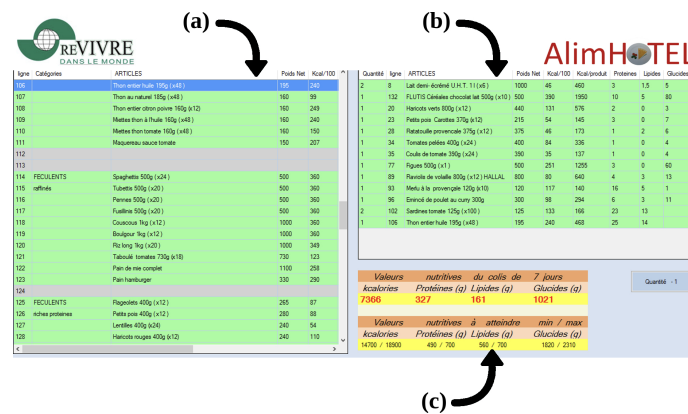


FIGURE 1. Le système d'aide à la décision AlimHotel supportant la préparation des colis : base d'aliments (a), sélection courante (b) et balance nutritionnelle (c)

2.2. La Fédération Française des Banques Alimentaires et un futur Système d'Information Géographique

La Fédération Française des Banques Alimentaires (FFBA) est une importante organisation d'entraide fondée en 1984 sur le modèle des Banques Alimentaires Américaines. C'est une fédération de 79 banques alimentaires régionales et de 31 filiales locales, créant un réseau qui couvre la totalité du territoire national français.

Comme beaucoup d'associations, la FFBA reconnaît que la couverture du territoire par l'aide alimentaire reste inégale entre les territoires urbains et ruraux. En effet, alors que les zones rurales représentent 80% du territoire national, l'offre d'aide alimentaire se concentre sur les territoires urbains et périurbains. Ainsi, en raison de contraintes logistiques évidentes, l'aide alimentaire souffre de l'existence de zones dites « blanches », c'est-à-dire des zones où les distributions et les réseaux d'aide classiques ne peuvent assumer la précarité et l'insécurité alimentaire. Ces zones sont des territoires ruraux isolés.

La FFBA a ainsi lancé un projet, toujours en cours, de créer un Système d'Information Géographique (SIG) pour comprendre et mieux répondre à la problématique de ces zones blanches. La première version de cette carte a été réalisée en s'appuyant sur le logiciel QGIS¹ (voir figure 2²). Il s'agit d'un SIG qui peut superposer plusieurs couches d'informations, après quoi la cartographie proposée permet de localiser les services d'aide alimentaire par rapport à diverses données comme le taux de pauvreté, le nombre d'allocataires sociaux, la densité de population, etc.

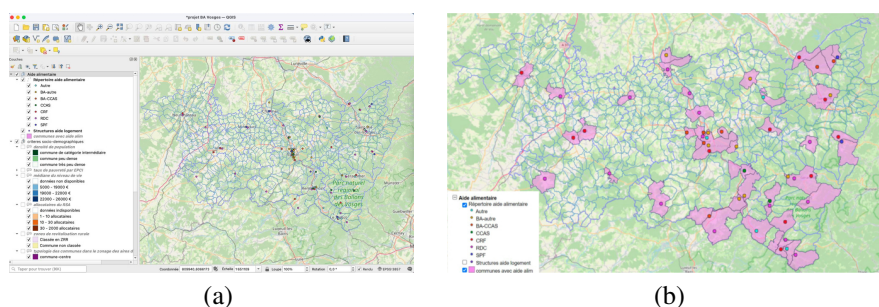


FIGURE 2. Un système d'information géographique pour agréger des informations (a) et montrer des inégalités de distribution (b)

Les fonctionnalités offertes par les SIG sont puissantes, même si elles ne résolvent pas le problème de l'accessibilité à l'aide alimentaire, elles permettent de le visualiser et d'aider à le caractériser pour apporter des solutions. Néanmoins, le potentiel de l'outil est évidemment subordonné à la qualité et à la disponibilité de l'information.

2.3. L'association HopHopFood et une application smartphone

HopHopFood est une association créée en 2018. L'objectif est de mettre en relation des personnes ayant besoin de nourriture avec des donateurs (particuliers ou commerçants) disposant de surplus de denrées. HopHopFood propose de lutter contre le gaspillage alimentaire au profit des populations défavorisées.

1. <https://www.qgis.org/>

2. source : <https://banquealim88.github.io/projetBA88/>, consulté le 28 avril 2023

HopHopFood cible les populations qui ne bénéficient pas, volontairement ou involontairement, du système d'aide alimentaire classique, notamment les étudiants et les migrants. HopHopFood les décrit comme des personnes vivant dans une « petite précarité », qui échappent souvent au radar des systèmes d'aide alimentaire traditionnels. L'objectif est de leur proposer une solution adaptée à leur profil. Cette population, souvent des jeunes technophiles équipés de smartphones, est estimée par HopHopFood à environ 10 millions de personnes. Parmi les solutions développées, nous nous sommes focalisés pour l'instant sur l'offre *business-to-consumer* qui permet aux vendeurs volontaires de proposer une partie de leurs invendus alimentaires. L'application pour smartphone fonctionne selon les mêmes principes que d'autres déjà existantes, à la différence près que HopHopFood est une association à but non lucratif.

Différents magasins (détaillants, restaurants, bouchers, boulangers, etc.) ayant des invendus alimentaires les mettent à disposition sur l'application. Les utilisateurs destinataires de l'aide alimentaire peuvent se connecter à l'application avec un code d'accès spécifique qui leur a été délivré par les travailleurs sociaux (Caisse d'Allocations Familiales notamment). Ainsi, un utilisateur disposant d'un code d'accès sélectionne les produits disponibles souhaités, les place dans son panier virtuel et se rend en magasin pour les retirer. HopHopFood précise que 70% des utilisateurs de ces codes sont des étudiants.

Les utilisateurs de ce service ne paient pas les colis qu'ils récupèrent. Les commerçants sont rémunérés par des déductions fiscales conformément aux lois françaises Garot et Egalim. Lorsqu'un utilisateur se rend chez un commerçant pour récupérer le colis qu'il a commandé, il présente son smartphone avec un justificatif de commande et le commerçant saisit son code spécifique. Cela permet ensuite au commerçant de déclarer les denrées alimentaires données, ce qui entraîne la création par la plateforme du formulaire administratif pour bénéficier de la déduction fiscale. En contrepartie, le commerçant s'engage à reverser à HopHopFood une partie de cette déduction fiscale, ce qui permet à l'association de continuer à se développer et de rester économiquement viable.

De plus, avec ce système, il n'y a pas de produits exclus du don (viande, fruits de mer, etc.). En effet, pour des raisons pratiques de sécurité alimentaire, les produits les plus chers ne peuvent pas être donnés à l'aide alimentaire classique. Dans le cas du service HopHopFood, ces règles restrictives ne s'appliquent pas car la chaîne d'approvisionnement vers les utilisateurs est similaire à celle des clients classiques. Ainsi, pour les utilisateurs, cette façon de récupérer les invendus est considérée comme beaucoup moins stigmatisante que de se rendre dans les centres de distribution d'aide alimentaire pour obtenir un colis alimentaire.

3. Conclusion et perspectives

L'aide alimentaire en France s'appuie sur le secteur associatif qui s'appuie lui-même sur le don de denrées alimentaires. Des auteurs comme Caillavet *et al.* (2021) ont mis en évidence certaines lacunes d'un tel système en soulignant comment les

filières d’approvisionnement de l’aide alimentaire deviennent un véritable casse-tête pour les associations d’aide alimentaire et pour le grand nombre de bénévoles mobilisés.

Les études de cas présentées dans cet article soulèvent de nombreuses questions liées à la durabilité du système d’aide alimentaire. Ce système a le mérite d’être une réponse adéquate à des formes spécifiques de vulnérabilité qui conduisent des individus ou des groupes d’individus à une situation d’insécurité alimentaire : par exemple, les personnes vivant dans les hôtels dans le cas de ReVivre, les personnes isolées dans les zones rurales dans le cas de la FFBA, et des personnes en « petite pauvreté » et *digital natives* dans le cas de HopHopFood. Les initiatives exposées, aussi intéressantes soient-elles, perpétuent la logique dominante dans l’approche de l’insécurité alimentaire, à savoir la distribution gratuite de nourriture partiellement issue du gaspillage alimentaire. Or, ce faisant, comme le soulignent Caillavet *et al.* (2021), on se place dans une logique curative pour réparer les défaillances d’un système, au lieu de privilégier une logique préventive pour éviter la survenue de l’insécurité alimentaire. De nouvelles propositions, évoquées notamment par Gundersen (2021) émergent et appellent à un revenu universel pour l’alimentation ou à une sécurité sociale de l’alimentation qui permettrait de s’attaquer aux causes profondes de l’insécurité alimentaire, du moins pour une partie de la population.

Remerciements

Les auteurs tiennent à remercier les étudiants et les associations d’aide alimentaire qui ont donné accès à leurs données de terrain et à leurs systèmes.

Bibliographie

- Caillavet F., Darmon N., Dubois C., Gomy C., Saïdi-Kabeche D., Paturol D. *et al.* (2021). *Vers une sécurité alimentaire durable: enjeux, initiatives et principes directeurs*. Rapport technique. Terra Nova, accessible à l’adresse <https://tnova.fr/societe/alimentation/vers-une-securite-alimentaire-durable-enjeux-initiatives-et-principes-directeurs/> (le 30 avril 2023).
- Gundersen C. (2021). A proposal to reconstruct the supplemental nutrition assistance program (snap) into a universal basic income program for food. *Food Policy*, vol. 101, p. 102096.
- Nations-Unies. (2021a). *The right to adequate food, human rights fact sheet series no. 34.* Rapport technique. United Nations Human Rights, accessible à l’adresse <https://www.ohchr.org/Documents/Publications/FactSheet34en.pdf> (le 30 avril 2023).
- Nations-Unies. (2021b). *The sustainable development goals report*. Rapport technique. Accessible à l’adresse <https://unstats.un.org/sdgs/report/2021/> (le 30 avril 2023).
- Nations-Unies, FAO, IFAD, UNICEF, WFP, WHO. (2021). *The state of food security and nutrition in the world 2021* (vol. 2021). Food & Agriculture Org, accessible à l’adresse <https://www.fao.org/3/cb4474en/cb4474en.pdf> (le 30 avril 2023).

Vers la conceptualisation du périmètre de la Responsabilité Numérique des Entreprises

Aiste Rugeviciute^{1,2}, Vincent Courboulay¹

¹L3i, La Rochelle Université

Technoforum | 23 avenue Albert Einstein | BP 33060 - 17031 La Rochelle, France

²Square Research Center, Square Management

173 Av. Achille Peretti | 92200 | Neuilly-sur-Seine | France

aiste.rugeviciute@univ-lr.fr

RESUME. Un nouveau concept de Responsabilité Numérique des Entreprises (RNE) a émergé pour répondre aux questions relatives aux performances sociales et environnementales d'une entreprise dans le cadre de ses activités numériques. Si un nombre de chercheurs discutent des motivations pour le déploiement de la RNE, son périmètre reste ambigu, ce qui complique son adoption. Sur la base des normes de responsabilité environnementale et sociale et en identifiant les enjeux numériques à partir de l'analyse de la littérature, ce projet propose de structurer le périmètre de la RNE. Dans cette optique, nous visons à contribuer aux débats sur la gouvernance des activités numériques au sein des organisations.

Mots-clés : Responsabilité Numérique des Entreprises, numérique responsable, ESRS

1. Introduction

Les technologies numériques présentent un fort potentiel pour contribuer au développement durable. Pourtant, elles s'accompagnent de coûts et de risques d'aggravation des problèmes environnementaux et sociaux. Compte tenu de la nature omniprésente des technologies numériques, les entreprises sont confrontées à un dilemme : comment prendre des décisions stratégiques responsables et tenir compte des effets économiques, environnementaux et sociaux du numérique ? En réponse à ces défis, un nouveau domaine, la Responsabilité Numérique des Entreprises (RNE), est apparu, faisant référence à la performance extra-financière des activités numériques des entreprises. Bien qu'il n'existe actuellement aucune définition universellement acceptée de la RNE, nous adhérons dans notre travail à celle proposée par Mihale-Wilson *et al.* (2021) qui affirment que la RNE vise à minimiser les effets négatifs du numérique tout en maximisant les impacts positifs des activités numériques des entreprises. Dans cette optique, la RNE cherche à garantir un développement, un déploiement et une utilisation éthiques et responsables du numérique. Même si la conceptualisation scientifique de la RNE en est encore à ses débuts, elle est perçue comme un sujet de plus en plus important pour tous les acteurs économiques dans la réalisation de leurs politiques de développement durable (Herden *et al.*, 2021; Lobschat *et al.*, 2021).

Si une littérature de plus en plus abondante présente les motivations pour la mise en place de la RNE, le périmètre de cette dernière reste encore ambigu (Carl *et al.*, 2022). L'absence de points de référence communs constitue un défi pour les entreprises qui souhaitent attribuer et mettre en place leurs stratégies RNE. Bien que

les chercheurs discutent des raisons pour lesquelles les organisations doivent implémenter la RNE (Lobschat *et al.*, 2021; Wirtz *et al.*, 2022), une littérature limitée a tenté de structurer et d'identifier les thèmes qui devraient être abordés dans ce cadre (Dörr, 2021; Herden *et al.*, 2021; Thorun *et al.*, 2017). Ces travaux restent à consolider afin de concrétiser le débat théorique et pratique sur la manière dont la RNE peut aider à relever les nouveaux défis numériques dans le contexte du développement durable (Elliott *et al.*, 2021).

Sur la base des propositions des normes de responsabilité environnementale et sociale (Environmental and Social Responsibility Standards - ESRS) et en identifiant les enjeux du numérique responsable à partir de l'analyse de la littérature, nous proposons de structurer le périmètre de la RNE. Il s'agit de la première étape de notre projet en cours qui vise à développer un modèle mathématique pour anticiper les impacts socio-écologiques d'une stratégie RNE, afin de promouvoir le développement, le déploiement et l'utilisation éthiques et responsables du numérique.

2. Contexte

Le terme de Responsabilité Numérique des Entreprises est relativement nouveau - l'une de ses premières occurrences se trouve dans le rapport d'Accenture de 2015 (Cooper *et al.*, 2015). Pourtant, les questions entourant les responsabilités et les risques numériques ne sont pas propres à la communauté RNE, mais une question à laquelle d'autres communautés de recherche sont confrontées depuis des décennies. Mueller (2022) affirme que les racines conceptuelles de la RNE se trouvent dans l'éthique des affaires, la Responsabilité Sociale/Sociétale des Entreprises (RSE) et l'éthique informatique. La reconnaissance du fait que les organisations ont la responsabilité d'atténuer les impacts négatifs des produits et services numériques qu'elles fabriquent, conçoivent, utilisent et déploient s'étend également à d'autres communautés de recherche en informatique. Les chercheurs des Systèmes d'Information (SI), surtout de « Green SI », ont proposé divers cadres et pratiques pour aider les organisations à élaborer des stratégies visant à atténuer les effets négatifs des produits et services numériques sur l'environnement (Deng *et al.*, 2017; Ereğ *et al.*, 2013; Molla, 2009). Les réflexions sur les risques et les opportunités du numérique et le rôle que les entreprises doivent jouer pour les aligner sur les stratégies de développement durable se retrouvent également parmi d'autres communautés de recherche plus spécialisées, comme l'IA éthique ou Sustainable Human-Computer Interaction (SHCI).

La pluridisciplinarité est donc au cœur de la Responsabilité Numérique des Entreprises. Elle se situe à l'intersection de la recherche sur l'éthique des affaires, le développement durable et les impacts des technologies numériques. Différentes définitions de la RNE ont été proposées mais elles partagent le même constat : la RNE doit jouer un rôle important dans la gouvernance et permettre aux entreprises de se conformer à une vision plus large du numérique responsable (Mueller, 2022). Même en l'absence de consensus sur ce que sont les responsabilités numériques ou sur les intérêts qu'elles doivent représenter, de plus en plus de chercheurs soulignent que les impacts du numérique sur l'environnement et la société doivent être pris en compte dans le cadre de la RNE (Dörr, 2021; Elliott *et al.*, 2021; Herden *et al.*, 2021; van der

Merwe et Al Achkar, 2022). Il y a donc en cela des similarités avec la RSE. En 2010, Harmon *et al.* (2010) ont affirmé que la RSE devrait être le moteur du numérique responsable, notamment parce qu'elle s'applique à l'impact des entreprises sur l'économie, l'environnement et la société dans son ensemble. La plupart des publications actuelles indiquent que la RNE et la RSE partagent des éléments similaires, mais que la RNE mérite une attention distincte en raison des défis uniques et des particularités d'un monde numérique (Carl *et al.*, 2022; Lobschat *et al.*, 2021; C. A. Mihale-Wilson *et al.*, 2021; van der Merwe et Al Achkar, 2022).

3. Définition du périmètre de la RNE

Compte tenu du fait que la RNE est perçue comme un sujet qui rejoint les thèmes de la RSE, nous proposons d'utiliser un outil RSE pour structurer son périmètre. Pour cela, nous avons décidé de choisir les nouvelles normes européennes « European Sustainability Reporting Standards » (ESRS) pour les raisons suivantes :

- Les normes ESRS tiennent déjà compte de la législation européenne (European Commission, n.d.) ainsi que d'autres instruments internationaux, tels que les Principes directeurs de l'OCDE à l'intention des entreprises multinationales. Dans l'UE, elles deviendront aussi obligatoires pour un large éventail d'entreprises ;
- L'utilisation de thèmes et de termes largement reconnus pourrait faciliter l'appropriation des concepts de la RNE au sein des entreprises et favoriser le dialogue entre les différentes parties prenantes ;
- Le positionnement de la RNE dans la lignée de l'ESRS pourrait permettre d'illustrer davantage le lien entre le numérique et le développement durable et faciliter l'alignement des stratégies d'informatique et de RSE.

Selon les propositions de l'ESRS, « un enjeu de développement durable est important du point de vue de l'impact lorsqu'il concerne les impacts matériels réels ou potentiels, positifs ou négatifs, de l'entreprise sur les personnes ou l'environnement à court, moyen et long terme » (EFRAG, 2022). Sur la base de cette définition, nous avons clarifié notre objectif : *identifier les enjeux du numérique qui concernent les impacts positifs ou négatifs sur les personnes ou l'environnement, résultant des activités numériques d'une entité.*

Notre analyse des premières propositions de l'ESRS (EFRAG, 2022) révèle que jusqu'à présent, les impacts numériques sont à peine abordés. Les impacts environnementaux des activités numériques peuvent être intégrés dans les thèmes classiques de la RSE, identifiés par l'ESRS comme *le climat, la pollution, les ressources en eau et les ressources marines, la biodiversité et les écosystèmes, et l'économie circulaire*. Cependant, de nombreux impacts sociaux issus de la transformation numérique sont de natures différentes et doivent encore être consolidés. Sur la base de l'analyse documentaire, nous avons tenté de compléter ce panorama et proposé une cartographie des enjeux numériques afin de définir les thèmes à aborder dans le périmètre de la RNE. Le tableau 1 présente les enjeux du numérique identifiés pour une dimension sociale/sociétale et précise le lien entre ces questions et les sujets existants de l'ESRS.

L'objectif de notre présentation du périmètre de la RNE ne prétend pas à l'exhaustivité, ni que les enjeux du numérique identifiés sont une représentation figée qui devraient être prise en compte dans toute stratégie RNE. La pertinence des thèmes de la RNE dépendra fortement du contexte. L'objectif est plutôt de fournir une

première synthèse des enjeux numériques actuellement considérés dans la littérature comme pertinents pour les entreprises et devant être traités dans le contexte plus large du développement durable. Le modèle est destiné à évoluer et à refléter les nouveaux défis et opportunités de la transition numérique pour les entreprises d'un point de vue socio-écologique.

Tableau 1. Périmètre RNE : dimension sociale/sociétale des enjeux numériques

Sujet ESRS	Sous-topique ESRS	Enjeux spécifiques du numérique
Effectif propre	Conditions de travail	<ul style="list-style-type: none"> • Bien-être numérique
	Égalité et de chances pour tous	<ul style="list-style-type: none"> • Compétences numériques • Gestion algorithmique
	Autres droits liés au travail	<ul style="list-style-type: none"> • Automatisation • Propriété des données et confidentialité
Travailleurs dans la chaîne de valeur	Conditions de travail	<ul style="list-style-type: none"> • Bien-être numérique
	Égalité et de chances pour tous	<ul style="list-style-type: none"> • Gestion algorithmique
	Autres droits liés au travail	<ul style="list-style-type: none"> • Propriété des données et confidentialité • Protection sociale des travailleurs indépendants du secteur numérique
Communautés concernées	Droits économiques, sociaux et culturels des communautés	<ul style="list-style-type: none"> • Biens communs numériques • Données ouvertes • Résilience et souveraineté numérique
	Droits civils et politiques des communautés	<ul style="list-style-type: none"> • Surveillance numérique • La cohésion sociale dans l'univers numérique • Liberté d'expression numérique
Consommateurs et utilisateurs finaux	Impacts liés à l'information pour les consommateurs et/ou les utilisateurs finaux	<ul style="list-style-type: none"> • Accès à l'information et autonomie • Transparence numérique • Propriété des données et confidentialité • Surveillance numérique
	Sécurité personnelle des consommateurs et/ou des utilisateurs finaux	<ul style="list-style-type: none"> • Bien-être numérique • Sécurité physique du produit • Sécurité des données et cyber-résilience
	Inclusion sociale des consommateurs et/ou des utilisateurs finaux	<ul style="list-style-type: none"> • Une IA impartiale et la non-discrimination • Accès aux services et infrastructures numériques • Compétences numériques

5. Conclusion et perspectives

Récemment, la communauté de recherche en SI a été appelée à intégrer et à étudier davantage l'impact des technologies numériques et leur signification dans la prise de décision des stratégies de développement durable (Lokuge *et al.*, 2021; Mann *et al.*, 2018; Santarius *et al.*, 2022). Le domaine de la RNE, nouvellement apparu, vise à répondre à ces questions. Cependant, l'adoption de la RNE dépend de la capacité à bien définir son périmètre, ainsi que comprendre sa relation avec la RSE classique et les avantages que les actions de la RNE peuvent apporter pour réduire les impacts négatifs et maximiser les impacts positifs des activités numériques des entreprises (C. Mihale-Wilson *et al.*, 2022). Cherchant à contribuer à l'avancement du débat scientifique sur la RNE, nous proposons une cartographie conceptuelle des thèmes de la RNE - les enjeux numériques, basée sur l'ESRS et enrichie d'observations tirées de la littérature existante. Dans cette optique, nous souhaitons contribuer à la clarification du champ d'application de la RNE, ce qui constitue un élément essentiel pour faciliter son application pratique et faire avancer la recherche académique.

En dépit de tous nos efforts, notre recherche a ses limites. Tout d'abord, une analyse de la littérature plus large pourrait fournir une liste plus exhaustive des thèmes RNE. En outre, l'interprétation des thèmes a inévitablement nécessité une certaine connaissance qualitative du contexte et n'a donc pas été exempte de jugement subjectif. Néanmoins, ce travail en cours peut servir de premier point de départ pour des pistes futures et d'inspiration pour la construction de cadres et de méthodologies pour l'adoption de la RNE et son évaluation.

Bibliographie

- Carl, K. V., Zilcher, T. M. C., & Hinz, O. (2022). Corporate Digital Responsibility and the current Corporate Social Responsibility standard: An analysis of applicability. In H. Roßnagel, C. Schunck, & S. Mödersheim (Eds.), *Open Identity Summit 2022* (pp. 75–86). Gesellschaft für Informatik e.V. https://doi.org/10.18420/OID2022_06
- Cooper, T., Siu, J., & Wei, K. (2015). *Corporate digital responsibility: Doing well by doing good*. Accenture. <https://www.criticaleye.com/inspiring/insights-servfile.cfm?id=4431>
- Deng, Q., Ji, S., & Wang, Y. (2017). Green IT practice disclosure: An examination of corporate sustainability reporting in IT sector. *J. Inf. Commun. Ethics Soc*, 15, 145–164. <https://doi.org/10.1108/JICES-12-2016-0046>
- Dörr, S. (2021). *Corporate Digital Responsibility. Managing Corporate Responsibility and Sustainability in the Digital Age* (1st ed.). Springer Berlin, Heidelberg.
- EFRAG. (2022). *First Set of draft ESRS - EFRAG*. <https://www.efrag.org/lab6>
- Elliott, K., Price, R., Shaw, P., Spiliotopoulos, T., Ng, M., Coopamootoo, K., & van Moorsel, A. (2021). Towards an Equitable Digital Society: Artificial Intelligence (AI) and Corporate Digital Responsibility (CDR). *Society*, 58(3), 179–188. <https://doi.org/10.1007/s12115-021-00594-8>
- Erek, K., Loeser, F., & Zarnekow, R. (2013). Reference Model for Sustainable Information Systems Management: Establishing a Holistic Research Agenda. *All Sprouts Content*, 511, 13.
- European Commission. (n.d.). *Corporate sustainability reporting*. https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en
- Harmon, R., Demirkan, H., Auseklis, N., & Reinoso, M. (2010). From Green Computing to Sustainable IT: Developing a Sustainable Service Orientation. *2010 43rd Hawaii*

- International Conference on System Sciences*, 1–10.
<https://doi.org/10.1109/HICSS.2010.214>
- Herden, C., Alliu, E., Cakici, A., Cormier, T., Deguelle, C., Gambhir, S., Griffiths, C., Gupta, S., Kamani, S., Kiratli, Y.-S., Kispataki, M., Lange, G., Matos, L., Moreno, L., Nunez, H., Pilla, V., Raj, B., Roe, J., Skoda, M., & Schons, L. (2021). “Corporate Digital Responsibility”: New corporate responsibilities in the digital age. *Sustainability Management Forum | NachhaltigkeitsManagementForum*, 29, 13–29. <https://doi.org/10.1007/s00550-020-00509-x>
- Lobschat, L., Mueller, B., Eggers, F., Brandimarte, L., Diefenbach, S., Kroschke, M., & Wirtz, J. (2021). Corporate digital responsibility. *Journal of Business Research*, 122, 875–888. <https://doi.org/10.1016/j.jbusres.2019.10.006>
- Lokuge, S., Sedera, D., Cooper, V., & Burstein, F. (2021). Digital Transformation: Environmental Friend or Foe? Panel Discussion at the Australasian Conference on Information Systems 2019. *Communications of the Association for Information Systems*, 48(1), 616–634. <https://doi.org/10.17705/1CAIS.04846>
- Mann, S., Bates, O., & Maher, R. (2018). Shifting the maturity needle of ICT for Sustainability. *5th International Conference on Information and Communication Technology for Sustainability*, 52, 209–226. <https://doi.org/10.29007/d6g3>
- Mihale-Wilson, C. A., Zibuschka, D. J., GmbH, R. B., Zibuschka, J., & Carl, K. V. (2021). Corporate Digital Responsibility – extended conceptualization and empirical assessment. *ECIS 2021 Research Papers*, 80, 17.
- Mihale-Wilson, C., Hinz, O., van der Aalst, W., & Weinhardt, C. (2022). Corporate Digital Responsibility: Relevance and Opportunities for Business and Information Systems Engineering. *Business & Information Systems Engineering*, 64(2), 127–132. <https://doi.org/10.1007/s12599-022-00746-y>
- Molla, A. (2009). Organizational motivations for green IT: developing and exploring a green IT matrix and motivation models. *Pacific Asia Conference on Information Systems*, 15.
- Mueller, B. (2022). Corporate Digital Responsibility. *Business & Information Systems Engineering*, 64, 689–700. <https://doi.org/10.1007/s12599-022-00760-0>
- Santarius, T., Bieser, J. C. T., Frick, V., Höjer, M., Gossen, M., Hilty, L. M., Kern, E., Pohl, J., Rohde, F., & Lange, S. (2022). Digital sufficiency: Conceptual considerations for ICTs on a finite planet. *Annals of Telecommunications*. <https://doi.org/10.1007/s12243-022-00914-x>
- Thorun, C., Vetter, M., Reisch, L., & Zimmer, A. K. (2017). *Indicators of consumer protection and empowerment in the digital world*. Institute for Consumer Policy.
- van der Merwe, J., & Al Achkar, Z. (2022). Data responsibility, corporate social responsibility, and corporate digital responsibility. *Data & Policy*, 4(e12). <https://doi.org/10.1017/dap.2022.2>
- Wirtz, J., Kunz, W., Hartley, N., & Tarbit, J. (2022). Corporate Digital Responsibility in Service Firms and Their Ecosystems. *Journal of Service Research*, 0(0). <https://doi.org/10.1177/10946705221130467>

OntoFiC : une ontologie pour la détection de fraude financière et la modélisation des comportements des clients

Lyliabrouk¹, Hamza Chergui^{1,2}, Hamid Ahaggach¹, Benjamin Auger¹, Dominique Cheron²

1. *Laboratoire d'Informatique de Bourgogne - EA 7534*

Université de Bourgogne

9, avenue Alain Savary, F-21078 Dijon - France

prénom.nom@u-bourgogne.fr

2. *SKAIZen Group*

hchergui,dcheron@skaizengroup.fr

RÉSUMÉ. La détection de fraude est un problème complexe pour les institutions financières. Dans cet article, nous présentons notre approche pour détecter les transactions frauduleuses dans le réseau SWIFT basée sur une ontologie du domaine. Nous présentons l'ontologie OntoFiC construite pour la modélisation des transactions et des acteurs SWIFT. Cette ontologie est peuplée avec un ensemble de données réelles. Nous avons développé une approche basée sur des règles associées à des scénarios de fraude.

ABSTRACT. Fraud detection is a complex issue for financial institutions. In this article, we present our approach to detect fraudulent transactions in SWIFT network based on the domain ontology. We present the OntoFiC ontology constructed for the modeling of SWIFT transactions and actors. This ontology is populated with a real dataset. We developed rules-based approach with rules associated to fraud scenarios.

MOTS-CLÉS : Ontologie, raisonnement, détection de fraude, finance

KEYWORDS: Ontology, reasoning, fraud detection, finance

1. Introduction

La détection de fraude est un problème complexe et constitue un véritable challenge pour les institutions financières. Les transactions frauduleuses ne sont pas fréquentes, mais les conséquences sur ces institutions peuvent aller du remboursement de la somme au client jusqu'à des amendes importantes dans les cas par exemple de blanchiment d'argent. Pour cela, les institutions financières doivent disposer d'outils

pour la prévention et détection de fraudes. L'entreprise SKAIZen Group développe un projet de recherche et d'innovation qui a pour but de modéliser les clients et les institutions financières en peuplant une base de connaissances à partir de différentes sources de données. Cette base permettra d'optimiser les moteurs de détection de fraudes dans les transactions financières. Le travail présenté dans cet article s'inscrit dans le cadre de notre collaboration avec SKAIZen Group dans le projet France relance (Auger *et al.*, 2022) qui a pour objectif la construction d'une base de connaissances alimentée par des données transactionnelles en prenant en compte les données financières par type (client, compte, institution financière) et par relation (bénéficiaire, débiteur, compte/client, etc.). Nous proposons dans ce travail une ontologie spécialisée sur les informations financières modélisant les informations des transactions bancaires dans le réseau SWIFT et les informations des acteurs : clients et institutions financières. Ce travail est basé sur le modèle SWIFT qui est devenu une norme ISO 20022¹ et notre base de connaissances clients KYC (Jabbari *et al.*, 2020). Cette ontologie est peuplée à partir de transactions bancaires de sources hétérogènes. La création d'une ontologie spécialisée sur les informations financières permet d'une part d'interroger la base de connaissances sur les relations entre les différentes entités (personnes ou organisations) et, d'autre part, sur la détection de fraudes à partir de règles.

La suite de l'article est organisée de la manière suivante : dans la section 2, nous dressons un état de l'art des techniques sémantiques de détection de fraudes financières basées sur les ontologies. Nous proposons notre approche dans la section 3. Afin de valider notre approche, nous avons réalisé des expérimentations à partir d'un jeu de données réel. Nous concluons notre travail dans la section 4 et nous donnons quelques perspectives.

2. État de l'art : Les techniques basées sur la sémantique

Les ontologies sont utilisées depuis plusieurs années pour la représentation des connaissances dans plusieurs domaines (commerce, santé, biologie, financier). Ces dernières années, plusieurs travaux ont utilisé les ontologies dans le domaine financier pour la représentation des connaissances et la détection des fraudes financières. L'ontologie permet d'une part la représentation du domaine et la possibilité de définir des règles afin d'analyser les transactions et ainsi de détecter des fraudes ou des comportements suspects. (Attigeri *et al.*, 2018) proposent une ontologie pour modéliser les fraudes bancaires. La méthode TF-IDF est utilisée pour trouver les phrases où l'expérience de fraude apparaît le plus. Le modèle Latent Dirichlet Allocation (LDA) est utilisé afin d'extraire les termes clés du domaine pour créer l'ontologie avec le thésaurus WordNet et définir les relations. L'ontologie est utilisée comme référence afin de détecter les nouvelles transactions frauduleuses. (El Orche *et al.*, 2018) proposent une approche pour prévenir et détecter des transactions frauduleuses dans des systèmes de paiement électroniques. L'approche est basée sur une ontologie contenant les acti-

1. <https://www.iso20022.org/iso-20022-message-definitions>

vités, les règles anti-fraudes, les risques et les activités transactionnelles. Les auteurs présentent le processus semi-automatique de détection de fraudes, mais n'expliquent pas comment l'ontologie est construite, ni comment le peuplement de l'ontologie est réalisé. (Ahmed *et al.*, 2021) proposent un algorithme de génération d'alertes basé sur des règles classées par sévérité. Dans ce travail, les auteurs créent l'ontologie *Financial Fraud Detection* (FFD), ils développent des règles de fraudes anti-blanchiment d'argent et un algorithme de génération d'alertes. (Hussaini *et al.*, 2022) développent une ontologie pour définir les fraudes financières avec l'objectif d'identifier des patterns pour la prévention et la détection de fraudes. Cette approche est composée de plusieurs étapes : la définition des types de fraudes, la modélisation du comportement des utilisateurs en se basant sur des ontologies existantes. La validation de cette approche est réalisée sur plusieurs jeux de données.

La détection de fraude ne doit pas dégrader la relation avec le client en bloquant les transactions de clients légitimes. Pour cela, le processus de détection de fraude doit également prendre en compte le client en intégrant son profil et ses usages. A notre connaissance, il n'existe pas de travaux sur les transactions du réseau SWIFT, et sur la modélisation des différents acteurs (clients ou agents). Le développement d'une approche basée sur des systèmes de règles avec une ontologie du domaine prenant en compte le profil client et les spécificités du réseau SWIFT pourrait aider les institutions financières à améliorer leur système de lutte contre la fraude.

3. Approche

Nous présentons dans cette section l'ontologie du domaine développée dans le cadre de nos travaux. Nous construisons dans un premier temps l'ontologie du domaine (concepts et propriétés). Cette dernière est peuplée à partir de transactions SWIFT. Nous définissons à l'aide d'experts des règles de fraudes basées sur les transactions et les clients. Cela nous permet d'étiqueter et de détecter des transactions frauduleuses de l'ontologie peuplée. Ensuite, nous exécutons des requêtes pour identifier les transactions frauduleuses et les présenter aux experts sous forme structurée à l'aide d'un outil de visualisation basé sur les graphes.

Nous avons sélectionné avec l'aide d'experts du domaine les informations les plus pertinentes pour la détection de fraude. Les informations concernant les clients ont été sélectionnées de la base de connaissances KYC. Dans l'ontologie proposée, les concepts présentés dans le tableau 1 sont divisés en trois concepts principaux : le **client**, l'institution financière qui est un **agent** débiteur, créateur ou intermédiaire et la **transaction**. Les propriétés de données et d'objet ont également été définis à l'aide des experts.

TABLEAU 1. *Concepts*

Name	Description
Activity	Le domaine d'activité
Agent	Institution financière (banque du client, banque de l'intermédiaire)
Agent:CreditorAgent	Institution financière attachée au client crédeur
Agent:DebtorAgent	Institution financière attachée au client débiteur
Agent:IntermediaryAgent	Institution financière intermédiaire entre deux institutions financières
Customer	Client crédeur et débiteur (n'est pas une institution financière)
Customer:Organization	Responsable de vente de biens et services
Customer:Organization:Association	Organisations, associations et ONG. Pas utilité à cause d'ambiguïté juridique.
Customer:Organisation:Company	Entreprise publique ou privée
Customer:Person	Personne
Transaction	Transfert d'argent entre deux clients

Nous présentons dans le tableau 2 quelques règles d'inférences avec le langage SWRL que nous avons défini avec les experts.

TABLEAU 2. *Règles*

Nom	Description
Règle 1 : PaysEtranger	La transaction est effectuée dans un pays différent de celui de l'agent
Règle 2 : ZoneUE-US	le client se situe en Europe ou aux États-Unis et il réalise une transaction avec une devise différente de l'euro ou du dollar.
Règle 3 : MontantJour	Examine le montant total des transactions effectuées
Règle 4 : NbDevises	Transactions dans plus de 5 devises
Règle 5 : Expiration	Vérifie la date et le montant
Règle 6 : Triangulaire	Vérifie le transfert d'argent

1: $\text{Transaction}(?t) \wedge \text{TransactionCountry}(?t, ?tc) \wedge \text{Agent}(?a) \wedge \text{AgentCountry}(?a, ?ac) \wedge \text{notEqual}(?tc, ?ac) \rightarrow \text{IsFraud}(?t, \text{true})$

2: $\text{Transaction}(?t) \wedge \text{currency}(?t, ?c) \wedge \text{TransactionCountry}(?t, ?tc) \wedge (\text{notEqual}(?c, \text{"euro"}) \vee \text{notEqual}(?c, \text{"dollar"})) \wedge (\text{equal}(?tc, \text{"Fr"}) \vee \text{equal}(?tc, \text{"Usa"})) \rightarrow \text{IsFraud}(?t, \text{true})$

3: $\text{Transaction}(?t) \wedge \text{amount}(?t, ?amt) \wedge \text{Agent}(?a) \wedge \text{DailyAmount}(?a, ?dam) \wedge \text{lessThan}(?amt, 100) \wedge \text{greaterThan}(?dam, 10000) \rightarrow \text{IsFraud}(?t, \text{true})$

4: $\text{Transaction}(?t) \wedge \text{Agent}(?a) \wedge \text{AgentCurrencyCount}(?a, ?count) \wedge \text{greaterThan}(?count, 5) \rightarrow \text{IsFraud}(?t, \text{true})$

5: $\text{Transaction}(?t) \wedge \text{Agent}(?a) \wedge \text{LastTransactionDate}(?ltd, ?a) \wedge$
 $\text{greaterThan}(\text{monthsBetween}(\text{now}(), ?ltd), 6) \wedge \text{amount}(?t, ?amt) \wedge$
 $\text{greaterThan}(?amt, 1000) \rightarrow \text{IsFraud}(?t, \text{true})$

6: $\text{Transaction}(?t1) \wedge \text{TransactionDate}(?t1, ?td1) \wedge \text{DebitorAgent}(?da) \wedge$
 $\text{hasTransactionWithSameCreditorAndDate}(?da, ?td1, ?t2List) \wedge$
 $\text{ListSize}(?t2List, ?size) \wedge \text{GreaterThant}(?size, 0) \wedge \text{CreditorAgent}(?t2, ?ca) \wedge$
 $\text{TransactionDate}(?t2, ?td2) \wedge \text{CreditorAgent}(?ca) \wedge \text{Equals}(?ca, ?da) \wedge$
 $\text{Equals}(?td2, ?td1) \wedge \text{Amount}(?t1, ?amount1) \wedge \text{ListContains}(?t2List, ?t2) \wedge$
 $\text{Amount}(?t2, ?amount2) \wedge \text{Equals}(?amount1, ?amount2) \rightarrow \text{IsFraud}(?t1, \text{true})$

Les expérimentations ont été réalisées avec un jeu de données de 1000000 de transactions provenant du réseau SWIFT. Il s'agit de transactions SWIFT où nous avons extrait les champs relatifs à la transaction et au client. Avant d'utiliser notre ontologie, il était important de s'assurer que l'ontologie était valide et cohérente. Pour ce faire, nous avons soumis notre ontologie à des experts financiers qui l'ont évaluée en termes de vocabulaire, de concepts, de hiérarchie des données et de sémantique. L'ontologie est créée avec protégé 5.5.0². Nous avons également utilisé les raisonneurs Fact++ et HermiT pour vérifier la consistance et la cohérence de notre ontologie financière. Les raisonneurs ont été utilisés pour vérifier que les données ont été correctement liées entre elles et que les inférences logiques sont cohérentes avec les données dans l'ontologie. Le peuplement d'ontologie financière est une étape cruciale pour analyser les données financières et identifier les transactions frauduleuses et suspectes. Il consiste à créer des instances, des concepts, des propriétés d'objet et des propriétés de données d'ontologie. Avec *Owlready2*³, nous avons peuplé notre ontologie avec 1000000 transactions financières. Après avoir peuplé l'ontologie, nous avons utilisé *SPARQL* pour exécuter des requêtes pour identifier les transactions suspectes ou frauduleuses. Finalement, pour faciliter l'analyse des données et visualiser les relations entre les transactions, nous avons utilisé la bibliothèque *Pyvis*⁴ pour afficher les résultats de nos requêtes *SPARQL* sous forme de graphe (figure 1).

4. Conclusion

Nous avons proposé dans cet article une ontologie de domaine pour les transactions interbancaires du réseau SWIFT avec les informations du client (KYC). Notre travail s'inscrit dans le cadre d'un projet de collaboration avec l'entreprise SKAIzen Group. La construction de l'ontologie OntoFiC permet la modélisation des transactions SWIFT et des clients. Cette ontologie est peuplée à partir des transactions issues d'un jeu de données réel. Notre approche de détection de fraudes repose sur le raisonnement basé sur les règles, nous proposons également de visualiser nos requêtes *SPARQL* avec un outil de visualisation et la bibliothèque *pyvis*. La validation de notre approche a été réalisée à travers des expérimentations, nous avons obtenu des résultats

2. <https://protege.stanford.edu>

3. <http://owlready2.readthedocs.io>

4. <https://pyvis.readthedocs.io>

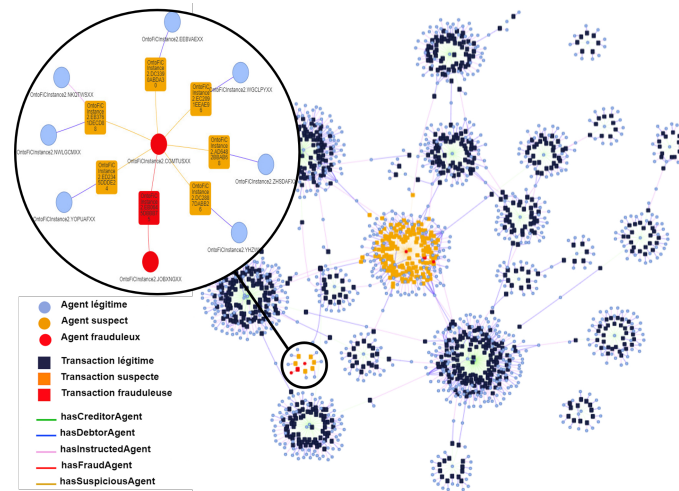


FIGURE 1. *Visualisation des transactions frauduleuses*

prometteurs. Cette approche peut être utilisée en complément d’outils de prévention et de détection de fraudes pour les institutions financières. Dans nos travaux futurs, nous prévoyons d’étendre notre ontologie, en générant dynamiquement de nouvelles règles de détection de transactions frauduleuses avec les techniques d’apprentissage automatique.

Remerciements Ce travail est soutenu à la fois par l’entreprise SKAIZen Group, l’ANRT et l’ANR (France Relance).

Bibliographie

- Ahmed M., Ansar K., Muckley C. B., Khan A., Anjum A., Talha M. (2021). A semantic rule based digital fraud detection. *PeerJ Computer Science*.
- Attigeri G., M M M., Pai R., Kulkarni R. (2018, 01). Knowledge base ontology building for fraud detection using topic modeling. *Procedia Computer Science*, vol. 135, p. 369-376.
- Auger B., Chergui H., Chehade Y., Kadri J. E., Abrouk L., Cabioch N. (2022). Construction d’une ontologie dans le domaine financier pour la détection de fraudes. In *Inforsid*, p. 157–162.
- El Orche A., Bahaj M., Ain Alhayat S. (2018). Ontology based on electronic payment fraud prevention. *Faculty of Sciences Technologies HASSAN 1*.
- Hussaini A., Guessoum Z., Laurent E. (2022, 09). Elaboration of financial fraud ontology. In, p. 277-285.
- Jabbari A., Sauvage O., Zeine H., Chergui H. (2020). A french corpus and annotation schema for named entity recognition and relation extraction of financial news. In *Proceedings of the 12th language resources and evaluation conference*, p. 2293–2299.

Caractérisation des fausses nouvelles

Une approche basée sur la modélisation conceptuelle

Nicolas Belloir^{1,2}, Wassila Ouerdane³, Oscar Pastor⁴

1. CReC St-Cyr, Académie Militaire de St Cyr Coëtquidan, Guer, France

2. IRISA, Vannes, France

3. MISC, Centrale Supélec, Université Paris-Saclay, Gif sur Yvette, France

4. PROS Research Group, Universitat Politècnica de València, Spain

Contact : nicolas.belloir@irisa.fr

RÉFÉRENCE DE L'ARTICLE INTERNATIONAL. Cet article est un résumé de: Belloir N., Ouerdane W., Pastor, O.: *Characterizing Fake News: A Conceptual Modeling-based Approach*. 41st International Conference on Conceptual Modeling (ER) 2022 : Hyderabad, India: 115-129

MOTS-CLES : Modélisation conceptuelle, fausses nouvelles, intelligence artificielle explicable

KEYWORDS : Conceptual Modeling, Fake News, Explainable Artificial Intelligences

La désinformation, et en particulier l'un de ses mécanismes les plus voyants, les fausses nouvelles, est devenue un réel problème de société. Être capable d'identifier puis de lutter contre des fausses nouvelles est une nécessité. Dans ce contexte, il ne suffit pas de pouvoir identifier une fausse nouvelle, il faut également pouvoir expliquer la raison qui amène à ce résultat. Même si de nombreuses actions louables telles que le « fact checking » existent, force est de constater que seule l'utilisation de processus automatisés permet de lutter contre la masse des fausses nouvelles.

De nombreuses approches visant à détecter des fausses nouvelles de manière automatique existent, utilisant en majorité des techniques d'intelligence artificielle (IA). Ces dernières, souvent basées sur des approches de « machine learning », sont inégales et forcément dépendantes du jeu de données avec lequel elles sont entraînées. De plus, si ces techniques peuvent dire si oui ou non une nouvelle est une fausse nouvelle, elles ne permettent pas d'expliquer pourquoi.

Dans ce contexte, nous défendons l'idée qu'il est nécessaire de disposer d'un cadre précis permettant de comprendre les concepts inhérents au fonctionnement des fausses nouvelles. L'objectif est de pouvoir s'appuyer sur ce cadre de manière à justifier par la description les mécanismes impliqués permettant de déterminer en quoi telle nouvelle est une fausse nouvelle. En effet, tout système d'information, destiné à traiter des informations sur les fausses nouvelles, doit identifier en détail

quelles sont les entités pertinentes qui caractérisent conceptuellement les différentes dimensions de ces dernières. Plus précisément, nous suivons un processus basé sur l'intelligence artificielle explicable (XAI) proposé par (Spreeuwenberg, 2019) pour faciliter la construction de modèles bien justifiés et explicables pour la détection de fausses nouvelles. Notre objectif est de proposer une approche qui soit compréhensible, fiable et gérable par les humains. Les différentes étapes suggérées par cette approche sont les suivantes : (i) Obtenir une compréhension partagée du domaine, (ii) comprendre la tâche à effectuer et sélectionner le bon périmètre, (iii) collecter les bonnes données et améliorer leur qualité, (iv) sélectionner les techniques d'IA qui donnent des résultats, (v) générer de bonnes explications, et (vi) faire évoluer la solution dans le temps. Notre contribution établit la fondation d'un tel processus en résolvant la première étape, essentielle, qui consiste à obtenir une compréhension commune de ce qu'est une fausse nouvelle à travers un modèle conceptuel. À partir de cette base conceptuelle solide, le reste du processus XAI proposé peut être appliqué de manière fiable. L'explicabilité de notre approche est guidée par le modèle conceptuel qui est conforme au cœur de la contribution : avoir une définition ontologiquement bien fondée de ce qu'est une fausse nouvelle, qui est directement dérivée du modèle conceptuel.

Ainsi, nous avons spécifié un modèle conceptuel caractérisant le concept de fausse nouvelle et ses principales caractéristiques. Dans (Belloir et al, 2022), nous avons identifié ces dernières à partir de la littérature SHS, sans toutefois les spécifier. Parmi celles-ci, la première concerne l'*origine* des fausses nouvelles. Le constat est fait qu'une fausse nouvelle est rarement créée par une personne isolée. Identifier sa chaîne de création permet de comprendre quel objectif elle sert, et donc de mieux la mesurer et la contrer. Deuxièmement, les fausses nouvelles sont généralement construites sur une *déformation de la réalité*. Identifier cette distorsion est un point essentiel pour permettre son explication. Troisièmement, afin de renforcer sa crédibilité, une fausse nouvelle s'appuie souvent sur une « autorité ». Cette dernière peut être réelle ou fictive, et son message peut lui-même être sorti de son contexte ou totalement inventé. Là encore, il est opportun de pouvoir l'identifier. Enfin, les fausses nouvelles sont construites pour influencer l'opinion de leur cible en générant une *charge émotionnelle*. La réaction de la cible l'amène à tirer certaines conclusions et à changer son opinion sur un sujet. Savoir identifier ce mécanisme permet ainsi de le pointer du doigt. Ainsi le modèle conceptuel que nous présentons est articulé autour des trois dimensions principales suivantes : (i) l'identification de l'origine des fausses nouvelles, (ii) la relation entre les faits vrais et faux, et (iii) la cible comprenant l'opinion et l'émotion. Le modèle conceptuel ainsi proposé a été modélisé en UML, en focalisant sur un niveau de détail plus fin et exploitable.

Bibliographie

- Belloir N, Ouerdane W, Pastor O, Frugier E et de Barmon L.A. (2022), A conceptual characterization of fake news: a positioning paper. *Actes de RCIS 2022*. Barcelone.
- Spreeuwenberg, S. (2019): *AIX: Artificial Intelligence needs eXplanation: Why and how transparency increases the success of AI solutions*. LibRT BV, Amsterdam.

ST-FRAUD : un outil de détection de transactions frauduleuses

Hamza Chergui^{1,2}, Lylia Abrouk¹, Nadine Cullot¹,
Nicolas Cabioch²

1. Laboratoire d'Informatique de Bourgogne - EA 7534
Université de Bourgogne

hamza.chergui,lylia.abrouk,nadine.cullot@u-bourgogne.fr

2. SKAIZen Group

hchergui, ncabioch@skaizengroup.fr

RÉSUMÉ. ST-Fraud est un outil de détection de transactions frauduleuses utilisant les techniques d'apprentissage automatique. Il est composé de trois modules : l'entraînement d'un modèle de classification capable de distinguer les transactions frauduleuses des légitimes, l'identification de types de fraudes pour améliorer le modèle, et le contrôle des transactions détectées frauduleuses avec des informations utiles pour les experts chargés de vérifier ces transactions.

ABSTRACT. ST-Fraud is a fraud transaction detection tool that uses machine learning techniques. It consists of three modules: training a classification model capable of distinguishing fraudulent transactions from legitimate ones, identifying types of fraud to improve the model, and monitoring detected fraudulent transactions with useful information for experts responsible for verifying these transactions.

MOTS-CLÉS : apprentissage automatique, outil, détection de fraude, finance

KEYWORDS: machine learning, tool, fraud detection, finance

1. Introduction

Ces dernières années, plusieurs travaux s'intéressent à la lutte contre la fraude financière. Les systèmes actuels ne sont pas assez efficaces, nos travaux portent sur l'amélioration de ces systèmes avec des techniques basées sur l'apprentissage automatique (Hamza *et al.*, 2023 ; Chergui *et al.*, 2023) pour les institutions financières faisant parties du réseau SWIFT.

SWIFT est un réseau international connectant plus de 11 000 organisations bancaires réparties dans près de 200 pays à l'aide de transactions. Les systèmes de détection de fraude actuels sont basés sur des règles prédéfinies, ils détectent un faible nombre de transactions frauduleuses et génèrent un nombre élevé de faux positifs (transactions

légitimes détectées frauduleuses). Les fraudeurs adaptent constamment leur manière de frauder afin de contourner ces règles. Pour renforcer ces systèmes, les techniques basées sur l'apprentissage automatique sont un réel atout : "*elles offrent la puissance numérique et la flexibilité fonctionnelle nécessaires pour identifier des modèles complexes*" (Dixon *et al.*, 2020).

Dans cet article, nous présentons notre outil ST-Fraud conçu pour détecter les transactions frauduleuses et faciliter leurs vérifications par les experts chargés de contrôler les transactions bloquées par les systèmes de lutte contre la fraude. Ce travail s'inscrit dans le cadre d'une collaboration avec l'entreprise SKAIZen Group sur des transactions du réseau SWIFT.

2. L'outil ST-Fraud

ST-Fraud est une application composée de trois modules :

1. L'import de données, le calcul de caractéristiques et l'entraînement de modèles
2. L'identification des types de fraudes et l'amélioration du modèle
3. L'analyse des transactions frauduleuses

2.1. Entraînement du modèle de classification :

Le premier module (figure 2.1) permet d'entraîner un modèle de classification binaire. Pour cela, un jeu de données labellisé importé est séparé en un jeu d'entraînement et un jeu de test. Les caractéristiques nécessaires pour distinguer les transactions légitimes des frauduleuses sont calculées sur des périodes prédéfinies. Ensuite, un algorithme de classification est sélectionné pour entraîner le modèle. Sa performance est évaluée avec le f1-score : une mesure de moyenne pondérée entre le rappel et la précision d'un modèle.

2.2. Analyse des clusters de transactions frauduleuses

Le deuxième module (figure 2.2) est dédié à l'identification des types de fraudes du jeu de données. Nous formons un sous-jeu de données composées des transactions prédites frauduleuses par le modèle entraîné dans la section 2.1. Ensuite, nous réalisons un clustering de ces transactions avec l'algorithme k-means. Puis, nous analysons les clusters à l'aide du modèle de classification et le framework SHAP capable d'interpréter ses prédictions. Nous avons choisi deux types de visualisations intégrés dans le framework SHAP: les heatmaps pour une visualisation locale des transactions au sein des clusters, et les beeswarm pour une visualisation globale des transactions entre les clusters. Ces visualisations vont permettre aux experts d'associer un type de fraude aux clusters. Le modèle est ensuite amélioré en un modèle de classification multi-classes capable d'attribuer aux transactions les classes légitimes et frauduleuses. Cela permet aux experts de faciliter leurs enquêtes pour bloquer ou libérer les transactions détectées frauduleuses par le système.

Model Training

Choose a file

Drag and drop file here
Limit 100GB per file

Browse files

Split training and testing set

October

January December

History Period

3D × 1M ×

Choose the classifier

Catboost Random Forest XGBoost

Name the model

model-V1

Submit

FIGURE 1. *Entraînement d'un modèle*

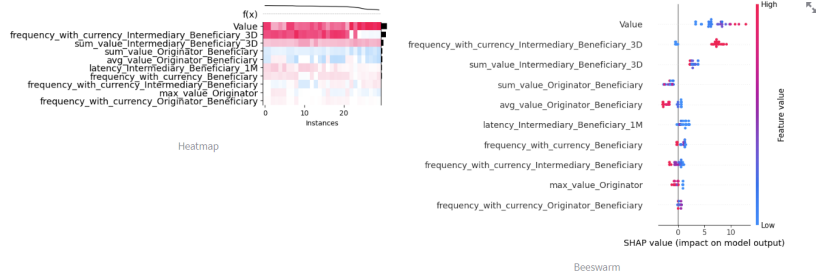
2.3. *Examen des transactions frauduleuses :*

Le dernier module de l'outil est réservé aux experts financiers chargés de contrôler les transactions frauduleuses. Nous présentons sur ce module (figure 2.3) les informations relatives à chaque transaction. Nous utilisons le modèle de classification multiclassées pour afficher le pourcentage d'appartenance aux classes légitimes et frauduleuses. Nous avons ajouté des informations non réelles sur les acteurs par manque de données. Un projet de recherche est mené pour la collecte d'informations sur les acteurs à partir de sources externes (DBpedia, articles) (Auger *et al.*, 2022).

Bibliographie

Auger B., Chergui H., Chehade Y., Kadri J. E., Abrouk L., Cabioch N. (2022). Construction d'une ontologie dans le domaine financier pour la détection de fraudes. In *Inforsid'23*, p. 157–162.

Cluster 2



Transactions number

30

Transactions average amount

1714702

Name the fraud type

DAF

Train Multi-Class Model

FIGURE 2. Identification d'un type de fraude

Check Fraudulent Transaction

Select model

model-V1

Amount: 878634 Currency: A4B

Date: 2019-01-01 09:46:00

Originator country IL to beneficiary country LK



BIC: BTBFILXX
 Name: Philippe Brassac
 Nationality: France
 Company: Chief_executive_officer

Transactions number

82070

Transactions average amount

625108



BIC: RJSGLKXX
 Name: Lakshmi Mittal
 Nationality: British_people
 Company: ArcelorMittal

Transactions number

1

Transactions average amount

878634

95.0% legit

Previous

5.0% DAF

Legitimate

0.0% SF

Fraudulent

0.0% LAF

Next

FIGURE 3. Examen d'une transaction frauduleuse

Chergui H., Abrouk L., Cullot N., Cabioch N. (2023). Réduction du risque du coût d'un modèle dans la détection de fraude financière. In *Extraction et gestion des connaissances, EGC 2023*, vol. E-39, p. 641–642.

Dixon M., Halperin I., Bilokon P. (2020). *Machine learning in finance: From theory to practice*. Springer International Publishing.

Hamza C., Lylia A., Nadine C., Nicolas C. (2023). Semi-supervised method to detect fraudulent transactions and identify fraud types while minimizing mounting costs. *International Journal of Advanced Computer Science and Applications*, vol. 14, n° 2.

La Sécurité Psychologique et ses Conséquences sur Les Equipes Scrum

Théo Moussaoui^{1,2}, Rébecca Deneckère²

1. Caisse des Dépôts et Consignations, 2 avenue Pierre Mendès France, 75914 Paris 13, France

2. Centre de Recherche en Informatique, Université Paris 1 Panthéon-Sorbonne, 90 rue de Tolbiac, 75013 Paris, France

RESUME. Les pratiques agiles sont de plus en plus utilisées et préconisent les interactions entre les différentes parties prenantes. Cependant, ces interactions ne peuvent se faire de manière efficace que lorsque les individus se sentent en confiance les uns avec les autres et n'ont pas peur des conséquences de certaines prises de risques. Cette caractéristique s'appelle la Sécurité Psychologique. Nous étudions ici la littérature sur la sécurité psychologique et comment elle peut influencer les éléments de Scrum, l'une des pratiques agiles les plus utilisées.

ABSTRACT. Agile practices are increasingly being used and promote interactions between different stakeholders. However, these interactions can only be effective when individuals feel confident with each other and are not afraid of the consequences of certain risk-taking. This characteristic is called Psychological Safety. We study here the literature on psychological safety and how it can influence the elements of Scrum, one of the most used agile practices.

Mots-clés : Pratiques Agiles, Scrum, Sécurité Psychologique.

KEYWORDS: Agile Practices, Scrum, Psychological Safety

1. Introduction

Dans le domaine du développement de logiciels, les pratiques évoluent constamment. L'un des principaux changements est la popularisation de l'approche agile depuis la création du Manifeste Agile en 2001. Cette approche est de plus en plus répandue, au point que 86% des équipes de développement de logiciels ont adopté des pratiques agiles en 2021¹. L'approche agile met l'accent sur les individus, leurs interactions et l'auto-organisation des équipes, ce qui place les personnes au

¹ 5th State of Agile Report : Agile adoption accelerates across the enterprise - Digital.ai, 2021 <https://info.digital.ai/rs/981-LQX-968/images/RE-SA-15th-Annual-State-Of-Agile-Report.pdf>

cœur du processus et les responsabilise². Parmi les équipes agiles, 66% utilisent le framework Scrum. Il s'agit d'un framework léger qui aide les individus, les équipes et les organisations à créer de la valeur grâce à des solutions adaptatives pour des problèmes complexes (Sutherland et Schwaber, 2020). Depuis 2010, Scrum est défini par le Scrum Guide, avec sept versions à ce jour, la dernière datant de 2020.

Les besoins d'interaction au sein de ce cadre sont très importants et il est donc raisonnable de supposer que la qualité de ces interactions a une influence sur la progression des équipes. Dans ce contexte, le concept de sécurité psychologique³, largement étudié depuis les années 1990, joue un rôle important. En effet, la sécurité psychologique décrit la perception des conséquences de la prise de risques interpersonnels dans un contexte particulier, comme le lieu de travail (Edmondson, 1999). Elle trouve son origine dans la recherche organisationnelle et est apparue dans les années 60, lors des réflexions sur les changements organisationnels. On avançait alors que la sécurité psychologique était essentielle pour que les personnes se sentent en sécurité et capables de changer leur comportement en réponse aux défis organisationnels changeants (Edmondson, 2014). Pour Schein (1993), la sécurité psychologique aide les individus à surmonter leur caractère défensif et leur peur de l'apprentissage lorsqu'on leur présente des données qui contredisent leur ressenti. En sécurité psychologique, les individus sont plus enclins à se concentrer sur les objectifs collectifs et à prévenir les problèmes plutôt qu'à s'auto-protéger. Plusieurs définitions de la sécurité psychologique existent et se ressemblent, mais nous avons choisi celle d'Edmondson (1999), pour qui *la sécurité psychologique est la croyance partagée par les membres d'une équipe que celle-ci est sûre quant à la prise de risques interpersonnels*. Dans un environnement accordant une grande importance à la communication, une sécurité psychologique élevée devrait être bénéfique pour tous. Mais alors, qu'est-ce qui influence cette sécurité psychologique et quelles sont ses conséquences ? Edmondson *et al.* (2004) ainsi que Newman *et al.* (2017) ont établi des cartographies sur lesquelles nous nous appuyons dans cette étude. Nous les avons croisées avec d'autres sources pertinentes. Marder *et al.* (2021) ont adopté une approche inverse en étudiant l'impact de l'utilisation du Framework Scrum sur la sécurité psychologique.

Dans cette étude, nous allons explorer la littérature afin de comprendre quels sont les éléments susceptibles d'influencer la sécurité psychologique (Section 2). Ensuite, nous allons nous concentrer sur les équipes Scrum et tenter de comprendre comment la sécurité psychologique peut influencer les éléments de Scrum (Section 3). Nous concluons dans la Section 4.

2. Les facteurs de variation de la sécurité psychologique

Plusieurs études ont tenté de cartographier les antécédents et les conséquences de la sécurité psychologique, notamment celle d'Edmondson, Kramer et Cook (2004) et

² Agile Alliance- Agile Manifesto, février 2001 <https://agilemanifesto.org/>

³ Le terme « Sécurité Psychologique » est la traduction de « Psychological Safety » que l'on trouve dans la littérature anglaise.

celle de Newman *et al.* (2017). La première se concentre sur les antécédents organisationnels qui influencent la sécurité psychologique de l'équipe et les conséquences sur les comportements d'apprentissage de l'équipe. La seconde a réalisé une segmentation en plusieurs niveaux pour étudier le problème de manière plus précise. Elle étudie la sécurité psychologique à trois niveaux différents avec des antécédents influant sur ces différents niveaux et des conséquences distinctes à chaque niveau. Elle souligne également les effets modérateurs de la sécurité psychologique en spécifiant qu'au niveau organisationnel, ce sont les pratiques organisationnelles qui influencent la sécurité psychologique. Au niveau individuel, les différences et les relations interpersonnelles influencent la sécurité psychologique sur les attitudes de travail et les comportements. Enfin, au niveau de l'équipe, le leadership, les caractéristiques d'équipe et les réseaux de relations influencent la sécurité psychologique de l'équipe qui a des impacts au niveau de l'équipe, organisationnel et individuel. On peut constater que la sécurité psychologique est influencée par différents paramètres selon le niveau considéré, et que les conséquences sont également différentes. Nous nous sommes appuyés sur ces cartographies existantes pour identifier le cadre illustré dans la figure 1.

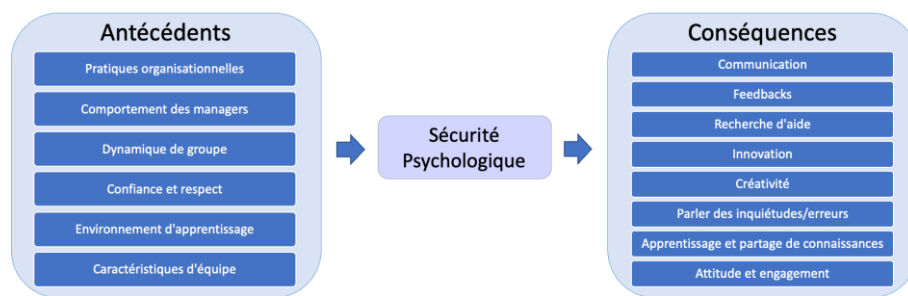


Figure 1. Antécédents et conséquences de la sécurité psychologique

2.1. Les antécédents de la sécurité psychologique

Il existe six antécédents de la sécurité psychologique : pratiques organisationnelles, comportement des managers, dynamique de groupe, confiance et respect, environnement d'apprentissage et caractéristiques de l'équipe.

Pratiques Organisationnelles. Les pratiques organisationnelles sont déterminantes pour la sécurité psychologique de l'organisation. Les audits internes, par exemple, qui évaluent le fonctionnement de l'entité étudiée peuvent avoir des répercussions négatives sur les relations émotionnelles du groupe étudié (Eden et Moriah, 1996). Ces audits impliquent fortement les parties prenantes à toutes les phases du processus (Carmeli et Zisu, 2009). On a constaté que le support organisationnel perçu et la confiance organisationnelle sont tous deux liés à la sécurité psychologique, et que cette dernière a une relation significative avec l'audit interne. Les études de (Schein, 1993) et (Eden et Moriah, 1996) sont les deux premières à avoir mesuré la sécurité psychologique au niveau organisationnel en se basant sur les perceptions des employés et en adaptant les axes de mesure proposés par

Edmondson (1999) pour les équipes. Cependant, la pertinence de ces études a été remise en question ultérieurement car il est difficile de parler de sécurité psychologique à l'échelle de toute l'organisation, étant donné que la perception de chaque individu peut être grandement influencée par le leadership et les caractéristiques de son équipe (Carmeli et al., 2010). Pourtant, au niveau individuel, il s'avère que les pratiques organisationnelles solidaires sont liées à des bénéfices pour les employés comme l'engagement auprès de l'organisation mais également de la performance, corrélé à la perception de la sécurité psychologique croissante (Newman *et al.*, 2017). Sur cette question de l'influence des organisations, Schein (2010) a mis en avant des éléments clés montrant que la culture d'une organisation au travers les valeurs et croyances qu'elle porte peut permettre de réduire l'anxiété des personnes face à l'ambiguïté et l'incertitude.

Comportement des Managers. Le comportement des managers est crucial, car les membres des équipes y sont sensibles (Tyler et Lind, 1992). Les managers ont en effet valeur d'exemple. S'ils n'admettent pas leurs erreurs, les membres de l'équipe ne seront pas enclins à admettre les leurs, craignant des conséquences négatives (Edmondson, 2004). Si, en revanche, les managers admettent leurs erreurs et partagent ouvertement qu'il faut en parler, alors les membres de l'équipe seront plus enclins à parler librement (Edmondson, 2001). Il s'avère que les mauvaises nouvelles ne sont pas souvent rapportées à la hiérarchie (Lee, 1993) et que les employés ont moins tendance à demander de l'aide à leur supérieur qu'à leurs collègues (Lee, 1997). Les managers qui encouragent les commentaires envoient un signal clair que l'avis de chacun est important. S'ils adoptent un comportement fermé, distant et directif, les membres de l'équipe ne se sentiront pas en sécurité psychologique. En revanche, s'ils adoptent une attitude d'ouverture, ils favorisent la sécurité psychologique. En sortant de leur propre perspective et en adoptant une posture de coach, ils deviennent des modèles et sont plus à même de comprendre les informations qui leur sont communiquées (Edmondson, 2004).

Dynamique de groupe. Les interactions entre les différents rôles et personnalités que les membres d'une équipe adoptent dans leur travail sont à l'origine de dynamiques informelles émergentes (Kahn, 1990). Que ce soit de manière consciente ou non, chaque individu incarne un rôle ou une posture qui influe sur la manière dont il interagit avec les autres membres de l'équipe, ce qui peut faciliter ou compliquer les relations interpersonnelles. L'ensemble de ces interactions crée des dynamiques de groupe informelles qui ont un impact sur la transmission d'informations. Ainsi, la sécurité psychologique d'une équipe est influencée par les dynamiques informelles qui se créent au sein de celle-ci (Edmondson et al., 2004).

Confiance et Respect. La notion de confiance revêt une grande importance étant donné que la gouvernance, et par conséquent les managers, ont un impact direct sur la résolution des conflits. Alors que cette résolution ainsi que les répercussions des conflits représentent 20% de leur temps (Chan et al., 2008) (Schermerhorn et al., 1998), un style de gestion des conflits plus collaboratif a tendance à accroître la confiance et la satisfaction entre le manager et les employés. Grâce à une telle gestion favorisant la confiance et la satisfaction, les employés sont davantage disposés à communiquer et ainsi à améliorer la sécurité psychologique (Erkutlu et

Chafra, 2015). Lorsque des personnes au sein d'une équipe se font confiance, la sécurité psychologique est significativement augmentée (May *et al.*, 2004). En effet, les membres d'une équipe auront plus facilement envie de parler de sujets (notamment sensibles) aux personnes en qui ils ont confiance. Les personnes ont plus tendance à partager des idées et des concepts quand ils pensent que les critiques vont être plus constructives que destructives (Kahn, 1990). Les personnes qui sentent leur légitimité remise en question se sentent plus jugées et craignent de porter préjudice à leur réputation (Moingeon et Edmondson, 1998). Sur ces aspects, si les relations au sein d'un groupe sont caractérisées par la confiance et le respect, alors les individus peuvent penser qu'on leur accordera le bénéfice du doute, une caractéristique de la sécurité psychologique (Edmondson *et al.*, 2004).

Environnement d'apprentissage. La création d'environnements d'apprentissage permet d'instaurer un climat de sécurité psychologique, particulièrement au travers de la mise en place de champs d'application, qui sont utilisés dans divers domaines tels que la pratique médicale, mais qui peuvent également être transposés à d'autres secteurs. Il est important de souligner que l'idée de l'environnement d'apprentissage est à la base de cette approche. Les champs d'application, qui consistent en des forums permettant la pratique et l'expérimentation avant la réflexion sur les résultats (Peter, 1990), s'avèrent être très bénéfiques car ils autorisent l'erreur lors de l'entraînement. Cette pratique n'est pas nouvelle et a déjà été utilisée dans le milieu professionnel du sport ou des pilotes d'avion qui s'entraînent régulièrement de cette façon (Edmondson *et al.*, 2004). Il est évident que les pilotes d'avion ne peuvent pas se contenter de lire le manuel de l'avion et s'élancer sans pratique préalable sur simulateur. Ce dernier leur permet d'avoir un environnement d'apprentissage sécurisé pour expérimenter, apprendre de leurs erreurs, identifier leurs faiblesses et savoir comment réagir face à l'imprévu (Edmondson *et al.*, 2004). Les champs d'application favorisent ainsi la sécurité psychologique en réduisant les risques mais aussi en mettant en lumière l'importance de l'apprentissage, ainsi que le fait qu'on ne peut pas se reposer sur une seule tentative pour réussir (Edmondson *et al.*, 2004).

Caractéristiques d'équipe. Des études ont cherché à mettre en évidence l'influence de la perception des caractéristiques de l'équipe sur la sécurité psychologique, tant au niveau de l'équipe que de ses membres. À titre individuel, la perception commune de certains éléments tels que la compréhension partagée des systèmes par l'ensemble de l'équipe (Bendoly, 2014) ou le climat d'amélioration continue de la qualité (Rathert *et al.*, 2009) a un impact sur la performance et l'engagement organisationnel en favorisant une augmentation de la sécurité psychologique au sein de l'équipe (Newman *et al.*, 2017). Au niveau des équipes, des caractéristiques telles que les récompenses d'équipe partagées (Chen et Tjosvold, 2012), les structures d'équipes formelles (Bresman et Zellmer-Bruhn, 2013) ou un engagement d'équipe dans le travail de délimitation (Faraj et Yan, 2009) tendent à améliorer la sécurité psychologique. Il a également été constaté que lorsque les membres de l'équipe avaient la responsabilité partagée de mauvaises décisions d'investissement, la sécurité psychologique leur donnait le courage d'admettre l'échec, contrairement au cas où ils étaient individuellement responsables et où le manque de sécurité psychologique augmentait leur engagement envers de mauvaises décisions

d'investissement. Les membres de l'équipe sont donc plus motivés pour sauver le projet et prendre les choses en main dès que possible (O'Neill, 2009). Malgré de nombreuses recherches sur l'influence des caractéristiques de l'équipe, il existe relativement peu de théories pour expliquer leurs effets (Newman *et al.*, 2017).

2.2. Les conséquences de la sécurité psychologique

Il y a 8 conséquences de la sécurité psychologique : Communication, Feedbacks, Recherche d'aide, Innovation, Créativité, Parler des inquiétudes/erreurs, Apprentissage et partage de connaissances, Attitude et engagement.

Communication. La sécurité psychologique a un impact positif sur la communication interpersonnelle, comme cela a été défini précédemment. Il a été observé qu'elle a un lien avec les silences au sein d'un groupe : dans un groupe où le niveau de sécurité psychologique est élevé, il y aura moins de silences, et inversement, selon Brinsfield (2013). En ce qui concerne les sujets difficiles à aborder, les personnes auront davantage tendance à en parler à des personnes en qui elles ont confiance et qui présentent un haut niveau de sécurité psychologique (Tynan, 2005). Les individus seront également plus enclins à remonter des erreurs à leur superviseur si celui-ci leur inspire confiance (Kakar, 2018). En somme, la sécurité psychologique a une incidence sur de nombreux aspects fondamentaux de la communication.

Feedbacks. Les retours spontanés sont plus sincères lorsqu'ils sont adressés à des personnes ayant un haut niveau de sécurité psychologique (Tynan, 2005). La vérité a plus de chances de surgir si les personnes sont moins sensibles aux critiques. La recherche de feedback est essentielle car elle reflète l'état d'esprit d'une personne ou d'un groupe. Si l'on recherche des feedbacks, cela indique que l'on est activement à la recherche d'amélioration. Cette recherche est plus développée lorsque la sécurité psychologique de l'équipe est élevée (Edmondson *et al.*, 2004).

Recherche d'aide. En cas de faible sécurité psychologique, la recherche d'aide peut sembler délicate car cela implique de reconnaître son ignorance. Si un collègue hiérarchiquement supérieur est présent, cet effet peut être amplifié (Lee, 1997). Le fait de demander de l'aide expose alors à un risque de paraître incompetent, mais une sécurité psychologique élevée peut encourager la recherche d'aide malgré cette crainte de jugement et de conséquences négatives (Edmondson *et al.*, 2004).

Innovation. Au sein des équipes de développement logiciel, il est observé que la sécurité psychologique, combinée à une bonne cohésion d'équipe, a un effet très positif sur l'innovation, en particulier sur le partage des connaissances favorisant l'innovation (Gu *et al.*, 2013). Plus la sécurité psychologique est élevée, plus le partage des connaissances est important. C'est pourquoi la sécurité psychologique est largement mise en avant, car il est clairement indiqué que le partage des connaissances ne peut être contraint, contrôlé ou imposé, mais doit être induit par la création d'un environnement d'équipe propice (Kakar, 2018). En d'autres termes, la sécurité psychologique joue un rôle important dans la relation entre l'innovation et la performance. En effet, les groupes qui sont peu innovants ont généralement une performance faible, mais ceux qui sont très innovants et ont un faible niveau de

sécurité psychologique ont une performance encore plus faible que les premiers. Cela montre que les entreprises ne peuvent pas se fier uniquement à l'innovation, car cela peut conduire à des efforts inutiles. Il est donc important de créer un environnement de travail qui favorise à la fois l'innovation et la sécurité psychologique pour améliorer la performance globale (Baer et Frese, 2003).

Créativité. La perception collective d'une sécurité psychologique est liée à la pensée créative et la prise de risque (Palanski et Vogelgesang, 2011). Lorsque les individus au sein d'une organisation ont la perception que la sécurité psychologique est élevée, la créativité est augmentée (Edmondson et al., 2004).

Parler des inquiétudes / erreurs. Au sein des relations interpersonnelles et d'équipe, la sécurité psychologique est considérée comme un moyen de faire pencher la balance en faveur des avantages de dire les choses plutôt que de leur coût (Edmondson et al., 2004). Il est donc souligné que la capacité à s'exprimer permet non seulement de mettre l'accent sur le positif, mais surtout d'aborder avec plus de confiance les erreurs. Au sein des équipes, cela se traduit par un défi du statu quo en abordant ouvertement les points d'échec ou en proposant des améliorations (Carmeli et Gittel, 2009). Cela déclenche alors une dynamique d'apprentissage.

Apprentissage et partage de connaissances. Il existe une corrélation entre la sécurité psychologique et les comportements d'apprentissage, que ce soit au niveau individuel (Liu et al., 2014) ou au niveau d'équipe (Ortega et al., 2010, Stalmeijer et al., 2007, Wong et al., 2010). Ces études montrent que la sécurité psychologique élevée améliore l'apprentissage car elle aide les personnes à apprendre de leurs échecs (Carmeli, 2007). De plus, la littérature sur l'apprentissage social et le traitement de l'information montre qu'un environnement psychologique sûr favorise l'apprentissage lorsque les membres de l'équipe sont encouragés à prendre des risques, à expérimenter et à partager leurs idées et leurs inquiétudes (Newman et al., 2017). La sécurité psychologique a également un impact sur le partage de connaissances, car un haut niveau influence positivement le partage de connaissances entre les membres (Xu et Yang, 2010).

Attitude et engagement. Au niveau individuel, des études ont montré que la sécurité psychologique a un effet positif sur les attitudes des personnes envers le travail d'équipe (Ulloa et Adams, 2004), ainsi que sur leur engagement organisationnel (Rathert *et al.*, 2009) (Chen *et al.*, 2014). Les chercheurs ont exploré le lien entre la sécurité psychologique et les attitudes de travail en se basant sur la théorie des échanges sociaux, selon laquelle la sécurité psychologique créée par des pratiques de soutien mène à une réciprocité en termes d'attitudes de travail. Ceci est particulièrement évident dans (De Clercq & Rius, 2007) et (Chen *et al.*, 2014).

3. Les effets de la sécurité psychologique sur les équipes Scrum

Nous allons nous intéresser à Scrum, tel que défini dans le Scrum Guide de 2020⁴, qui s'appuie sur plus de 25 ans d'expérience de la communauté. Afin d'illustrer notre

⁴ Les citations en italique et entre guillemets sont tirées directement du Scrum Guide 2020 (Sutherland et Schwaber, 2020).

propos, nous allons simuler un faible niveau de sécurité psychologique en supposant les effets que cela peut avoir, que nous désignerons par "insécurité psychologique", en nous basant sur les éléments précédemment relevés : les piliers, les valeurs, les événements, les artefacts et engagements, ainsi que la Scrum team (figure 2).

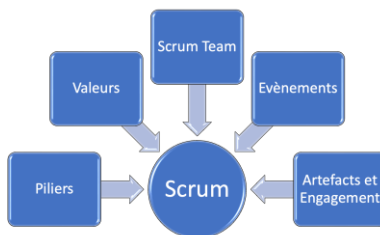


Figure 2. Éléments de Scrum à prendre en compte pour la sécurité psychologique

Les piliers. « Scrum est fondé sur l'empirisme et la pensée Lean. L'empirisme affirme que la connaissance provient de l'expérience et que la prise de décision s'appuie sur l'observation de faits. La pensée Lean réduit le gaspillage et se focalise sur l'essentiel ». Scrum est profondément marqué par ces deux concepts, si bien que tout ce qui peut les entraver, directement ou indirectement, affectera Scrum. Cependant, nous n'allons pas rentrer ici dans les détails de la pensée Lean puisque nous nous concentrons sur le Scrum Guide. Les piliers de l'empirisme sont la transparence, l'inspection et l'adaptation (voir tableau 1).

Tableau 1. Effets sur les Piliers de Scrum

Pilier	Analyse
Transparence. « Le processus et le travail émergents doivent être visibles pour ceux qui effectuent le travail ainsi que pour ceux qui le reçoivent ».	Un niveau de sécurité psychologique faible entraîne une communication moins efficace et plus de silence. Les membres de l'équipe sont alors moins enclins à rendre visible le travail émergent, ce qui nuit à la transparence et crée de l'opacité. Une transparence trompeuse peut engendrer du gaspillage.
Inspection. « Les artefacts Scrum et les progrès vers les objectifs convenus doivent être inspectés fréquemment et avec diligence pour détecter des écarts ou des problèmes potentiellement indésirables ».	Si le niveau de sécurité psychologique est faible, cela peut détériorer la communication et rendre la détection des problèmes ou des écarts plus difficile lors de l'inspection. De plus, l'inspection d'éléments peu transparents est rendue difficile, voire peu efficace.
Adaptation. « Si certains aspects d'un processus s'écartent des limites acceptables ou si le produit résultant est inacceptable, alors le processus appliqué ou les éléments produits doivent être adaptés. Une Scrum Team doit s'adapter dès lors que l'inspection révèle quelque chose de nouveau ».	L'effet boule de neige de l'insécurité psychologique perturbe le fonctionnement de l'empirisme de Scrum, entravant la transparence, l'inspection et l'adaptation. Les éléments biaisés ou incomplets issus d'une transparence altérée peuvent entraîner des adaptations inappropriées ou une absence d'adaptation. En outre, la remise en question nécessaire lors de l'adaptation est compromise si les membres de l'équipe ne se sentent pas en sécurité pour exprimer leurs inquiétudes.

Les valeurs. « Lorsque ces valeurs sont incarnées par la Scrum Team et les personnes avec lesquelles elle travaille, alors les piliers empiriques Scrum de transparence, d'inspection et d'adaptation émergent en consolidant la confiance ». En partant de ce constat, il est possible d'affirmer que si la sécurité psychologique altère les valeurs, alors l'application de Scrum risque d'être compromise. Le Tableau 2 souligne que 4 des 5 valeurs de Scrum peuvent être directement affectées.

Tableau 2. Effets sur les Valeurs de Scrum

Valeur	Analyse
Engagement sur l'atteinte de ses objectifs et à se soutenir mutuellement	Il a été mentionné précédemment que la sécurité psychologique peut avoir un impact sur l'engagement au sein de l'équipe. Dans ce contexte, l'insécurité psychologique peut remettre en question la valeur de l'engagement.
Focus [sur les objectifs à atteindre]	Comme la communication est essentielle pour que l'équipe reste concentrée sur ses objectifs, une insécurité psychologique peut compromettre la valeur du focus.
Ouverture sur le travail et les défis à relever	Dans une approche empirique comme celle de Scrum, le travail émerge au fil du temps, ce qui implique qu'il y ait des essais et des erreurs. L'insécurité psychologique peut cependant décourager les membres de l'équipe de parler ouvertement de leurs échecs et de leurs défis, ce qui peut les empêcher de résoudre efficacement les problèmes.
Respect mutuel au sein de l'équipe et envers elle	Le respect au sein de la Scrum Team est étroitement lié à une bonne communication, qui favorise la confiance et le respect mutuel. Ainsi, lorsque la sécurité psychologique est présente, on peut s'attendre à ce que le respect soit également présent, car cela favorise une bonne communication.
Courage de mener les bonnes actions et de travailler sur des problèmes difficiles	La valeur du courage, qui permet de dire non, de demander de l'aide, d'aborder les sujets difficiles, est également étroitement liée à la sécurité psychologique. Les feedbacks, la demande d'aide et la discussion des inquiétudes et des erreurs sont tous des éléments clés qui y sont liés. Dans le cas d'une insécurité psychologique, la valeur du courage peut être altérée.

La Scrum Team. « La Scrum Team est une seule et même unité stable, composée de professionnels focalisés sur un seul objectif à la fois : le Product Goal ». L'importance de la valeur de focus est mise en avant. Le Scrum Guide souligne également que les équipes Scrum sont autogérées et prennent des décisions internes sur qui fait quoi, quand et comment. Cette autogestion implique nécessairement une forte communication, qui est compromise en cas d'insécurité psychologique. Au-delà de la simple communication, presque tous les sujets mentionnés dans la littérature ont un impact sur l'autogestion. On peut donc supposer qu'un faible niveau de sécurité psychologique peut nuire à l'autogestion de la Scrum Team. Il est important de noter que le Scrum Guide est peu prescriptif en ce qui concerne les activités liées aux responsabilités de cette dernière. En effet, le Scrum Master, le Product Owner, mais surtout les Developers auront des activités différentes en fonction du contexte (Tableau 3).

Tableau 3. Effets sur la Scrum Team

Team	Analyse
Developers	<p>« Les Developers s'engagent à réaliser tous les aspects d'un incrément utilisable à chaque Sprint. » Puisque l'insécurité psychologique peut compromettre l'engagement des membres de l'équipe, ils peuvent ne pas être en mesure de réaliser le travail nécessaire Sprint après Sprint. Ils sont responsables de la création d'un plan pour le Sprint, qui doit être élaboré en collaboration avec l'équipe, impliquant une communication claire et une évaluation critique. En présence d'une insécurité psychologique qui altère la communication et l'engagement, en plus de favoriser le silence, on peut remettre en question leur capacité à co-construire un plan de qualité. De plus, « ils sont redevables d'adapter leur plan chaque jour par rapport au Sprint Goal », ce qui nécessite à la fois de parler des inquiétudes, de demander de l'aide en cas de blocage et d'avoir une réelle attitude d'équipe. En ce qui concerne l'attitude d'équipe et l'engagement, altérables par l'insécurité psychologique, leur redevabilité à se tenir mutuellement responsables en tant que professionnels (Sutherland et Schwaber, 2020) sera également affectée.</p>
Scrum Master	<p>« Le Scrum Master est redevable de la bonne mise en place de Scrum », mais comme la plupart des éléments de Scrum peuvent être impactés, cela peut complètement la dégrader. Pour la réaliser, « il aide à comprendre Scrum en théorie et en pratique, autant au niveau de la Scrum Team que de l'organisation ». Sur ce point, la sécurité psychologique intervient sur le partage de connaissance. Dans son activité d'accompagnement de la Scrum Team sur les questions d'auto-gestion et de pluridisciplinarité, il doit parler ouvertement et partager ses connaissances. La notion de transparence est ici au cœur de l'activité et se retrouve également dans sa responsabilité à retirer les obstacles se dressant sur le chemin de l'équipe. Il a besoin que les obstacles soient visibles pour pouvoir les retirer, la parole ouverte est donc essentielle. Toutes les activités d'accompagnement ont les mêmes risques puisqu'elles touchent toutes au partage de connaissances et à l'attitude de travail d'équipe. Sur ses activités vis-à-vis de l'organisation, la posture étant similaire, on peut dresser les mêmes suppositions, en s'intéressant cette fois-ci à une échelle plus large.</p>
Product Owner (PO)	<p>« Le PO est redevable de la gestion efficace du Product Backlog ce qui implique le développement du Product Goal et la communication explicite de celui-ci ». Il est crucial de rendre l'objectif transparent en l'évoquant ouvertement et régulièrement, mais cette tâche devient plus difficile lorsque la sécurité psychologique est faible. De même, la responsabilité du Product Owner de créer et de communiquer les éléments du Product Backlog et de les ordonner au sein de celui-ci peut également être altérée par une faible sécurité psychologique, puisque cela implique une communication ouverte et transparente en équipe. « Le PO peut déléguer tout ou partie des activités pour lesquelles il est responsable mais il en restera toujours la personne redevable ». Cette délégation demande de la confiance et un certain esprit d'équipe. « Au cours d'un Sprint, le périmètre peut être clarifié et renégocié avec le PO selon ce que l'on apprend ». Dans ce cadre, il faut que la parole soit libre pour parler des apprentissages. Le Scrum Guide nous expose les conditions de réussite du PO : « pour qu'il réussisse, toute l'organisation doit respecter ses choix ». En termes de dépendance avec l'extérieur, il est évident que la sécurité psychologique joue un rôle important. On peut également supposer que cette sécurité exerce une influence sur la capacité à respecter les choix du PO, tout comme la valeur de respect. Les choix et les actions du PO seraient donc moins libres et auraient un impact moindre dans un environnement d'insécurité psychologique.</p>

Les événements. Chaque événement dans Scrum représente une opportunité d'inspection et d'adaptation des artefacts Scrum. Ils sont expressément conçus pour favoriser la transparence. « *Les événements sont utilisés dans Scrum dans le but de créer de la régularité, minimisant le besoin d'avoir d'autres réunions* ». Ils sont ainsi propices à mettre en évidence les piliers de l'empirisme (voir le Tableau 4).

Tableau 4. Effets sur les événements de Scrum

	Analyse
Sprint	« <i>Scrum combine quatre événements formels pour l'inspection et l'adaptation dans un événement conteneur, le Sprint</i> ». Ce sont ces éléments que nous allons étudier.
Sprint Planning	« <i>Il démarre le Sprint en présentant le travail à effectuer. Le plan qui en résulte est créé par le travail collaboratif de toute la Scrum Team</i> ». Il est raisonnable de penser qu'une attitude d'équipe est indispensable, sinon il est possible qu'une partie ou la totalité de l'équipe ne soit pas suffisamment impliquée, ce qui pourrait conduire à une définition insuffisante du plan pour l'itération. « <i>Les trois sujets abordés lors du Sprint Planning sont : Pourquoi ce Sprint est-il important ? Que peut-on faire durant ce Sprint ? Comment le travail choisi sera-t-il réalisé ?</i> » Ces trois questions suivent un ordre logique et chacune a un impact sur la suivante. Si le Sprint Goal est mal défini, le contenu du Sprint ne sera peut-être pas adapté, et la Scrum Team devra réfléchir à des solutions pour des éléments qui ne sont peut-être pas les plus importants. Des éléments prévus dans le Product Backlog peuvent créer des inquiétudes quant à leur faisabilité, mais en cas d'insécurité psychologique, elles peuvent ne pas être abordées, ce qui peut créer des problèmes en cours de Sprint
Daily Scrum	« <i>L'objectif du Daily Scrum est d'inspecter la progression vers le Sprint Goal et l'adaptation du Sprint Backlog si nécessaire, par l'ajustement des futurs travaux planifiés</i> ». Le Daily Scrum représente l'inspection et l'adaptation quotidienne de la Scrum Team. Toutefois, il est possible que l'insécurité psychologique ait un impact considérable sur la progression de l'équipe, étant donné que cette réunion a lieu chaque jour. Si les membres de l'équipe ne se sentent pas en confiance, cela peut les empêcher de partager certaines informations, comme des apprentissages ou des inquiétudes, pouvant remettre en cause l'atteinte du Sprint Goal. Le Scrum Guide souligne l'importance du Daily Scrum pour améliorer la communication, identifier les obstacles, favoriser la prise de décision rapide et éviter d'autres réunions. Toutefois, en cas d'insécurité psychologique, la communication peut être réduite et la qualité du partage d'informations, d'apprentissages et d'inquiétudes limitée, ce qui pourrait impacter les bénéfices de cette réunion.
Sprint Review	« <i>L'objectif de la Sprint Review est d'inspecter le résultat du Sprint et de déterminer les adaptations futures. La Scrum Team présente les résultats de son travail aux principales parties prenantes et les progressions vers le Product Goal sont discutées</i> ». Une fois de plus, l'événement est en parfaite conformité avec les principes empiriques de Scrum et on peut envisager les effets d'un faible niveau de sécurité psychologique. « <i>La Scrum Team et les parties prenantes passent en revue ce qui a été accompli durant le Sprint et ce qui a changé dans leur environnement. Ils collaborent sur la marche à suivre et sur les décisions à prendre</i> ». En l'absence de partage des acquis et d'une attitude collaborative, la revue de Sprint peut se limiter à une simple démonstration, négligeant l'inspection et l'adaptation.
Sprint Retrospective	« <i>Elle consiste à réfléchir à des pistes pour améliorer la qualité et l'efficacité</i> ». C'est peut-être là où l'insécurité psychologique peut exercer la plus grande influence, en particulier les discussions des problèmes et des pistes d'amélioration. En effet, « <i>la Scrum Team discute de ce qui s'est bien passé durant le Sprint, des problèmes</i>

	rencontrés et de la manière dont ces problèmes ont été (ou pas) résolus ». Un faible niveau de sécurité psychologique peut conduire à un manque de communication autour des problèmes, rendant difficile l'identification de pistes d'amélioration, ce qui peut remettre en question l'utilité, voire l'efficacité, de la Sprint Retrospective.
--	---

Les artefacts et leurs engagements. Le Scrum Guide indique que « *des artefacts peu transparents peuvent mener à des décisions qui diminuent la valeur et augmentent le risque* ». Nous avons vu qu'une sécurité psychologique insuffisante peut réduire la transparence et donc que les artefacts Scrum et les engagements associés remplissent moins efficacement leur fonction (voir Tableau 5).

Tableau 5. Effets sur les artefacts et les engagements de Scrum

Artefact	Analyse
Product Backlog & Product Goal	« <i>Le Product Backlog est une liste ordonnée et émergente de ce qui est nécessaire pour améliorer le produit</i> ». Le caractère émergent du Product Backlog implique une collecte fréquente de retours auprès des parties prenantes, ce qui suppose une recherche active de feedback. Or, nous avons observé que cette recherche peut être entravée par l'insécurité psychologique, ce qui risque de réduire la pertinence du Product Backlog. L'affinement du Product Backlog constitue un élément clé de son évolution. « <i>Il consiste à décomposer et à définir davantage les éléments du en éléments plus fins et plus précis</i> ». La qualité de l'affinement du Product Backlog dépend largement de la collaboration et des discussions pertinentes au sein de l'équipe, permettant de clarifier les sujets. En conséquence, une sécurité psychologique insuffisante pourrait compromettre cet affinement et donc l'évolution adéquate du Product Backlog. « <i>Les Developers effectuant le travail étant responsables du dimensionnement</i> » il est possible de considérer comme indispensable la mise en commun des connaissances, dans le cas où des Developers ont déjà été confrontés à un sujet similaire, pour faire part de leur expérience et ainsi contribuer à une estimation plus précise. « <i>Le Product Goal décrit un état futur du produit qui peut servir de cible à la Scrum Team pour planifier.</i> » Il est question ici de la transparence. La définition du Product Goal revêt une importance capitale et dépend du Product Owner. Néanmoins, il représente l'objectif à long terme que la Scrum Team se doit d'atteindre (ou d'abandonner) avant de s'atteler au suivant. Remettre en question le Product Goal exige des échanges pouvant s'avérer difficiles si la ligne directrice principale n'est pas facilement remise en question. En cas d'insécurité psychologique, même si certains membres de l'équipe ont des raisons de croire que le Product Goal n'est plus pertinent, ils ne l'évoqueront pas forcément devant les autres membres de l'équipe.
Sprint Backlog & Sprint Goal	« <i>Le Sprint Backlog est composé du Sprint Goal (le pourquoi), de l'ensemble des éléments du Product Backlog choisis pour le Sprint (le quoi), ainsi que d'un plan d'action pour la réalisation de l'Incrément (le comment)</i> ». On remarque qu'il est très complet. Il est mis à jour tout au long du Sprint. Cependant, pour qu'il soit suffisamment détaillé, il est nécessaire que les Developers soient engagés et participent activement aux discussions. Dans un contexte d'insécurité psychologique, on peut s'attendre à ce qu'il soit incomplet et manque donc de transparence, risquant de manquer des détails pouvant entraîner un échec dans l'atteinte du Sprint Goal. Ce dernier est le dénominateur commun du Sprint Backlog et sa progression doit être inspectée chaque jour. Il est donc nécessaire qu'il soit bien compris et partagé par tous les Developers. Comme pour le Product Goal, il est possible de ne pas l'atteindre ou de se rendre compte qu'il est obsolète. Dans ce cas,

	des discussions doivent avoir lieu pour permettre la transparence. Toutefois, dans un climat d'insécurité psychologique, il est possible que la Scrum Team ne traite pas le sujet et réalise un Sprint de moindre qualité.
Incrément & Definition of Done (DoD)	L'Incrément représente le travail accompli par les Developers pour créer le produit. Cependant, les effets négatifs de l'insécurité psychologique se répercutent plutôt sur les éléments mentionnés précédemment, qui auront ensuite des répercussions sur l'Incrément. Par exemple, un problème non signalé lors du Daily Scrum par les Developers peut entraîner un incrément incomplet qui ne répond pas nécessairement au Sprint Goal. Des retards dans le partage d'informations ou dans l'apprentissage peuvent ralentir les développements et donc la création d'Incréments. « <i>La DoD est une description formelle de l'état de l'Incrément lorsqu'il satisfait les mesures de qualité requises pour le produit</i> ». Une faible sécurité psychologique peut entraîner un manque de transparence qui pourrait limiter la compréhension commune de la DoD, affectant ainsi la qualité de l'Incrément. Étant donné que la DoD est inspectée et adaptée lors de la Sprint Retrospective, un manque de transparence à cet égard rendra son adaptation plus difficile. En outre, étant donné que la qualité ne doit pas baisser, il est important pour la Scrum Team de discuter régulièrement de la DoD.

Récapitulatif. La représentation graphique ci-dessous résume les éléments impactés par les conséquences du manque de sécurité psychologique. Il est à noter que l'Incrément et le Sprint sont les seuls éléments qui ne sont pas directement touchés par ces effets. Cependant, cela ne signifie pas qu'ils ne peuvent pas être altérés d'une manière ou d'une autre dans un contexte d'insécurité psychologique, car de nombreux éléments sont interdépendants et peuvent s'influencer mutuellement.

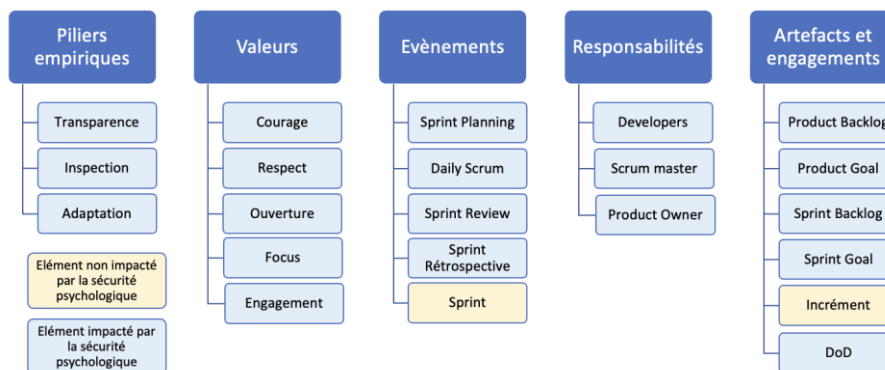


Figure 1. Éléments de Scrum impactés par la sécurité psychologique

4. Conclusion

La sécurité psychologique a des effets non négligeables sur le travail en équipe. Pour le cas des équipes Scrum, le constat est sans appel. En prenant simplement le cadre tel que défini dans le Scrum Guide, il est possible de mettre en avant de très probables dysfonctionnements causés par une insécurité psychologique. Que ce soient les valeurs, qui permettent le succès de Scrum, les piliers empiriques sur lesquels s'appuie Scrum ou tous les autres éléments cités ici, presque aucun

n'échappe à l'influence de la sécurité psychologique. Le niveau de sécurité psychologique d'une équipe est finalement un des facteurs principaux la menant à la réussite du fait de tous les aspects du travail en équipe qu'elle touche, d'autant plus lorsqu'elles évoluent dans un cadre tel que Scrum. La littérature nous montre que de nombreux éléments peuvent la faire varier et il est judicieux pour les entreprises adoptant le cadre Scrum de s'intéresser à ces facteurs pour réussir. Les conséquences sur la réussite peuvent notamment être accompagnées de conséquences humaines puisqu'en cas d'insécurité psychologique, tous les membres de la Scrum Team peuvent être freinés dans l'accomplissement de leurs activités.

Il conviendrait maintenant de tester cette analyse à la réalité du terrain pour valider ou invalider certains des éléments avancés. Comme nous avons seulement pris en compte les éléments définis par le Scrum Guide, nous n'avons pas abordé les pratiques complémentaires courantes dans les équipes Scrum et il serait pertinent de comprendre l'impact sur ces pratiques. Avec l'avènement massif du télétravail, la communication n'est plus la même qu'avant et il faudrait voir si cela a une incidence sur la sécurité psychologique. De plus, on peut se demander si l'adoption de Scrum basée sur ses valeurs ne crée pas justement un environnement propice à la sécurité psychologique, ce qui signifierait que les effets existent dans les deux sens. Il faudrait également identifier des outils pour mesurer la sécurité psychologique d'un individu ou d'une équipe qui ne soient pas trop invasifs pour ne pas perturber les tâches de ceux-ci et éviter les biais d'observation. De même, il faudrait étudier les corrélations d'impacts entre les différents antécédents, les conséquences et les éléments de Scrum de manière plus précise.

Bibliographie

- Baer M., Frese M. (2003). Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance. *J. Organ. Behav. Int. J. Ind. Occup. Organ. Psychol. Behav.*, vol.24, n°1, p. 45-68.
- Bendoly E. (2014). System dynamics understanding in projects: Information sharing, psychological safety, and performance effects. *Prod. Oper. Manag.*, vol.23, n°8.
- Bresman H., Zellmer-Bruhn M. (2013). The structural context of team learning: Effects of organizational and team structure on internal and external learning. *Organ. Sci.*, vol.24, n°4.
- Brinsfield C. T. (2013). Employee silence motives: Investigation of dimensionality and development of measures. *J. Organ. Behav.*, vol.34, n°5, p. 671-697.
- Carmeli A. (2007). Social capital, psychological safety and learning behaviours from failure in organisations. *Long Range Plann.*, vol.40, n°1, p. 30-44.
- Carmeli A., Gittel J. H. (2009). High-quality relationships, psychological safety, and learning from failures in work organizations. *J. Organ. Behav. Int. J. Ind. Occup. Organ. Psychol. Behav.*, vol. 30, n° 6, p. 709-729.
- Carmeli A., Reiter-Palmon R., Ziv E. (2010). Inclusive leadership and employee involvement in creative tasks in the workplace: The mediating role of psychological safety. *Creat. Res. J.*, vol.22, n°3, p. 250-260.
- Carmeli A., Zisu M. (2009). The relational underpinnings of quality internal auditing in medical clinics in Israel. *Soc. Sci. Med.*, vol.68, n°5, p. 894-902.

- Chan K. W., Huang X., Ng P. M. (2008). Managers' conflict management styles and employee attitudinal outcomes: The mediating role of trust. *Asia Pac. J. Manag.*, vol.25, n°2, p. 277-295.
- Chen C., Liao J., Wen P. (2014). Why does formal mentoring matter? The mediating role of psychological safety and the moderating role of power distance orientation in the Chinese context. *Int. J. Hum. Resour. Manag.*, vol.25, n°8, p. 1112-1130.
- Chen G., Tjosvold D. (2012). Shared rewards and goal interdependence for psychological safety among departments in China. *Asia Pac. J. Manag.*, vol.29, n°2, p. 433-452.
- De Clercq D., Rius I. B. (2007). Organizational commitment in Mexican small and medium-sized firms: the role of work status, organizational climate, and entrepreneurial orientation. *J. Small Bus. Manag.*, vol.45, n°4, p. 467-490.
- Eden D., Moriah L. (1996). Impact of internal auditing on branch bank performance: a field experiment. *Organ. Behav. Hum. Decis. Process.*, vol.68, n°3, p. 262-271.
- Edmondson A. (1999). Psychological safety and learning behavior in work teams. *Adm. Sci. Q.*, vol.44, n°2, p. 350-383.
- Edmondson A. C., Bohmer R. M., Pisano G. P. (2001). Disrupted routines: Team learning and new technology implementation in hospitals. *Adm. Sci. Q.*, vol.46, n°4, p. 685-716.
- Edmondson A., Moingeon B. (1999). Learning, trust and organizational change. *Organ. Learn. Learn. Organ.*, p. 157-175.
- Edmondson A. C., Kramer R. M., Cook K. S. (2004). Psychological safety, trust, and learning in organizations: A group-level lens. *Trust Distrust Organ. Dilemmas Approaches*, vol.12, n°2004, p. 239-272
- Edmondson A.C. , Lei Z. (2014). Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct. *Annu. Rev. Organ. Psychol. Organ. Behav.*, vol.1, n°1.
- Erkutlu H., Chafra J. (2015). The mediating roles of psychological safety and employee voice on the relationship between conflict management styles and organizational identification. *Am. J. Bus.*
- Faraj S., Yan A. (2009). Boundary work in knowledge teams. *J. Appl. Psychol.*, vol.94, n°3.
- Gu Q., Wang G. G., Wang L. (2013). Social capital and innovation in R&D teams: the mediating roles of psychological safety and learning from mistakes. *Rd Manag.*, vol.43, n°2.
- Kahn W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Acad. Manage. J.*, vol.33, n°4, p. 692-724.
- Kakar A. K. (2018). How do team cohesion and psychological safety impact knowledge sharing in software development projects?. *Knowl. Process Manag.*, vol.25, n°4.
- Lee F. (1993). Being Polite and Keeping MUM: How Bad News is Communicated in Organizational Hierarchies 1. *J. Appl. Soc. Psychol.*, vol.23, n°14, p. 1124-1149.
- Lee F. (1997). When the going gets tough, do the tough ask for help? Help seeking and power motivation in organizations. *Organ. Behav. Hum. Decis. Process.*, vol. 72, n°3.
- Leroy H., Dierynck B., Anseel F., Simons T., Halbesleben J.R.B., McCaughey D., Savage G.T., Sels L. (2012). Behavioral integrity for safety, priority of safety, psychological safety, and patient safety: A team-level study. *J. Appl. Psychol.*, vol.97, n°6, p. 1273.
- Liu S., Hu J., Li Y., Wang Z., Lin X. (2014). Examining the cross-level relationship between shared leadership and learning in teams: Evidence from China. *Leadersh. Q.*, vol.25, n°2.

- Marder B., Ferguson P., Marchant C., Brennan M., Hedler C., Rossi M., Black S., Doig R. (2021). 'Going agile': Exploring the use of project management tools in fostering psychological safety in group work within management discipline course., *Int. J. Manag. Educ.*, vol.19, n°3, p. 100519.
- May D. R., Gilson R. L., Harter L. M. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *J. Occup. Organ. Psychol.*, vol.77, n°1, p. 11-37.
- Moingeon B., Edmondson A. (1998). 12. Trust and Organisational Learning. *Trust Econ. Learn.*, p. 247.
- Newman A., Donohue R., Eva N. (2017). Psychological safety: A systematic review of the literature. *Hum. Resour. Manag. Rev.*, vol.27, n°3, p. 521-535.
- O'Neill O. A. (2009). Workplace Expression of Emotions and Escalation of Commitment 1. *J. Appl. Soc. Psychol.*, vol.39, n°10, p. 2396-2424.
- Ortega A., Sánchez-Manzanares M., Gil F., Rico R. (2010). Team learning and effectiveness in virtual project teams: The role of beliefs about interpersonal context. *Span. J. Psychol.*, vol.13, n°1, p. 267-276.
- Palanski M. E., Vogelgesang G. R. (2011). Virtuous creativity: The effects of leader behavioural integrity on follower creative thinking and risk taking. *Can. J. Adm. Sci. Can. Sci. Adm.*, vol.28, n°3, p. 259-269.
- Peter S. (1990). The fifth discipline. Art Pract. Learn. Organ. Doubleday Currence N. Y.
- Rathert C., Ishqaidif G., May D. R. (2009). Improving work environments in health care: test of a theoretical framework. *Health Care Manage. Rev.*, vol.34, n°4, p. 334-343.
- Schein E. H. (1993). On dialogue, culture, and organizational learning. *Organ. Dyn.*, vol.22, n°2, p. 40-52.
- Schermerhorn J. R., Hunt J. G., Osborn R. N. (1998). Basic organizational behavior. J. Wiley Hoboken, NJ, USA.
- Sutherland J., Schwaber K. (2020). The Scrum Guide. [En ligne]. Disponible sur: <https://scrumguides.org/>
- Schein E. H. (2010). Organizational culture and leadership, vol.2. John Wiley & Sons.
- Stalmeijer R. E., Gijsselaers W. H., Wolfhagen I. H., Harendza S., Scherpbier A. J. (2007). How interdisciplinary teams can create multi-disciplinary education: the interplay between team processes and educational quality. *Med. Educ.*, vol. 41, no 11.
- Tyler T. R., Lind E. A. (1992). A relational model of authority in groups. *Advances in experimental social psychology*, vol.25, Elsevier, p. 115-191.
- Tynan R. (2005). The effects of threat sensitivity and face giving on dyadic psychological safety and upward communication 1. *J. Appl. Soc. Psychol.*, vol.35, n°2, p. 223-247.
- Ulloa B. C. R., Adams S. G. (2004). Attitude toward teamwork and effective teaming. *Team Perform. Manag. Int. J.*, vol.10, n°7/8, p. 145-151.
- Wong A., Tjosvold D., Lu J. (2010). Leadership values and learning in China: The mediating role of psychological safety. *Asia Pac. J. Hum. Resour.*, vol.48, n°1, p. 86-107.
- Xu Y., Yang Y. (2010). Student learning in business simulation: An empirical investigation. *J. Educ. Bus.*, vol.85, n°4, p. 223-228.

Enrichissement et visualisation de trajectoires de visiteurs dans un musée

Jérémy Richard¹, Cyril Faucher¹, Karell Bertet¹

Laboratoire L3i, Institut LUDI, Avenue Michel Crépeau
17042 La Rochelle Cedex 1, France

jeremy.richard2@univ-lr.fr, cyril.faucher@univ-lr.fr, karell.bertet@univ-lr.fr

RÉSUMÉ. Cet article présente une démonstration d'un module d'un système d'information conçu pour gérer le contenu des applications de visite mobile pour les musées. Notre objectif est d'offrir une meilleure compréhension des comportements et préférences des visiteurs lors de leur visite. Dans cette démonstration, nous nous concentrons sur la présentation des avantages de l'enrichissement des trajectoires des visiteurs de musée en combinant les données de logs des applications mobiles avec les données de localisation grâce à une visualisation Web.

ABSTRACT. This article presents a demonstration of one module of an information system designed to manage the content of mobile visit applications for museums. Our goal is to offer a better understanding of the behaviors and preferences of visitors during their visit. In this demonstration, we focus on showing the benefits of enriching museum visitor trajectories by combining mobile application log data with location data through a web-based visualization.

MOTS-CLÉS : Enrichissement sémantique - trajectoire sémantique - positionnement en intérieur

KEYWORDS: Semantic enrichment - semantic path - indoor positioning

1. Motivations

De nombreux musées disposent d'applications mobiles permettant de visiter ceux-ci. Les contenus de ces applications sont généralement constitués de parcours thématiques regroupant les oeuvres, chacune est décrite et illustrée par des médias. Au vu des coûts de ces applications et parfois du matériel à maintenir pour le bon fonctionnement de celui-ci, l'évaluation de ces outils devient un enjeu pour les conservateurs des musées. En partenariat avec des musées de Nouvelle-Aquitaine nous développons des outils numériques répliquables permettant de valoriser les collections des musées. Dans cet article nous présentons un des modules d'un système d'information permettant de gérer le contenu d'applications mobiles de visite pour les musées. A terme, nos méthodes et outils doivent fournir une aide au pilotage à destination des conservateurs de musées avec une maîtrise de bout en bout allant de la gestion de contenus (liés aux oeuvres) jusqu'à l'analyse du comportement des visiteurs permettant de valider ou d'invalidier des choix d'organisation. Cette démonstration a pour objectif de mettre en évidence l'intérêt d'une approche enrichie de l'analyse des trajectoires des visiteurs de musées, qui combine des données de localisation avec des données de logs provenant d'une application mobile. En effet, l'analyse des mouvements des visiteurs seule ne permet pas toujours d'obtenir une compréhension détaillée de leur comportement. Il est ainsi possible d'observer un afflux d'activités dans une zone particulière sans pouvoir conclure de manière certaine que l'individu est intéressé par les œuvres présentées ou qu'il les apprécie. En intégrant ces données de localisation avec d'autres informations issues de l'application mobile, il est possible de mieux cerner le comportement des visiteurs et de déduire leur niveau d'engagement dans leur visite.

2. Trajectoires sémantiques

Une trajectoire enrichie de diverses informations, apportant du contexte à celle-ci, est appelée une trajectoire sémantique. La trajectoire sémantique a la particularité de combiner les dimensions temporelle, sémantique et spatiale. (Spaccapietra *et al.*, 2008) représentaient dès 2008 une trajectoire comme une séquence de "Stop and Move". Cette représentation de trajectoire sous forme de séquences est toutefois commune dans l'état de l'art. Elle associe à un symbole (tel un lieu ou un type de mobilité) une donnée temporelle et cette succession de symboles prend ainsi la forme d'une séquence temporelle. Le modèle APM (Activity, POI et Move model) (Moreau *et al.*, 2019), représente ici une trajectoire sémantique sous la forme d'une séquence d'épisodes. Un épisode est défini comme étant un couple $(s, [l, u])$ où s est un symbole et $[l, u]$ est un intervalle avec l est la date de début et u la date de fin. Des traces de log d'une application mobile de visite dans un musée apportent une information qualitative du comportement du visiteur que la trajectoire seule de déplacement ne peut pas fournir. Notre objectif est d'enrichir les données quantitatives avec des données qualitatives en nous appuyant sur la dimension temporelle des données. Il s'agit d'un processus d'annotation de la trajectoire brute afin de lier les données qualitatives aux données quantitatives et leur donner une sémantique. Une annotation est une donnée additionnelle associée à la trajectoire, à une sous-trajectoire ou à un point de la trajectoire (Parent *et al.*, 2013). La trajectoire ainsi enrichie devient une trajectoire sémantique. Par le biais de la trajectoire sémantique, nous pouvons dès lors mieux comprendre le comportement d'un individu de par ses déplacements, mais aussi ses agissements. Nous nous intéresserons ici à la trajectoire de visite dans un musée. Par le biais d'une application mobile, "Visite Musée", nous viendrons enrichir les déplacements des personnes déambulant dans le musée avec des informations de leur activité sur cette application. Ce faisant, l'objectif est de préciser l'activité des visiteurs et de fournir un journal de bord reposant sur l'aspect spatial et contextuel de la visite.

3. Constitution du jeu de données

L'application "Visite Musée" est une application qui a été développée par la licence professionnelle spécialisée dans le développement d'application mobile de l'IUT de La Rochelle. Celle-ci se veut agir comme un compagnon de visite, accompagnant le visiteur lors de son parcours du musée. Elle est actuellement utilisée dans bon nombre de musées de la région Nouvelle-Aquitaine. Cette application se compose de trois vues principales, montrées figure 1.

- "Vue accueil" : la vue principale de l'application. Elle permet de choisir le parcours du musée et de naviguer vers les autres vues. (figure 1.a)
- "Vue plan" : permet de localiser les oeuvres à l'intérieur du musée.(figure 1.b)
- "Vue détail" : permet d'avoir plus d'informations sur une oeuvre enregistrée dans l'application, on pourra ainsi avoir plus de contextes sur une oeuvre, son auteur, son histoire. (figure 1.c)

L'application "Visite musée" utilise le récepteur à signaux Bluetooth BLE du téléphone ou de la tablette qui servira ici afin de localiser le parcours de l'individu lors de sa déambulation dans le musée. Cette expérimentation a été réalisée sur une journée entière lors des Journées du Européennes du Patrimoine en 2021 au Museum d'Histoire Naturelle de La Rochelle. Nous avons pu prêter plusieurs tablettes à des visiteurs volontaires pour participer à l'expérimentation et par conséquent à utiliser l'application "Visite musée" durant tout ou partie de leur déambulation dans le musée. Nous avons ainsi récupéré 32 visites complètes avec les informations de localisation des visiteurs, ainsi que les logs d'activités sur l'application. L'ensemble des données sont transmises au système d'information muséal en temps réel via internet.



a) Vue accueil

b) Vue plan

c) Vue détail

FIGURE 1. Les différentes vues de l'application

4. Modélisation de la trajectoire en intérieur

La modélisation et la visualisation de trajectoires enrichies par des données de contexte a été effectuée sur les données récoltées afin de pouvoir combiner les données de déplacements et les données d'utilisation d'application mobile durant la période de visite des volontaires à l'étude. Les données issues de l'application, sous une forme sémantique retranscrivent l'enchaînement des différentes vues au sein de celle-ci. Nous représentons cette modélisation figure 2.

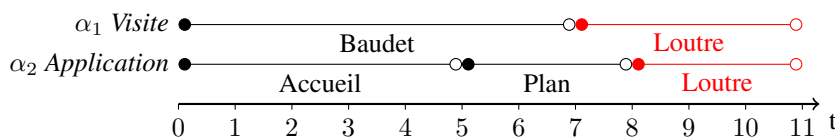


FIGURE 2. Modélisation de la trajectoire sémantique au sein du musée

Nous notons une dimension "déplacement" ainsi qu'une dimension "Application". Les données de déplacement sont directement issues d'un algorithme de positionnement en intérieur et mis sous la forme d'une séquence temporelle d'intervalles retraçant le parcours du visiteur. La dimension "Application" est quant à elle formalisée de la même façon, en créant des intervalles temporels basés sur le temps passé sur une des vues de l'application.

5. Démonstrateur

Nous avons conçu une visualisation afin de montrer de façon cohérente plusieurs informations, dans l'objectif d'offrir un "journal de bord" de la visite d'un utilisateur à partir d'une trajectoire sémantiquement enrichie. Les figures 4 à 7 illustrent l'interface utilisateur du démonstrateur. L'IHM est ainsi fondée sur le formalisme vu précédemment. On y retrouve une visualisation des deux types de données synchronisées. Les informations suivantes sont retranscrites sur une même visualisation :

- La trajectoire de la personne, dans le musée ;

- L'utilisation de l'application, avec le visuel de la vue;
- Le rapport au temps, afin de dresser le lien entre les déplacements et l'utilisation de l'application;
- Dessiner, au fur et à mesure, le trajet de la personne afin de montrer son utilisation de l'espace muséal;

Plus particulièrement, elles montrent la représentation d'une trajectoire d'un visiteur, non seulement l'utilisation de l'espace, mais aussi le type de comportement à chaque pas temporel : Recherche d'oeuvre à découvrir, découverte ou bien simple déplacement. On y retrouve une frise chronologique synthétisant les intervalles d'activité et un aperçu du contenu consulté sur l'application pour un intervalle temporel donné. Quand le contenu de l'application et le déplacement sont concordants, nous le signifions en entourant le cercle de position d'une épaisse bordure noire. L'intensité en couleur du point est directement reliée au temps passé à la position indiquée. Cette visualisation a été pensée pour un affichage dynamique du résumé de la visite, on peut y voir le graphe de déplacement se construire peu à peu, et la frise d'utilisation de l'application défile en fonction du temps qui passe. En ajoutant une autre dimension contextuelle dans la visualisation d'une visite nous avons un regard plus "proche" de la réalité et plus fidèle à l'expérience de visite de l'individu.

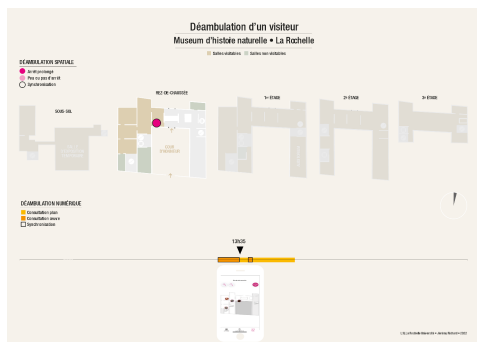


FIGURE 4. Instantané de la trajectoire d'un visiteur (1)

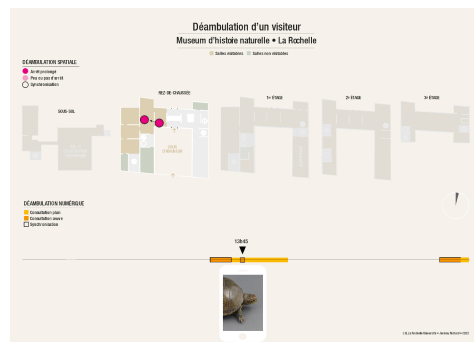


FIGURE 5. Instantané de la trajectoire du visiteur (2)

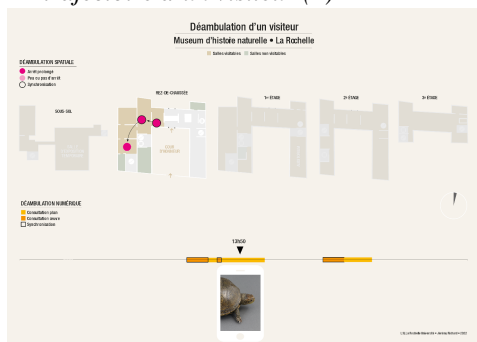


FIGURE 6. Instantané de la trajectoire du visiteur (3)

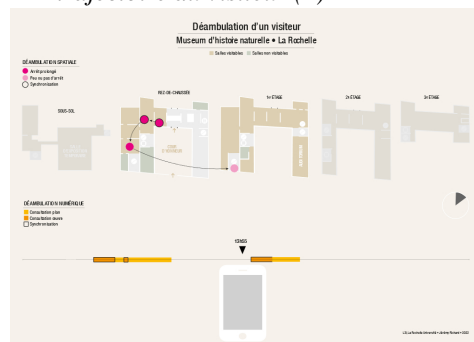


FIGURE 7. Instantané de la trajectoire du visiteur (4)

FIGURE 3. Les traces de mobilités d'un visiteur

6. Conclusion

Nous proposons une démonstration pour mettre en évidence l'intérêt de notre méthode d'enrichissement à base de trajectoires sémantiques. L'outil présenté est un module de visualisation qui combine à la fois une vue spatiale et une vue chronologie de données de déplacement enrichies par données de logs d'activité provenant d'une application mobile dédiée à la découverte de collections muséales. Les travaux futurs consisteront à mettre en avant les parcours les plus communs en utilisant notamment des algorithmes issus de l'analyse formelle des concepts.

Bibliographie

- Moreau C., Devogele T., Etienne L. (2019). Calcul de similarité sémantique entre trajectoires. *Revue Internationale de Géomatique*, vol. 29, n° 1, p. 107-127.
- Parent C., Spaccapietra S., Renso C., Andrienko G., Andrienko N., Bogorny V. *et al.* (2013, août). Semantic trajectories modeling and analysis. *ACM Comput. Surv.*, vol. 45, n° 4, p. 1–32. Consulté sur <https://dl.acm.org/doi/10.1145/2501654.2501656>
- Spaccapietra S., Parent C., Damiani M. L., Macedo J. A. de, Porto F., Vangenot C. (2008). A conceptual view on trajectories. *Data and Knowledge Engineering*, vol. 65, n° 1, p. 126 - 146.

Éthique de la gestion du consentement au traitement de données personnelles : une analyse au prisme des *dark patterns*

Robert Viseur¹

1. Service TIC, FWEG, UMONS
17 place Warocqué, B-7000 Mons, Belgique
robert.viseur@umons.ac.be

RÉSUMÉ. Face au développement du big data et à son application aux données à caractère personnel, le législateur européen a conçu un cadre juridique protecteur : le « Règlement général de protection des données » (RGPD). Son entrée en application le 25 mai 2018 a ramené au cœur des préoccupations des acteurs de la publicité ciblée et de l'analyse de performances la question du recueil du consentement préalable à tout traitement de données à caractère personnel. En a découlé l'apparition de prestataires spécialisés dans la création d'interfaces de recueil de consentement, les Consent Management Platforms (CMP), mais aussi la multiplication des dark patterns visant à forcer l'obtention dudit consentement. Dans cette recherche exploratoire, nous utilisons une typologie de dark patterns pour classer les pratiques identifiées puis discuter, d'une part, leur légalité, d'autre part, leur éthique (du point de vue des approches utilitariste et déontologique), enfin les meilleures manières de lutter contre les dérives observées. Nous montrons en particulier l'existence d'une zone grise permettant aux professionnels de maximiser, parfois provisoirement, la quantité de données à caractère personnel collectées.

ABSTRACT. Faced with the development of big data and its application to personal data, the European legislator has provided a protective legal framework: the "General Data Protection Regulation" (GDPR). When it came into force on May 25, 2018, the issue of obtaining consent prior to any processing of personal data was at the heart of the concerns of targeted advertising, particularly programmatic advertising, and of performance analysis. This has led to the emergence of service providers specialized in the creation of consent collection interfaces, the Consent Management Platforms (CMP), but also to the multiplication of dark patterns aiming at forcing the obtaining of such consent. In this exploratory research, we use a typology of dark patterns to classify the identified practices and then discuss, on the one hand, their legality, on the other hand, their ethics (from the point of view of utilitarian and deontological approaches), and finally, the best ways to fight against the observed drifts. In particular, we show the existence of a grey area allowing professionals to maximize, sometimes temporarily, the amount of personal data collected.

Mots-clés : vie privée, éthique, cookie, dark pattern, RGPD, CMP.

KEYWORDS: privacy, ethics, cookie, dark pattern, DGPR, CMP

1. Introduction

Avec une capitalisation boursière cumulée de l'ordre de dix mille milliards de dollars, les GAFAM sont en vingt cinq ans devenus incontournables dans le paysage économique. Au cœur de leurs modèles d'affaires : la collecte massive (*big data*) de données à caractère personnel permettant le ciblage de la publicité (Facebook, Google, Microsoft), le classement personnalisé d'informations (Facebook, Google, Microsoft) et la suggestion de recommandations d'achats (Amazon). Cette évolution a amené le législateur européen à proposer, avec le « *Règlement général de protection des données* » (RGPD), un cadre harmonisé de protection des données à caractère personnel pour le citoyen européen.

Le RGPD impacte les GAFAM mais d'une manière générale toute entreprise ou association recueillant des données à caractère personnel, et ce quel qu'en soit le volume. Sont notamment impactés la myriade d'acteurs impliqués dans les dispositifs de publicité ciblée, en particulier dans celui, fragmenté, de la publicité programmatique (Allary & Balusseau, 2018), mais aussi les entreprises ayant mis en œuvre des modèles d'affaires basés sur l'accès gratuit à des contenus en ligne, dès lors valorisés par la publicité en ligne et la revente de données. Face à la complexité d'obtention du consentement et au risque accru de refus de la part des utilisateurs, les éditeurs de sites web recourent, d'une part, aux services d'acteurs spécialisés dans la création d'interfaces de recueil du consentement (Hils et al., 2020), soit des CMP (*Consent Management Platforms*), d'autre part, à la mise en œuvre de *dark patterns* visant à « extorquer » le consentement à l'aide d'artifices techniques et visuels à la légalité et à l'éthique discutables (Nouwens et al., 2020). Cette recherche exploratoire propose dès lors une analyse de ces *dark patterns* appliqués aux interfaces de recueil de consentement (CMP) et aux conditions d'utilisation des services (CGU) fixant les finalités de la collecte de données.

Notre article est organisé en quatre sections. Dans une première section, nous proposons un état de l'art sur la publicité ciblée, la collecte de données, le *tracking*, le RGPD et le concept de *dark pattern*. Dans une seconde section, nous présentons succinctement notre méthodologie. Dans une troisième section, nous analysons les techniques trompeuses appliquées au recueil de consentement, les classons en utilisant une typologie de *dark patterns* puis évaluons leurs caractères légal et éthique. Dans une quatrième section, et avant de conclure, nous discutons l'impact de ces *dark patterns* sur le caractère libre et éclairé du consentement fourni et proposons plusieurs pistes d'action en matière de régulation permettant de davantage respecter la vie privée des utilisateurs de services en ligne.

2. État de l'art

Mesguish et Thomas (2013) distinguent quatre âges du Web. Le premier, s'étendant de 1994 à 1996, est baptisé « *Web des pionniers* ». Cette expression désigne le développement d'un Web encore réduit en taille alimenté par des pionniers technophiles. De 1996 à 2004, le « *Web des documents* » s'accompagne d'une explosion du nombre de sites permise par la facilité des nouveaux outils d'édition de contenu et alimentée par les débuts du commerce électronique. Le « *Web social* », parfois appelé Web 2.0, s'étend de 2004 à 2010. Il voit une implication plus importante des utilisateurs dans la création et l'enrichissement des

contenus. Dès 2010, le « *Web temps réel* » se développe avec la part croissante des réseaux sociaux (audience) ainsi que le développement des *smartphones* et des tablettes. Enfin, l'essor des objets connectés annonce un cinquième âge du web, que nous baptiserons « *Web ubiquitaire* » permettant la création d'un double numérique sous la forme d'un profil et ouvrant de nouvelles perspectives en termes de services individualisés. Cette extraction continue de données personnelles conduit à la mise en place d'un « *capitalisme de surveillance* » (Zuboff, 2019) couvrant à la fois les mondes virtuels (p. ex. moteurs de recherche) et réels (p. ex. objets connectés).

Cette évolution s'est accompagnée d'une mutation de la publicité en ligne sous des formes de plus en plus ciblées (Peyrat, 2009), jusqu'à la publicité comportementale cherchant à coller au plus près des centres d'intérêt immédiats des consommateurs tels que révélés par leur historique de navigation. Cette personnalisation avancée suppose un travail permanent de *tracking* (p. ex. *cookies*) et d'analyse de données (profilage) par les régies publicitaires. Les plus connus sont Google et Facebook mais d'autres acteurs plus petits sont notamment actifs en publicité programmatique (Allary & Balusseau, 2018). Cette dernière organise, au travers de plates-formes dédiées, la rencontre en quasi temps réel des offres d'espaces publicitaires mis aux enchères par des éditeurs de sites web, et les demandes émises par les annonceurs à la recherche d'une clientèle spécifique. L'objectif est dès lors de coller aux plus près des préoccupations (commerciales) des internautes de manière à maximiser les conversions (p. ex. achat d'un produit) et la rentabilité des campagnes publicitaires.

La collecte de données personnelles à partir du navigateur va s'appuyer sur des traceurs variés incluant les *cookies* mais aussi des empreintes (*fingerprinting*) calculées sur base des caractéristiques du navigateur et de la machine sur laquelle il est installé (Viseur, 2021). Ces identifiants peuvent donc être déterministes (p. ex. *login* et *cookies*) ou probabilistes (p. ex. *fingerprinting* et adresse IP) c'est-à-dire utilisables par croisement de plusieurs données. Les entreprises vont par ailleurs combiner des données propres (*first party data*), notamment issues de leur logiciel CRM (*Customer Relationship Management*) et du Web (p. ex. site web et réseaux sociaux), des données issues de partenaires (*second party data*) et des données achetées auprès de tiers tels que les courtiers en données (*third party data*). Le croisement de ces données accroît les possibilités d'association à un profil. Aux États-Unis, le triplet composé du genre, de la date de naissance et du code postal permet ainsi la ré-identification dans 87 % des cas (Sweeney, 2000). D'autres données, telles que l'historique des requêtes dans les moteurs de recherche ou les données de géolocalisation, se prêtent également à la ré-identification (Narayanan et al., 2016).

Dans un monde où le coût de l'accès à l'information tend vers zéro, l'objet rare n'est plus l'information mais bien l'attention. Le concept d'économie de l'attention a fait l'objet d'un effort de théorisation de la part d'Emmanuel Kessous (2012). Ce dernier décrit la transition d'un marketing de segmentation vers un marketing des traces renforçant l'emprise des offreurs sur les consommateurs en l'absence d'un contrôle fort des données à caractère personnel par les individus. Ce constat d'asymétrie des forces entre les entreprises du numérique, en particulier les GAFAM, et les utilisateurs de services numériques a motivé l'Union européenne à accroître la protection des citoyens européens grâce au « *Règlement général de protection des données* » (RGPD), publié le 27 avril 2016, d'application depuis le 25 mai 2018 (Banck, 2018). Le RGPD définit le concept de « *donnée à caractère*

personnel » comme « toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée « personne concernée »); est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale », ce qui recouvre à la fois les identifiants déterministes et les identifiants probabilistes. Le RGPD est « neutre sur le plan technologique », ce qui signifie notamment qu'il s'applique à tout type de traceur.

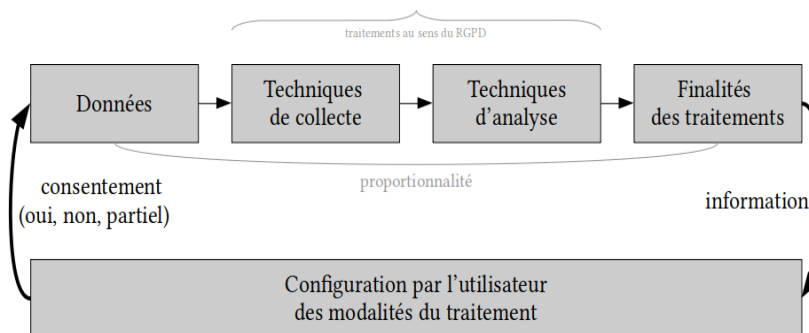


Figure 1. Fonctionnement du RGPD.

Le RGPD définit le concept de « traitement » de manière large comme « toute opération ou tout ensemble d'opérations effectuées ou non à l'aide de procédés automatisés et appliquées à des données ou des ensembles de données à caractère personnel, telles que la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction » (cf. Figure 1). Le RGPD exige que le consentement, libre et éclairé, se manifeste par un acte positif de la part de l'utilisateur. Il impose par ailleurs que les données collectées soient cohérentes au regard des finalités annoncées (principe de proportionnalité). Sur cette base, l'utilisateur peut refuser ou accepter, en tout ou partie, les collectes sollicitées. Les violations à ce règlement sont instruites par des agences nationales de protection de données telles que la [CNIL](#) en France.

La question du recueil du consentement est cruciale pour les entreprises actives dans la publicité ciblée et dans l'analyse de performances dès lors qu'elles recourent à des données qui ne sont pas anonymisées. En effet, si l'on s'en tient aux *cookies*, ces derniers se révèlent utiles, voire nécessaires, pour identifier un utilisateur (au travers de son navigateur), ce qui permet, premièrement, de limiter son exposition à une campagne de publicité (principe du *capping*), deuxièmement, de cibler les publicités qui lui sont envoyées, soit par le suivi d'une session, soit par le biais de son identification puis de son rattachement à un profil, troisièmement, de dresser des statistiques de fréquentation fiables (p. ex. comptabilisation des visiteurs uniques) (Viseur, 2021). Le RGPD a donc deux conséquences pour ces entreprises. D'une

part, elles doivent se mettre en conformité avec les dispositions du règlement, d'autre part, elles sont exposées au refus de consentir au traitement des données par les utilisateurs. En ont résulté plusieurs conséquences. Premièrement, les entreprises ont dû veiller à centraliser les demandes de consentement de manière à éviter leur multiplication, dès lors réduire le risque de refus. En particulier, le secteur de la publicité programmatique est composé d'une myriade d'acteurs réconciliant leurs données par le biais de techniques comme le *cookie syncing* (Papadopoulos et al., 2019). Deuxièmement, des acteurs spécialisés sont apparus pour gérer les interfaces de recueil de consentement en proposant des outils standards : les CMP (*Consent Management Platform*) (Nouwens et al., 2020). Troisièmement, et cela pré-existait au RGPD, le recours aux *dark patterns* s'est multiplié dans les interfaces de recueil de consentement.

Tableau 1. Typologie de *dark patterns* (basé sur Gray et al., 2018).

Nom	Description
Harcèlement	Appliquer une redirection des fonctionnalités attendues qui persiste au-delà d'une ou plusieurs interactions.
Obstruction	Rendre un processus plus compliqué que nécessaire de sorte à dissuader certaines actions.
Sournoiserie	Tenter de cacher, travestir ou retarder la divulgation d'une information importante pour l'utilisateur.
Interférence d'interface	Manipuler l'interface utilisateur de manière à favoriser certaines actions au détriment d'autres actions.
Action forcée	Contraindre l'utilisateur à réaliser certaines actions pour accéder (ou continuer à accéder) à certaines fonctionnalités.

Le terme « *dark pattern* » désigne la situation où un designer utilise sa connaissance du comportement humain (p. ex. psychologie) et les désirs des utilisateurs finaux pour mettre en œuvre des fonctionnalités trompeuses qui ne sont pas dans l'intérêt de l'utilisateur (Gray et al., 2018). En ce sens, il s'inscrit dans la logique des *sludges* définis par Thaler (2018) par opposition aux *nudges* (Thaler et Sunstein, 2010) supposés, dans la logique du paternalisme libertaire, aider les citoyens à faire les meilleurs choix pour eux-mêmes sans les priver de pouvoir faire s'ils le souhaitent un choix alternatif. Gray et al. (2018) proposent une typologie de *dark patterns* (cf. Tableau 1) incluant le harcèlement (*nagging*), l'obstruction (*obstruction*), la sournoiserie (*sneaking*), l'interférence d'interface (*interface interference*) et l'action forcée (*forced action*). Les exemples fournis couvrent notamment le partage de données à l'image du *privacy zuckering* par lequel l'utilisateur est amené à partager plus d'informations que nécessaire.

3. Méthodologie

Cette analyse préliminaire des caractères légal et éthique des interfaces de recueil de consentement et des conditions générales d'utilisation précisant les données collectées ainsi que les finalités s'appuie, d'une part, sur une première sélection de cas identifiés par l'auteur, d'autre part, sur des exemples documentés par le site [Pixel de tracking](#) et le compte Twitter [Pixel de Tracking](#) qui l'accompagne. Cette

sélection a permis d'associer un ensemble de pratiques courantes à la typologie de *dark patterns* proposée par Gray et al. (2018). Au total, 34 *dark patterns* ont été identifiés et repris dans une grille d'analyse (sous LibreOffice.org Calc).

Différents outils ont été exploités pour approfondir cette analyse. D'une part, l'outil « *Outils de développement web* » de Firefox a été utilisé pour valider le *dark pattern* de sournosierie (*sneaking*) consistant à collecter des données avant le recueil du consentement. D'autre part, un script Python, permettant d'obtenir automatiquement une capture d'écran de la page d'accueil d'une liste de sites avec l'affichage du CMP, a été développé, et testé, offrant des perspectives d'automatisation des analyses par pays ou par secteur. Ce programme a notamment été utilisé sur les sites des principaux journaux belges (dont le CMP est généralement fourni par [Didomi](#)) et français, soit un total de 32 sites web.

4. Résultats

Nous proposons dans cette section, d'une part, de catégoriser les *dark patterns* appliqués au CMP (*Consent Management Platform*) et aux CGU (Conditions Générales d'Utilisation), d'autre part, d'approfondir leur mise en œuvre par les sites Web.

4.1. Catégorisation des *dark patterns*

Sur base de la typologie de *dark patterns* proposée par Gray et al. (2018), nous proposons de catégoriser les *dark patterns* observés, d'une part sur les CMP, d'autre part sur les CGU. Nous utiliserons les noms francophones pour désigner ces techniques.

Tableau 2. Application de *dark patterns*, par type, aux CMP et CGU.

Technique	CMP	CGU
Harcèlement	Demande récurrente d'autorisation (p. ex. géolocalisation).	Validation récurrente des CGU modifiées d'un service.
Obstruction	Principe des <i>cookie walls</i> allégés (p. ex. multiplication des <i>sliders</i> sans refus global).	Page interminable centralisant les CGU de tous les services proposés par une même firme.
Sournosierie	Collecte de données même en cas de refus. Difficulté de retrait du consentement (principe du <i>roach hotel</i>).	Texte très long ¹ noyant l'information, ou écrit dans un langage volontairement cryptique.
Interférence d'interface	Cases de consentement pré-cochées. Mise en évidence du bouton pour accepter.	Masquage des clauses les plus problématiques (p. ex. utilisation des données d'un formulaire à des fins

¹ La longueur excessive des CGUs des réseaux sociaux a été mise à l'honneur par l'artiste [Dima Yarovinski](#) dans l'exposition « I agree ». La durée de lecture des CGUs d'Instagram y était ainsi évaluée à 1 heure 30 environ (cf. [\[url\]](#)). McDonald et Cranor (2008) ont pour leur part évalué le temps annuel de lecture des politiques de confidentialité à 244 heures.

Technique	CMP	CGU
	Moindre visibilité du bouton pour refuser. Multiplication des étapes pour refuser.	commerciales).
Action forcée	Principe des <i>cookie walls</i> (accès conditionné à l'acceptation ou au paiement). Fourniture de la date de naissance pour valider une limite d'âge (principe du <i>privacy zuckering</i>).	<i>na</i>

4.2. Analyse des dark patterns

4.2.3. Dans le cas des CMPs

Le harcèlement paraît légal puisque l'utilisateur garde la possibilité de donner ou non son consentement. Il n'est par contre pas éthique dès lors qu'il tente de l'obtenir « à l'usure » et est par ailleurs contraire aux recommandations de la CNIL (délibération [n°2020-092](#)) qui prévoient de conserver les choix « pendant un certain laps de temps » (recommandation : 6 mois) en fonction de la nature du site ou de l'application. L'obstruction, la sournoiserie et l'action forcée débouchent généralement sur des dispositifs à la fois contraires à l'éthique et, souvent, à la loi, par exemple au titre de l'absence de consentement (p. ex. collecte avant acceptation) ou de la non-proportionnalité des données collectées au regard des finalités (p. ex. *privacy zuckering*), si l'on s'en tient aux exemples fournis dans le Tableau 2.

Le principe, largement répandu, des *cookie walls* (p. ex. refus du site d'accéder aux contenus sans acceptation de la collecte ni paiement d'une contrepartie financière), soit un cas d'obstruction ou d'action forcée (suivant le degré de blocage), se révèle contraire à la liberté du consentement (avis du Comité européen de la protection des données) mais ne doit pas être systématiquement interdite, par exemple dès lors qu'un accès à une version minimale du contenu est offerte (décision du Conseil d'État ; cf. [url](#)).

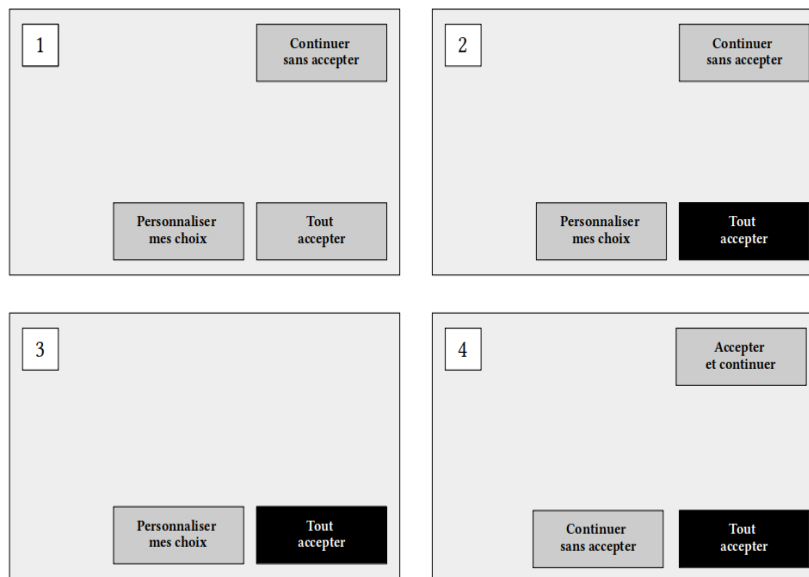


Figure 2. Dark patterns de type interférence d'interface (2, 3, 4).

Les interférences d'interface donnent lieu à une mise en œuvre plus diversifiée (cf. Figure 2). L'affichage de cases pré-cochées est clairement considéré comme illégal (cf. [url]). L'illustration n°1 représente un exemple d'interface de CMP conforme aux recommandations de la CNIL. Les illustrations n°2, 3 et 4 représentent des variations couramment observées. L'illustration n°2 consiste à mettre en évidence un choix qui arrange l'éditeur du site web car conduisant à l'acceptation de tous les traceurs. La facilité pour accepter ou refuser les traceurs est cependant équivalente. Elle s'éloigne par contre de la recommandation de la CNIL de ne pas mettre visuellement un choix davantage en évidence qu'un autre. L'illustration n°4 se révèle particulièrement vicieuse, à défaut d'être illégale, puisqu'elle inverse la position couramment utilisée pour les boutons d'acceptation totale et de refus afin d'induire en erreur l'utilisateur. Ce *dark pattern*, observé notamment sur le site du magazine Marianne, en a par la suite disparu, une conséquence possible de la bronca déclenchée sur les réseaux sociaux par cette découverte.

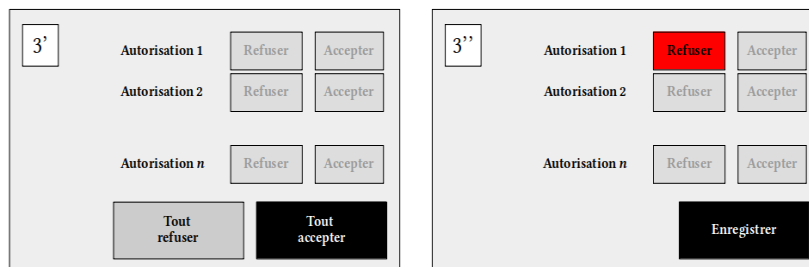


Figure 3. Dark patterns de type interférence d'interface (3', 3'').

L'illustration n°3 (cf. Figure 2) viole clairement les recommandations de la CNIL en ce sens qu'elle débouche typiquement sur l'enchaînement d'écrans illustré à la Figure 2 consistant à ne pas proposer de refus global et à multiplier les étapes pour personnaliser les choix par finalité. De plus, le consentement tend à être forcé en usant d'artifices visuels en proposant une seconde fois un bouton d'acceptation globale (cf. Figure 3 ; illustration n°3') ne disparaissant qu'à la configuration des finalités (cf. Figure 3 ; illustration n°3''). En février 2023, ce type d'interface restait par exemple majoritaire sur les sites de la presse belge (à l'exception du site de Roularta Media Group).

4.2.3. Dans le cas des CGUs

À défaut d'être clairement illégal, le harcèlement pose un problème éthique, puisqu'il tente d'obtenir un consentement « à l'usure ». Cette pratique se traduit par une modification fréquente des termes du contrat, avec une demande de validation à la clef. L'obstruction (p. ex. longueur d'un contrat) et la sournoiserie (p. ex. complexité excessive d'un contrat et morcellement entre plusieurs documents) posent ici un problème de conformité au règlement puisqu'ils touchent directement au caractère éclairé du consentement.



Figure 4. Dark patterns de type interférence d'interface (CGU).

Les interférences d'interface dans les CGU donnent lieu à une mise en œuvre particulièrement insidieuse (cf. Figure 4). Si la fermeture de sections spécifiques peut se justifier par des raisons de lisibilité (illustration n°2), le masquage pur et simple sans moyen simple d'identifier les sections rendues visibles dans un second temps (illustration n°1) pose clairement un problème éthique à défaut d'être une pratique illégale (car le lecteur garde la possibilité de déplier ces sections avant sa lecture... ce qu'il ne fera généralement pas dans la mesure où le lien d'affichage se révèle discret). Reste que, dans les deux cas, ces artifices sont généralement un moyen de masquer des clauses problématiques (p. ex. réutilisation des données encodées à des fins marketing dans un outil de formulaires en ligne).

5. Discussion

Nous discuterons dans cette section de l'éthique des *dark patterns*, de la dépendance à la publicité ciblée, de la régulation des pratiques ainsi que des limitations et des perspectives.

5.1. Éthique des dark patterns

Trois types de pratiques pourraient être distinguées : des pratiques clairement légales, des pratiques clairement illégales (p. ex. sournoiserie) et des pratiques légales mais peu ou prou éthiques, dont certaines en sursis (p. ex. harcèlement) du fait des évolutions de la jurisprudence et des publications des APDs (Autorités de Protection des Données), soit des lignes directrices, soit des recommandations.

Certaines pratiques comme le harcèlement (p. ex. demande répétée de consentement) et l'obstruction (p. ex. *cookie walls*) se révèlent menacées, en particulier suite à la publication de recommandations par la CNIL (cf. [délibération n°2020-092](#) du 17 septembre 2020). Ainsi le harcèlement est explicitement ciblé par la CNIL : « *De manière générale, la Commission recommande que le choix exprimé par les utilisateurs, qu'il s'agisse d'un consentement ou d'un refus, soit enregistré de manière à ne pas les solliciter à nouveau pendant un certain laps de temps* » (Délibération n° 2020-092 du 17 septembre 2020). Idem pour la sournoiserie sous la forme d'une entrave au retrait de consentement : « *Les utilisateurs ayant donné leur consentement à l'utilisation de traceurs doivent être en mesure de le retirer à tout moment. La Commission rappelle qu'il doit être aussi simple de retirer son consentement que de le donner* » (Délibération n° 2020-092 du 17 septembre 2020).

La légalité du principe des *cookie walls* au sens strict semble également en sursis et, dans l'attente d'une clarification (p. ex. futur règlement *eprivacy*) et d'une homogénéisation des décisions, se juge au cas par cas. Au niveau européen : « *L'EDPB a constaté la nécessité de nouvelles précisions, en particulier en ce qui concerne la validité du consentement fourni par la personne concernée lorsqu'elle interagit avec un « accès subordonné à l'acceptation de cookies » ou « cookie walls »* » (cf. [url](#)). Au niveau français : « *Par la décision du 19 juin 2020, le Conseil d'État a jugé que l'exigence d'un consentement « libre » ne pouvait toutefois pas justifier une interdiction générale de la pratique des « murs de traceurs » : la liberté du consentement des personnes doit être appréciée au cas par cas, en tenant compte notamment de l'existence d'alternative réelle et satisfaisante proposée en cas de refus des cookies* » (cf. [url](#)). Au niveau belge : « *Ne conditionnez pas la fourniture de vos produits ou services (même gratuits) à l'acceptation du traitement de données à caractère personnel non-nécessaires à la prestation du service ou à la fourniture du produit. N'essayez pas de forcer ou d'inciter, de quelque manière que ce soit, les personnes concernées, à vous fournir leur consentement à ces traitements* » (cf. [url](#), page 61).

Les pratiques observées évoluent parfois rapidement au gré des recadrages des autorités ou des dénonciations sur les réseaux sociaux. Certains médias (p. ex. Le Monde, Le Figaro et L'Équipe) ont ainsi persévéré dans l'utilisation du *cookie wall*. D'autres se sont adaptés (p. ex. Libération et 20 Minutes) : après avoir sollicité le paiement d'un abonnement, ces sites de journaux français permettent ensuite de consulter leur site avec un bandeau de rappel en bas de page, soit une succession de deux *dark patterns* (obstruction puis harcèlement). Cette modalité est (peut-être) elle-même en sursis : « *De plus, la Commission recommande que, lorsque le refus peut être manifesté par la poursuite de la navigation, le message sollicitant le consentement (par exemple, la fenêtre ou le bandeau) disparaisse au bout d'un laps de temps court, de manière à ne pas gêner l'utilisation du site ou de l'application et à ne pas, ainsi, conditionner le confort de navigation de l'utilisateur à l'expression de son consentement au traceur* » (cf. [url](#), page 10). Reste que la CNIL a très

récemment adouci sa position en prenant en compte, d'une part, la nécessité d'« *une juste rémunération* », par exemple via la publicité ou via un abonnement, et, d'autre part, l'acceptabilité du « *cookie wall* » dès lors qu'il existe « *une alternative réelle et équitable* » et qu'il est limité aux finalités permettant cette juste rémunération (cf. [url]). Signalons également, par exemple chez L'Équipe et Le Monde, la présence d'un *sludge* sous la forme d'une réduction sur le prix de l'abonnement en cas de connexion via un compte Google.

L'éthique dans le contexte du marketing se réfère à un ensemble de principes et de normes morales qui régissent les pratiques commerciales et publicitaires. Nantel et Weeks (1996) distinguent deux approches de l'éthique en marketing. La première, la plus ancienne, est utilitariste et se concentre sur la satisfaction du client. La seconde, plus récente, privilégie la déontologie. Sont dès lors mis en avant, par exemple, au sein de codes d'éthiques, le respect de l'esprit et de la lettre des législations, une présentation honnête des caractéristiques des produits vendus ainsi que le bannissement de toute pratique de vente et de publicité tendancieuse ou trompeuse. Si l'on se place sur un plan purement utilitariste, certaines pratiques d'obstruction ou d'action forcée peuvent se justifier dès lors qu'elles tendent à préserver un accès équitable à des services en ligne (satisfaction du client) ; sur un plan déontologique, par contre, ces pratiques ne respectent pas l'esprit porté par le RGPD. Quelle que soit l'approche éthique retenue, les *dark patterns* de harcèlement ou de sournoiserie paraissent difficilement justifiables tout en étant de plus en plus critiqués sur un plan légal. Quant aux interférences d'interface, leur éthique est contestable mais leur condamnation paraît parfois difficile même si certaines d'entre elles s'apparentent à la sournoiserie (p. ex. dissimulation de texte au sein des CGUs).

Tableau 3. Caractère éthique ou légal des *dark patterns* appliqués aux CMPs ou aux CGUs.

Technique	Éthique	Légalité
Harcèlement	Non car tentative de forcer le consentement « à l'usure ».	Oui mais en sursis du côté des APDs (cf. recommandations).
Obstruction	Non (déontologie) car volonté d'épuiser l'utilisateur mais discutable (utilitarisme) dès lors qu'elle concerne l'obtention d'une juste rémunération.	Non car condamné (p. ex. absence de bouton pour tout refuser) sauf pour les <i>cookies walls</i> (en discussion) ; oui pour les CGUs (p. ex. longueur).
Sournoiserie	Non car volonté d'obtenir le consentement en trompant l'utilisateur.	Non pour la sournoiserie appliquée aux CMPs mais oui pour de nombreuses techniques appliquées aux CGUs (p. ex. longueur et complexité).
Interférence d'interface	Non car généralement révélatrice d'une volonté de piéger l'utilisateur, en particulier quand le <i>dark pattern</i> tend vers la sournoiserie (cf. Figure 3 /4 par exemple).	Non pour un large ensemble de techniques (p. ex. cases pré-cochées et multiplication des étapes pour refuser) mais oui pour certaines techniques plus insidieuses (p. ex. déplacement de boutons standards).
Action forcée	Idem que l'obstruction.	Non observé.

Cette zone grise, composée de dispositifs en pratique tolérés (cf. Tableau 3), permet aux entreprises (p. ex. régies publicitaires) de « jouer la montre » sur base d'un calcul bénéfice-risque (probabilité et gravité d'une sanction). Si l'on assiste à un assainissement des pratiques, des dispositifs illégaux persistent cependant (p. ex. sournoiserie : collecter des données à caractère personnel sans avoir le consentement explicite des usagers). Par ailleurs, des dispositifs peu ou prou éthiques existent. Ils sont, soit en sursis (p. ex. harcèlement : demandes répétées de consentement), soit difficiles à réguler sauf à imposer des modèles d'interfaces (p. ex. interférences visuelles), soit tolérés par les APDs (sous conditions) (p. ex. obstruction : consentir ou payer).

5.2. Régulation des pratiques

Les producteurs de contenus en ligne sont aujourd'hui largement dépendants des revenus issus de la publicité, en particulier de la publicité ciblée. C'est par exemple le cas des nouveaux médias émergents sur des plates-formes telles que Youtube (Cauche, 2019). En résulte une dépendance aux GAFAM, et en particulier à Google, qui dominent le marché de la publicité en ligne (avec une part de marché cumulée sur le marché français d'environ 75 % ; Viseur, 2021). Certes, il existe des modes de rémunération alternatifs : les abonnements ou les dons récurrents (Bessière et al., 2017) par exemple. Cependant, ils semblent rester insuffisants pour remplacer totalement les revenus issus de la publicité. Les systèmes de *paywalls* se révèlent ainsi incapables de stimuler significativement les revenus de la presse (Myllylahti, 2014). Les médias recourent en outre à des *widgets* valorisés par la collecte de données (p. ex. AddThis ; Viseur, 2021) et incluent, sans doute par facilité, des objets également avides de données (p. ex. vidéos Youtube).

Cette dépendance aux données à caractère personnel, directe pour les régies publicitaires, indirecte pour les producteurs de contenus, eux-mêmes dépendants des régies publicitaires, explique l'échec des mécanismes d'*opt-in* comme le champs d'en-tête HTTP DNT *Do Not Track* (cf. [url1] et [url2] pour plus d'informations). Sont dès lors difficiles à mettre en œuvre en l'état des mécanismes de traitement automatisé des demandes de consentement (sur base d'un format structuré et de profils utilisateurs standards) par les navigateurs comme évoqué par Nouwens et al. (2020). Face à cette situation, des entreprises comme Apple se sont positionnées, avec une certaine crédibilité, comme défenseurs de la vie privée de leurs clients, en mettant en œuvre un mécanisme de filtrage de *cookies tiers* (ITP) dont le principe s'est ensuite étendu à Firefox (Viseur, 2021).

Découlent donc de cette dépendance des pratiques, parfois agressives, souvent légales mais à l'éthique discutable (*sludge*), pour obtenir le consentement au traitement de données personnelles. La complexité inhérente à l'octroi d'un consentement libre et éclairé s'oppose à l'idéal de gestion individuelle (Kröger et al., 2021). Les appels répétitifs et parfois vicieux (*dark patterns*) conduisent à la résignation de nombreux utilisateurs, ce que Solove (2020) a théorisé sous l'expression « *mythe du paradoxe de la vie privée* ». Le concept de « *privacy paradox* », c'est-à-dire le décalage entre l'intention et le comportement de divulgation de données à caractère personnel, a été introduit par Norberg et al. (2007). Si Waldman (2020) l'explique notamment par la rationalité limitée des consommateurs, Solove (2020) y voit au contraire un comportement rationnel. Si les utilisateurs valorisent leur vie privée mais agissent au contraire de leurs intérêts, ce

n'est pas la démonstration d'une faible valeur accordée, en réalité, à la vie privée mais bien un comportement rationnel de résignation face à l'arsenal de techniques insidieuses mises en œuvres pour extorquer leur consentement, soit un constat de faible contrôlabilité de ces dispositifs malgré le mécanisme légal de consentement.

Dans ce contexte, les associations et les collectifs ont un rôle qui devrait être mis en valeur. Premièrement, ils permettent la mise en évidence et l'objectivisation de comportements contre lesquels des logiciels ou des réglementations devraient pouvoir lutter. Ce sont par exemple [Open Terms Archive](#), pour la traçabilité des modifications de CGU, [TOS:DR](#) pour leur évaluation sur une échelle standardisée, [Data Experience](#) d'[Hestia Labs](#), pour la mise en évidence des critères de profilage, ou encore l'extension [Privacy Badger](#), de l'[EFF](#), pour le filtrage des traceurs. Deuxièmement, ils peuvent agir publiquement en dénonçant les comportements contraires à l'éthique, en exploitant la volonté des entreprises de conserver leur réputation en ligne. Troisièmement, ils agissent comme des intermédiaires entre les utilisateurs et les APD nationales pour construire les dossiers visant à lutter contre les infractions au RGPD (p. ex. [La Quadrature du Net](#) en France et la coupole européenne [EDRI](#)). Quatrièmement, ils pourraient, à la manière de la [FSF](#) pour les licences logicielles ou de la [Creative Commons](#) pour la culture libre, contribuer à la construction de contrats standardisés (Wylie, 2019), stables et compréhensibles, permettant dès lors l'octroi d'un consentement réellement éclairé.

Enfin, la complexité supplémentaire amenée par le RGPD a conduit à l'émergence de prestataires spécialisés dans la mise en œuvre de *Consent Management Platforms* (CMP). Ces entreprises sont par exemple : [Azeptio](#), [ConsentManager](#), [Cookiebot](#), [Didomi](#), [SFBX](#) et [Sirdata](#). Cette concentration accrue des interfaces de recueil de consentement entre les mains de quelques prestataires offre donc aux APDs un moyen d'action sur des acteurs ayant un impact significatif sur les pratiques auxquelles sont confrontés les utilisateurs.

Tableau 4. Protection des utilisateurs contre la collecte abusive de données.

	Avant	Actuellement	Futur possible
Macro	Lois nationales (<i>privacy</i>).	RGPD (EU). APDs.	Cadre juridique mondial émergent (RGPD, CCPA...). APDs (renforcement).
Meso	Modèle publicitaire dominant.	Développement d'une niche <i>pro-privacy</i> .	Régulation des GAFAM et des prestataires CMP. Soutien aux initiatives de standardisation et de labellisation (p. ex. D-Seal). Relais via des associations professionnelles (p. ex. IAB). Soutien aux associations <i>pro-privacy</i> .
Micro	Technologies anti-tracking (anonymisation, extensions...).	Technologies anti-tracking (anonymisation, extensions...). Consentement libre et éclairé.	Technologies anti-tracking (anonymisation, extensions...). Consentement libre et éclairé.

En pratique, s'appuyer sur le consentement libre et éclairé ne permet pas une protection optimale des utilisateurs face à la collecte abusive de données (cf. Tableau 4). En effet, le respect du RGPD s'accompagne de *dark patterns* légaux à défaut d'être éthiques. La recherche d'un impact maximal sur les pratiques de terrain peut dès lors passer par la régulation des acteurs conditionnant les pratiques d'un grand nombre d'acteurs (p. ex. GAFAM et CMP) ainsi que par le relais négocié via les associations professionnelles (p. ex. [IAB](#) ; car recommandations aux professionnels via le [TCF](#) : *Transparency & Consent Framework*). Davantage de visibilité pourrait également être accordée aux initiatives de standardisation et de labellisation (p. ex. [D-Seal](#)) (Schade, 2023). De plus, les collectifs et associations *pro-privacy* se révèlent des soutiens utiles pour les APD en aidant à la construction des dossiers en violation du RGPD.

5.3. Limitations et perspectives

Cette recherche exploratoire présente trois perspectives principales. Premièrement, la problématique des *dark patterns* a fait l'objet d'une analyse approfondie par l'EDPB (*European Data Protection Board*) dans un document, postérieur à cette recherche, daté du 14 mars 2022 ([EDPB, 2022](#)). Les *dark patterns* y sont assimilés à « *des interfaces et des expériences utilisateurs qui amènent les utilisateurs à prendre des décisions, involontaires et potentiellement préjudiciables, concernant le traitement de leurs données personnelles* » (page 2). Les catégories proposées diffèrent de Gray et al. (2018) et se concentrent davantage sur les interférences graphiques. Les catégories proposées sont la surcharge (comparable au harcèlement) ; l'omission, la confusion et la dissimulation (assimilables à des formes d'interférence d'interface) ; et enfin l'entrave (recouvrant la sournoiserie et l'action forcée). Cette classification devra probablement être prise comme référence et être suivie de clarifications de la part de l'[EDPB](#) ainsi que des autorités nationales. L'adaptation de cette recherche à la typologie européenne se révélera dès lors sans doute indispensable. Deuxièmement, et compte tenu de la rapidité de l'évolution du design des CMPs, notamment suite à leurs adaptations au contexte réglementaire, il serait intéressant de sauvegarder ces interfaces sur une base périodique à l'aide d'un robot (Python). De la sorte, il serait possible de déterminer leur positionnement (conformité plus ou moins importante à l'esprit du règlement), d'analyser leur évolution en relation avec les règlements, jurisprudences, lignes directrices et recommandations publiés au fil du temps ainsi que les actions de la société civile pour en dénoncer les pratiques. Troisièmement, l'utilisation d'outils d'*eye tracking*, couplée à des entretiens semi-directifs, permettrait de mieux comprendre la réaction des utilisateurs face à ces *dark patterns*, pour ensuite en évaluer l'efficacité et émettre des recommandations quant aux contre-mesures appropriées.

6. Conclusion

Dans cette recherche exploratoire, nous avons présenté l'évolution du marketing en ligne vers une approche plus ciblée occasionnant une collecte massive de données personnelles. Nous avons ensuite montré comment la question du recueil du consentement, cruciale pour les régies publicitaires et les courtiers en données (donc aussi pour la presse en ligne), avait motivé l'utilisation de *dark patterns* pour « extorquer » le consentement des utilisateurs. Sur base d'une typologie de *dark patterns*, nous avons alors analysé, sur un ensemble de CMPs et de CGUs, comment

ces *dark patterns* s'appliquaient concrètement au recueil de consentement. Cette analyse nous a permis de discuter les limites actuelles du RGPD, coexistant avec des dispositifs visant à forcer le recueil du consentement (*sludge*). Nous avons donc discuté leurs caractères légal et éthique (du point de vue utilitariste et déontologique). Cela nous a permis de montrer l'existence d'une zone grise exploitée lucrativement par les professionnels. Les moyens d'actions disponibles pour pallier ces limitations ont enfin été développés.

7. Références

- Allary J., & Balusseau V. (2018). La publicité à l'heure de la data. Adtech et programmation expliqués par des experts, Dunod.
- Banck A. (2018). RGPD : la protection des données à caractère personnel, Gualino.
- Bessière, V., & Stéphany, É. (2017). Le crowdfunding: fondements et pratiques. De Boeck Supérieur.
- Cauche, R. (2019). Professionnalisation des modes de diffusion sur YouTube: pour une exploration des outils de mise en ligne. Mise au point. Cahiers de l'association française des enseignants et chercheurs en cinéma et audiovisuel, (12).
- EDPB (2022). Dark patterns in social media platform interfaces: How to recognise and avoid them. Guidelines 3/2022. 14 mars 2022. European Data Protection Board.
- Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-14).
- Hils, M., Woods, D. W., & Böhme, R. (2020). Measuring the emergence of consent management on the web. In Proceedings of the ACM Internet Measurement Conference (pp. 317-332).
- Kessous E. (2012). L'attention au monde. Sociologie des données personnelles à l'ère numérique, Armand Colin.
- Kröger, J. L., Lutz, O. H. M., & Ullrich, S. (2021). The myth of individual control: Mapping the limitations of privacy self-management. Available at SSRN 3881776.
- McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. ISJLP, Vol. 4:3, 543-568.
- Mesguish V. & Thomas A. (2013). Net recherche 2013. De Boeck.
- Myllylahti, M. (2014). Newspaper paywalls—the hype and the reality: A study of how paid news content impacts on media corporation revenues. Digital journalism, 2(2), 179-194.
- Nantel, J., & Weeks, W. A. (1996). Marketing ethics: is there more to it than the utilitarian approach?. European journal of marketing, 30(5), 9-19.
- Narayanan, A., Huey, J., & Felten, E. W. (2016). A precautionary approach to big data privacy. Data protection on the move: Current developments in ICT and privacy/data protection, 357-385.
- Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. Journal of consumer affairs, 41(1), 100-126.
- Nouwens, M., Liccardi, I., Veale, M., Karger, D., & Kagal, L. (2020). Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-13).

- Papadopoulos, P., Kourtellis, N., & Markatos, E. (2019). Cookie synchronization: Everything you always wanted to know but were afraid to ask. In *The World Wide Web Conference* (pp. 1432-1442).
- Peyrat, B. (2009). *La publicité ciblée en ligne*. Rapport, CNIL (Commission Nationale de l'Information et des Libertés).
- Schade, F. (2023). Dark Sides of Data Transparency: Organized Immaturity After GDPR?. *Business Ethics Quarterly*, 1-29.
- Solove, D. J. (2021). The myth of the privacy paradox. *Geo. Wash. L. Rev.*, 89, 1.
- Sweeney L. (2000). *Simple Demographics Often Identify People Uniquely*. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.
- Thaler, R. H. (2018). Nudge, not sludge. *Science*, 361(6401), p. 431.
- Thaler, R., Sunstein, C. (2010). *Nudge. Comment inspirer la bonne décision*. Vuibert.
- Viseur, R. (2021). Du tracking, des contre-mesures et de leur efficacité dans la publicité ciblée. *Revue ouverte d'ingénierie des systèmes d'information*, vol. 2, n°1.
- Waldman, A. E. (2020). Cognitive biases, dark patterns, and the 'privacy paradox'. *Current opinion in psychology*, 31, 105-109.
- Wylie, C. (2019). *Mindfuck: Le complot Cambridge Analytica pour s'emparer de nos cerveaux*. Grasset.
- Zuboff S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.

Proposition d'une architecture utilisant le *trace clustering* pour recommander un parcours d'apprentissage

Wiem Hachicha^{1,2}, Leila Ghorbel¹, Ronan Champagnat²,
Mourad Rabah², Samuel Nowakowski³, Corinne Amel Zayani¹

1. MIRACL - Université de Sfax
Tunis Road Km 10 PB. 242
3021 Sfax, Tunisie
corinne.zayani@fss.usf.tn

2. L3i - Université de La Rochelle
Avenue Michel Crépeau
17 042 La Rochelle, France
nom.prenom@univ-lr.fr

3. LORIA - Université de Lorraine
Campus Scientifique, 615 rue du jardin-botanique
54 506 Vandœuvre-lès-Nancy, France
samuel.nowakowski@loria.fr

RÉSUMÉ. Les systèmes d'informations pédagogiques permettent d'observer les traces d'apprentissage des apprenants et de mener des analyses sur leurs pratiques ou de prédire leur réussite. Dans ces travaux nous étudions comment la fouille de processus, qui permet d'extraire des modèles de comportement des utilisateurs dans un système d'information, peut être utilisée dans un système de recommandation contextuel. Nous nous concentrons plus particulièrement sur le *trace clustering* qui vise à regrouper des traces possédant des dynamiques proches. Notre apport porte sur la définition d'une architecture pour la recommandation qui utilise le *trace clustering*. Nous validons notre proposition sur les données collectées d'un cours d'introduction à la programmation d'IHM.

ABSTRACT. Educational information systems make it possible to observe learners' learning traces and to carry out analyses of their behaviour. In this work, we study how Process Mining can be used in contextual recommender systems. Our contribution concerns the definition of an architecture for recommendation that uses trace clustering. We validate our proposal on data collected from an introductory course in UI programming.

MOTS-CLÉS : SI pédagogique, fouille de processus, recommandation, trace clustering

KEYWORDS: Intelligent Tutor System, Process Mining, Recommendation, Trace Clustering

1. Introduction

Les systèmes d'information pédagogiques se sont rapidement développés avec la mise en place d'environnements numériques de travail. La quantité de données produites et de traces laissées par les utilisateurs de ces systèmes offre l'opportunité de fournir des tableaux de bord d'apprentissage et des analyses sur les apprenants (Cordier *et al.*, 2013).

La personnalisation des apprentissages est devenue un facteur essentiel pour la réussite des apprenants. Plusieurs projets se sont développés afin de traiter cette problématique. Ces projets s'attaquent au problème de l'offre de formation et de parcours d'apprentissage personnalisé, avec un accompagnement des étudiants dans leur projet de formation et d'insertion professionnelle. Toutefois, peu d'outils sont capables de fournir des indicateurs pertinents afin de recommander des parcours personnalisés ou des actions de remédiation pour les étudiants en difficulté.

La découverte des parcours d'apprentissage reste un défi à relever dans le domaine pédagogique. Toutefois, dans des domaines connexes, on trouve des travaux approchants. En particulier dans le domaine de la fouille de processus des recherches se sont focalisées, à partir de l'analyse des traces d'exécution réelles des utilisateurs (enregistrement des chemins de navigation), sur l'extraction de connaissances sur le parcours de l'utilisateur (Leblay *et al.*, 2018).

Le parcours d'apprentissage revient à sélectionner et ordonner les activités à réaliser pour acquérir des connaissances et compétences. Ce parcours d'apprentissage correspond à un scénario pédagogique suivi par un apprenant. Il peut être modélisé par un processus métier. Ce type de processus possède la particularité d'être faiblement structuré. C'est-à-dire que l'utilisateur dispose de degrés de liberté importants. L'objectif est alors de déterminer pour chaque utilisateur ou par groupe d'utilisateurs le comportement adopté lors de l'utilisation du système.

Notre objectif est d'étudier la possibilité de définir une méthodologie de recommandation basée sur les processus extraits des traces utilisateurs et d'implémenter l'architecture logicielle correspondante. Pour atteindre cet objectif, nous proposons de construire un modèle utilisant la fouille de processus à partir des observations recueillies lors des expériences des utilisateurs précédents avec le système d'information pédagogique. Ce modèle représente l'enchaînement des étapes et leurs impacts sur les états du processus global et sera par la suite utilisé pour recommander l'étape la plus appropriée pour guider l'utilisateur ou l'apprenant actuel. Un premier défi, contribution présentée dans le présent article, consiste à classer les différentes trajectoires que nous aurons identifiées au sein de systèmes d'activité sélectionnés dans un contexte d'apprentissage, afin de pouvoir ensuite déterminer, de manière automatisée, à quelle trajectoire type correspond le parcours et le développement d'un apprenant dans un environnement numérique dépourvu de processus métiers bien identifiés.

Notre approche suggère une voie corrective, si nécessaire. Dans (Ho *et al.*, 2016), les auteurs ont introduit une méthode pour amener les utilisateurs à prendre les bonnes

décisions en fonction de certaines informations extraites d'un modèle de données qui décrit les activités possibles à réaliser pour atteindre l'objectif et leur impact. Ce modèle de données est une entrée de la méthode, dans le sens où il est construit a priori par des experts du domaine. Dans notre cas, notre approche calcule le modèle de données à partir des informations disponibles avant de recommander l'étape suivante.

Dans la section suivante nous présentons le domaine de la fouille de processus. Nous décrivons le principe, les contraintes, les techniques de fouilles et les critères de qualité pour estimer la pertinence du modèle extrait. Nous terminons cette section par une présentation du *trace clustering*.

Nous faisons, ensuite, un état de l'art de l'utilisation de la fouille de processus dans le domaine de l'éducation. Puis nous proposons notre architecture qui s'appuie sur le *trace clustering* pour classer les parcours d'apprentissage. Nous validons cette architecture sur les données d'un cours universitaire de programmation des IHM.

2. Fouille de processus

La fouille de processus est utilisée dans de nombreux domaines, notamment pour la modélisation des comportements des utilisateurs qui recherchent des informations dans une bibliothèque numérique (Trabelsi *et al.*, 2019a ; 2019b), dans le domaine médical afin de cartographier les processus healthcare (Pika *et al.*, 2019), ou dans les réseaux sociaux pour modéliser les parcours utilisateurs (Li, De Carvalho, 2019)... L'objectif de la fouille de processus est de découvrir, superviser et améliorer des processus métier existants en extrayant de la connaissance à partir des journaux d'événements facilement disponibles dans les systèmes d'information actuels.

Chaque événement dans un tel journal fait référence à une activité (une étape bien définie d'un processus) et est lié à un cas particulier (une instance de processus). Les techniques de fouille de processus utilisent des informations supplémentaires telles que la ressource (une personne ou un équipement) qui exécute ou lance l'activité, l'horodatage de l'événement ou des éléments de données enregistrés avec l'événement (par exemple, la quantité commandée).

Les techniques de fouille de processus sont apparues au cours des années 1990. (Cook, Wolf, 1998) et (Agrawal *et al.*, 1998) ont proposé des algorithmes de découverte de modèle de processus afin de pouvoir analyser une organisation ou comparer des exécutions de processus métier à partir des traces observées dans le système d'information. (Van der Aalst, 2016) a popularisé la fouille de processus et développé de nombreux algorithmes.

Une première utilisation de la fouille de processus consiste à découvrir un modèle de processus. Un algorithme de découverte analyse un journal d'événements et construit un modèle à partir des relations de précédences entre activités. Parmi les algorithmes de découverte nous pouvons citer Inductive Miner, Fuzzy Miner, Heuristic Miner et Alpha Miner. La sortie de ces algorithmes est un modèle tel que les réseaux de Petri, BPMN...

Une deuxième utilisation de la fouille de processus est la vérification de la conformité. Il s'agit ici de comparer un modèle existant avec les journaux. L'objectif est de déterminer si la réalité enregistrée dans les journaux d'événements est conforme vis-à-vis du modèle et réciproquement (le modèle est conforme vis-à-vis de journaux observés).

2.1. Journaux d'événements

Un journal d'événements contient un ensemble d'événements. Chaque événement correspond à la réalisation d'une activité. Les informations minimales afin d'extraire des connaissances en utilisant la fouille de processus sont : un identifiant permettant de rattacher l'activité à un processus (CaseID), un identifiant de l'activité (Activity) et un horodatage (timestamp) (Van der Aalst, 2016).

Chaque variant d'un processus est décrit par une séquence d'activités (par exemple $\langle Activity1, Activity2, Activity1 \rangle$). L'ensemble du journal d'événements peut s'écrire sous la forme $L = [\langle Activity1, Activity2, Activity1 \rangle^2, \langle Activity1, Activity2, Activity3 \rangle^1, \langle Activity1, Activity1 \rangle^3]$ où la multiplicité des séquences donne le nombre de fois que ce variant est présent dans le journal.

2.2. Algorithmes de découverte

Les algorithmes de découverte visent à extraire des modèles de processus à partir des informations contenues dans les journaux d'événements. Ils se concentrent sur les aspects contrôle. Les techniques de découvertes sont variées, mais toutes sont bâties sur les relations entre les activités dans les journaux d'événements.

Les algorithmes observent en particulier les relations de causalité entre deux activités. Ils se basent sur l'intuition que si une activité se trouve toujours après une autre, c'est qu'il y a certainement une relation de causalité entre les deux. Entre deux activités, a_1 et a_2 , quatre relations sont considérées :

1. *Succession directe*, $a_1 > a_2$ s'il existe un variant tel que a_1 est immédiatement suivie par a_2 ;
2. *Causalité*, $a_1 \rightarrow a_2$, si $a_1 > a_2$ et $a_2 \not> a_1$;
3. *Parallèle*, $a_1 \parallel a_2$, si $a_1 > a_2$ et $a_2 > a_1$;
4. *Choix*, $a_1 \# a_2$, si $a_1 \not> a_2$ et $a_2 \not> a_1$.

L'algorithme α (Aalst *et al.*, 2004) construit un raisonnement sur ces quatre relations pour déduire un réseau de Petri qui modélise les processus. D'autres algorithmes, tel *Heuristic Miner* (Weijters, Ribeiro, 2011) se basent sur le graphe des successions directes observées dans les journaux. *Heuristic Miner* vise à résoudre le problème de log contenant du bruit et obtenus à partir de processus faiblement structurés. Cet alo-

grithme utilise une métrique basée sur la fréquence afin de déterminer la confiance dans la relation de causalité entre deux activités, a_1 et a_2 calculée comme suit :

$$a_1 \Rightarrow a_2 = \left(\frac{|a_1 > a_2| - |a_2 > a_1|}{|a_1 > a_2| + |a_2 > a_1| + 1} \right)$$

Une autre stratégie de découverte, utilisée par *Fuzzy Miner* (Van der Aalst, 2016), consiste à travailler sur la visualisation d'un graphe en s'inspirant de ce qui est fait pour les cartes routières et en donnant des règles de zoom et d'agrégations de chemin en fonction de leur popularité.

2.3. Qualité des modèles découverts

L'objectif de la fouille de processus est de découvrir des modèles qui caractérisent la dynamique d'un système d'information à partir des traces. Plusieurs mesures ont été proposées afin de caractériser la qualité du modèle découvert. Ces mesures font appel aux notions d'*overfitting*, le fait qu'un modèle ne représente que les variants présents dans les logs (un nouveau variant ne sera pas représenté par le modèle), et d'*underfitting*, le fait qu'un modèle représente potentiellement beaucoup (trop) de variants.

Quatre mesures sont définies par la communauté (Buijs *et al.*, 2012) :

- *Fitness* : représente la capacité à expliquer les processus observés ;
- *Generalisation* : caractérise le fait que le modèle est capable de prendre en compte de nouveaux variants ;
- *Precision* : caractérise le fait que le modèle n'est pas trop général, c'est-à-dire qu'il ne prend pas en compte tous les variants ;
- *Simplicity* : caractérise la complexité et les spécificités du modèle.

Ces mesures visent à déterminer la capacité d'un modèle à expliquer et généraliser les dynamiques observées dans les journaux. Un bon modèle doit trouver un équilibre entre ces mesures.

Pour chaque mesure plusieurs critères ont été proposés. Nous utiliserons ceux définis par (Adriansyah, 2014) qui se basent sur la notion d'alignement entre une trace observée dans les log et une trace d'exécution du modèle. Pour chaque trace on totalise le nombre de transformations à faire pour passer de l'une à l'autre. La *Fitness* est calculée comme suit :

$$F(L, M) = 1 - \frac{\delta(\lambda_{opt}^M(L))}{\delta(\lambda_{worst}^M(L))}$$

Où δ est la fonction de coût, $\lambda_{worst}^M(L)$ est le cas le plus défavorable où il n'y a aucune synchronisation possible entre la trace et le modèle. $\lambda_{opt}^M(L)$ représente les coûts obtenus pour chaque alignement optimal.

La précision est calculée par :

$$P(T, M) = \frac{1}{|E|} \sum_{e \in E} \frac{en_T(e)}{en_M(e)}$$

Où E est l'ensemble des événements dans les logs T , A l'ensemble des activités, $en_T(e) \subseteq A$ est l'ensemble des activités présentes dans les traces et $en_M(e) \subseteq A$ l'ensemble des activités présentes dans le modèle.

2.4. Trace clustering

La fouille de processus est apparue dans le domaine des processus métiers. Ces processus sont généralement bien définis et cadrés au niveau du système d'information. Or, dans le cas de système d'information pédagogique, une partie du processus métier est entre les mains de l'apprenant. Nous avons donc des processus faiblement ou partiellement structurés. L'utilisation des algorithmes de découverte dans un tel contexte amène à des modèles complexes difficilement exploitables.

Les travaux de (Trabelsi *et al.*, 2021) montrent que l'utilisation du *trace clustering* aboutit à l'extraction de connaissance et permet d'identifier des dynamiques d'usages typiques en réduisant la complexité des modèles découverts. Le but du *trace clustering* est de regrouper des variants de scénarios d'apprentissage qui possèdent des caractéristiques, au niveau du processus décrit, similaires.

Le *trace clustering* consiste à regrouper les traces avant d'appliquer un algorithme de découverte (Diamantini *et al.*, 2016). Quatre approches ont été développées :

- le *trace-based clustering* regroupe les traces en fonction de leur similarité syntaxique. La mesure de similarité est inspirée de la distance de Levenshtein ;
- le *feature-based clustering* calcule un vecteur en fonction des caractéristiques de chaque trace. Parmi les caractéristiques couramment retenues, il y a la fréquence d'une activité dans un variant, la fréquence de succession directes, les sous-séquences maximales, les sous-séquences fréquentes...
- le *model-based clustering* qui réalise le regroupement sur les qualités des modèles minés, et
- l'*hybrid-based clustering* qui combine les approches précédentes (Song *et al.*, 2008 ; Zandkarimi *et al.*, 2020).

Nous venons de présenter le domaine de la fouille de processus en terminant par une introduction au *trace clustering*. Il s'agit de méthodes visant à regrouper des traces d'exécutions qui décrivent des comportements similaires. Dans la section suivante, nous présentons un état de l'art des approches utilisant la fouille de processus dans un contexte d'apprentissage.

3. État de l'art

Il existe de nombreuses techniques d'analyse de données issues de systèmes d'information pédagogiques telles que l'Educational Data Mining (EDM) (Berland *et al.*, 2014) et l'Educational Process Mining (EPM) (Romero, Ventura, 2013). L'Educational Process Mining permet de découvrir des modèles de processus d'apprentissage à partir de journaux d'événements à des fins différentes telles que la prédiction des performances, l'adaptation et la recommandation des étudiants.

La fouille de processus a été également utilisée pour analyser les parcours d'apprentissage des étudiants.

(Beemt *et al.*, 2018) explorent la relation entre le comportement d'apprentissage et les progrès d'apprentissage dans les MOOC afin de mieux comprendre comment les étudiants qui réussissent et ceux qui échouent répartissent leurs activités au cours des semaines de cours. Pour trouver les modèles de comportement d'apprentissage des étudiants dans le MOOC, une analyse des trajectoires d'apprentissage est réalisée en utilisant la fouille de processus. Le clustering hiérarchique est utilisé afin d'extraire la connaissance. Ils ont trouvé quatre groupes d'étudiants significatifs, chacun représentant un comportement spécifique allant du simple début à la fin du cours. Les techniques d'exploration de processus montrent que les étudiants qui réussissent affichent un comportement d'apprentissage plus régulier.

Dans (Romero *et al.*, 2008), une tâche de prétraitement est effectuée pour regrouper les utilisateurs en fonction de leur type d'interactions avec le cours. Cette étude permet de découvrir les comportements de navigation les plus spécifiques en utilisant uniquement les données groupées plutôt que le jeu de données complet en appliquant l'algorithme *Heuristic Miner*.

(Real *et al.*, 2020) ont présenté les résultats de l'utilisation de techniques de fouille de processus pour valider les parcours d'apprentissage des étudiants dans un cours d'introduction à la programmation. Ils ont utilisé un journal d'événements Moodle contenant 24605 événements soumis par 73 étudiants de premier cycle. Les résultats ont révélé que, dans l'ensemble, les étudiants qui ont réussi et ceux qui ont échoué ont emprunté des chemins différents pour réaliser les activités du cours. Ils ont également obtenu les flux de contrôle et les fréquences des activités et des connexions pour identifier les dépendances et les ressources qui ont démarré ou terminé le processus. L'analyse de ces résultats fournit des informations générales et spécifiques sur les parcours d'apprentissage des étudiants, ainsi que la possibilité pour les enseignants d'observer les comportements et les progrès des étudiants.

(Martinez *et al.*, 2021) a examiné les trajectoires d'apprentissage des étudiants dans leurs études d'informatique. L'étude s'est concentrée sur la modélisation des caractéristiques qui influencent les taux d'abandon. Par conséquent, les trajectoires des étudiants ayant abandonné leurs études et de ceux qui les ont terminées sont analysées et comparées à l'aide d'outils de fouille de processus. Les auteurs ont constaté que les cours jugés difficiles et entravant la progression académique ont été identifiés, tout

comme les derniers cours suivis par les étudiants avant l'abandon ou l'obtention du diplôme. Ils ont également constaté que les étudiants en décrochage terminent généralement après la première année et suivent des cours de programmation.

Enfin, nous présentons des travaux qui utilisent la fouille de processus avec un objectif de recommandation d'activités pédagogiques.

En utilisant la fonctionnalité de journalisation expérimentale de l'outil de modélisation JMermaid et les techniques de fouille de processus, (Sedrakyan *et al.*, 2014) analyse le modèle de comportement (données d'événements de modélisation conceptuelle de 20 cas et 10.000 événements). Les résultats de ce travail comprennent des modèles qui indiquent une performance d'apprentissage pire/meilleure. Les auteurs soutiennent que les résultats aident à améliorer les conseils d'enseignement pour la modélisation conceptuelle visant à fournir un feedback orienté vers le processus et fournissent des recommandations sur le type de données qui peuvent être utiles pour observer le comportement de modélisation du point de vue des résultats d'apprentissage.

Dans (Leblay *et al.*, 2018), les auteurs ont cherché à proposer une méthode d'assistance basée sur le parcours d'apprentissage pour aider les apprenants à construire leur parcours universitaire. Ils utilisent la découverte de processus pour extraire les parcours d'apprentissage des apprenants précédents, puis utilisent ce modèle pour recommander l'étape la plus appropriée pour guider l'utilisateur ou l'apprenant actuel.

La fouille de processus a été utilisée pour expliquer des comportements, analyser des parcours d'apprentissage et également avec un objectif de recommandation. Ces derniers, travaux sur la recommandation, se basent sur l'analyse du modèle global d'apprentissage. Aucun travail n'aborde la recommandation à travers le *trace clustering* utilisé pour classer les scénarios d'apprentissage.

Dans la section suivante, nous proposons une architecture exploitant le *trace clustering* pour la recommandation de parcours pédagogique.

4. Architecture pour la recommandation

Notre architecture se base sur les travaux de (Ghorbel *et al.*, 2015) qui traitent de l'adaptation de contenu pédagogiques pour la personnalisation des enseignements. Cette architecture est organisée en quatre couches : *source* (qui stocke les objets d'apprentissage et les données de profil), *interoperability* (qui gère l'hétérogénéité des données et propose des mécanismes de fusion), *adaptation* (qui propose une adaptation du scénario pédagogique) et *client* (chargée des interactions utilisateur).

Cette architecture permet : i) l'échange de données entre des profils d'apprenants hétérogènes sur la base d'une couche d'interopérabilité et ii) d'adapter aux apprenants la navigation dans les ressources d'apprentissage sur la base d'une couche d'adaptation. Dans (Hachicha *et al.*, 2021), nous étendons cette architecture avec deux autres couches qui utilisent la fouille de processus pour la recommandation. Nous étendons

cette architecture en ajoutant, au niveau de la couche fouille de processus, une étape de *trace clustering*.

L'architecture que nous proposons (Figure 1) est composée de quatre couches : client, recommandation, fouille de processus et source. Nous introduisons brièvement les 2 premières et présentons plus en détail les 2 dernières, au coeur de la présente communication.

4.1. La couche client

La couche client permet l'interaction entre l'apprenant et les systèmes d'apprentissage en ligne (LMS). Ainsi, l'apprenant peut envoyer une requête en cliquant sur les liens fournis à travers différents types de dispositifs (PC, mobile...).

4.2. La couche recommandation

La recommandation se fait en cours d'exécution et consiste à déterminer, à partir du modèle des processus et de la séquence réalisée jusque là par l'utilisateur, quelle activité doit être réalisée pour finir le processus en optimisant des critères préalablement définis. Les mécanismes de recommandation mis en place sont en dehors de la portée du présent article

4.3. La couche source

La couche source contient des bases de données distribuées de journaux d'événements, de modèles de processus, de profils globaux et de ressources d'apprentissage. Les journaux d'événements (*Event Logs*) sont des fichiers qui contiennent une grande quantité de données brutes sur l'interaction des apprenants avec le LMS. Comme les données dans les journaux d'événements sont souvent bruités, cette base de données nécessite une étape de prétraitement. Cette couche contient toutes les traces, laissées par l'utilisateur, que nous allons exploiter afin d'extraire son scénario d'apprentissage.

La base de données de *Global Profil* enregistre un profil pour chaque apprenant qui contient une vue globale de ses données qui sont distribuées dans les différents systèmes d'apprentissage en ligne. La construction du profil global est une étape préliminaire de la couche d'interopérabilité détaillée dans (Troudi *et al.*, 2020). Le profil contient plusieurs caractéristiques : des données personnelles, des données démographiques et des données sociales qui caractérisent les personnes connues (amis) et les intérêts (les intérêts de l'apprenant et ceux de ses amis). Les intérêts représentent les objets d'apprentissage sur lequel l'apprenant prend plaisir à passer du temps à apprendre. Nous étendons ce profil avec des informations sur le type de situation d'apprentissage (à distance, en face à face ou hybride), l'aspect social (travail individuel ou collaboratif) et le style d'apprentissage (aspect théorique ou pratique).

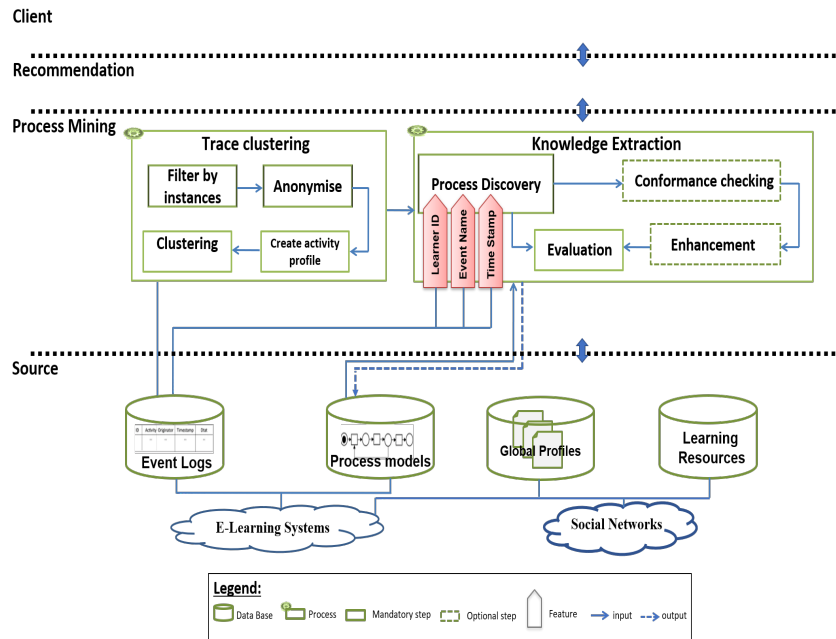


FIGURE 1. Architecture logicielle pour la recommandation d'activités basée sur le trace clustering

La base de données des ressources d'apprentissage, *Learning Resources* contient tout élément impliqué dans le processus d'apprentissage tel que les cours, quiz, pages web, images, vidéos...

4.4. La couche fouille de processus

La couche de fouille de processus permet de découvrir des modèles de processus d'apprentissage basés sur les journaux d'événements. Elle comprend une étape de *trace clustering* afin d'identifier des groupes homogènes de modèles d'apprenants (instances de processus possédant des caractéristiques proches). Nous nous sommes basés sur l'approche *feature-based clustering* en considérant la fréquence des activités par variant.

Notre apport porte sur l'ajout de cette étape de *trace clustering*. Nous cherchons à regrouper les étudiants en fonction de leurs scénarios d'apprentissage préférentiels avant d'extraire des informations qui nous aideront pour la couche de recommandation.

Comme le montre Figure 1, cette étape commence d'abord par le filtrage du journal des événements par instances pour nous assurer que seules les activités des apprenants

soient conservées. Nous éliminons ainsi le bruit (activités d'administration du cours Moodle), les *outliers*, et les traces incomplètes (abandons en cours de formation)

Nous avons ensuite créé le profil d'activité qui est utilisé pour diviser le journal des événements. Le profil d'activité est une matrice $N \times M$, où N représente le nombre d'apprenants et M le nombre d'activités. Chaque ligne de cette matrice correspond à une trace vectorielle qui est composée de fréquences d'activités.

Enfin, nous appliquons un algorithme de *clustering* pour diviser les traces en fonction du profil d'activité. Afin de sélectionner l'algorithme de *clustering* le plus adapté à notre cas nous avons réalisé des essais avec *DBSCAN*, *Agglomerative Clustering*, *Gaussian Mixture Model* et *k-means*. Le détail est donné dans la section suivante (5).

Les résultats du *trace clustering* constituent l'entrée de l'algorithme de découverte des processus. Ils sont constitués du scénario pédagogique suivi par les étudiants d'un même regroupement. Il est alors possible d'appliquer les algorithmes de découvertes de processus afin d'obtenir le modèle de scénario pédagogique pour un groupe.

Chaque modèle de processus montre le comportement d'utilisation le plus courant des apprenants dans un LMS. L'analyse du modèle de processus permet de visualiser et de reproduire le comportement réel de l'apprenant, de trouver des *patterns* dans le comportement d'apprentissage des apprenants, et plus encore de proposer une explication sur le scénario d'apprentissage et ainsi de la recommandation faite.

5. Validation

Afin de valider notre architecture nous avons extrait les journaux d'événements des apprenants ayant suivi le cours « Introduction aux interfaces homme-machine (IHM) » créé sur la plateforme Moodle à l'Université de La Rochelle (France). Un scénario pédagogique a été établi permettant aux apprenants d'atteindre l'objectif final. Les journaux d'événements comprennent 42 438 événements de 100 étudiants (c'est-à-dire 100 traces) ayant suivi le cours pendant un semestre. Chaque événement correspond à une activité réalisée par un apprenant.

Nous nous concentrons, dans cet article sur l'étape de regroupement des traces pour déterminer s'il est possible d'obtenir des ensembles de scénarios pédagogiques pertinents pour la recommandation.

Afin d'extraire des clusters de traces homogènes, nous avons créé le profil d'activité, puis nous avons appliqué l'algorithme de clustering. Dans notre cas particulier, le profil d'activité est une matrice 100×28 , où 100 représente le nombre d'apprenants et 28 représente le nombre d'activités (par exemple *Course module viewed* ou *Quiz attempt started*). Nous avons calculé le nombre d'occurrences de chaque activité dans chaque trace.

Nous avons effectué plusieurs expériences avec différents algorithmes de clustering tels que *DBSCAN* (Ester *et al.*, 1996), *Agglomerative Clustering* (Müllner, 2011), *Gaussian Mixture Model (GMM)* et *k-means* (Hachicha *et al.*, 2023).

Pour découvrir le modèle de processus pour chaque cluster, nous avons appliqué l'algorithme Heuristique Miner, car il a donné de meilleurs résultats dans nos travaux précédents (Hachicha *et al.*, 2021). Un exemple est donné par Figure 2 pour le cluster C2.

Nous appliquons les mesures de qualité classiques (Fitness (F), Précision (P) et Généralisation (G)) sur les modèles découverts.

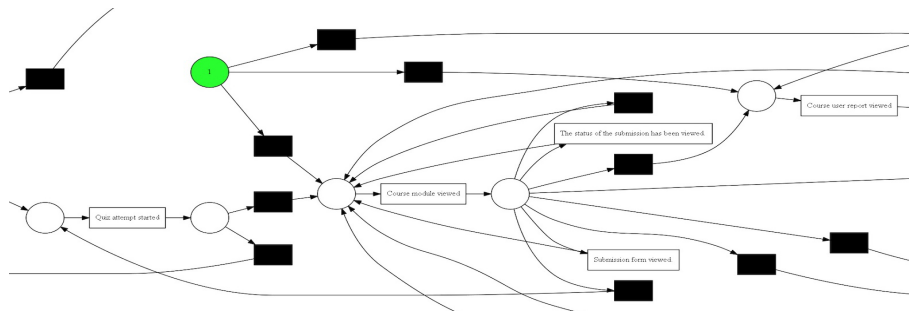


FIGURE 2. Extrait du modèle de parcours d'apprentissage découvert pour le cluster C2

Nous avons fixé le nombre de clusters à 5 après une étude expérimentale (*Elbow method*). Table 1 présente les résultats des mesures après l'application des algorithmes de clustering¹ *k-means*, *DBSCAN*, *Agglomerative Clustering* et *Gaussian Mixture Model* respectivement. Les cinq clusters sont identifiés par les labels C0, C1, C2, C3 et C4. Les clusters diffèrent en fonction des activités des apprenants. Dans C3 le comportement décrit correspond uniquement à utiliser moodle pour effectuer les quiz. Dans C2 les étudiants vont visualiser le cours avant de faire le quiz. Le cluster C4 correspond à des étudiants qui en plus ont téléchargé des archives disponibles sur la page moodle...

Puis, nous avons calculé la moyenne de chaque métrique afin d'identifier l'algorithme de clustering le plus performant. Nous avons utilisé pour cela la bibliothèque PM4Py² qui implémente le calcul basé sur la notion d'alignement. La moyenne de chaque métrique mesurée est abrégée par Avg.

Les résultats ont révélé que les modèles de processus découverts basés sur les clusters générés par *Gaussian Mixture Model* sont les meilleurs comparés aux autres algorithmes. Nous avons obtenu 0,9837 pour la valeur de Fitness, ce qui implique que le modèle rend mieux compte des comportements présents dans le journal des événements. En outre, nous avons enregistré la valeur la plus élevée en termes de précision, ce qui indique que le taux d'activités dans les journaux d'événements est de 0,3489 par rapport au total des activités détectées dans le modèle de processus. Nous

1. Les implémentations des calculs de clustering ont été faites avec la bibliothèque scikit-learn

2. <https://pm4py.fit.fraunhofer.de/>

avons enregistré 0,7692 pour la valeur de Généralisation, ce qui confirme la capacité du modèle à généraliser les comportements présents dans le journal d'événements.

TABLEAU 1. Critères de qualité obtenus après utilisation de différents algorithmes de clustering

	k-means			DBSCAN			Agglomerative			GMM		
	F	P	G	F	P	G	F	P	G	F	G	
C0	0.9922	0.1951	0.7718	0.9887	0.1441	0.7852	0.9892	0.2037	0.8019	0.9862	0.2393	0.8429
C1	0.9847	0.2218	0.8235	0.9841	0.2479	0.8554	0.9796	0.2960	0.7447	0.9799	0.1972	0.7597
C2	0.9652	0.1624	0.6668	0.9865	0.1909	0.6172	0.9923	0.1485	0.7963	0.9652	0.1624	0.6668
C3	0.9838	0.2884	0.7663	0.9332	1	0.7836	0.9652	0.1624	0.6668	0.9923	0.1459	0.7607
C4	0.9923	0.1486	0.7938	0.9802	0.1586	0.6003	0.9909	0.1623	0.7553	0.9949	1	0.8162
Avg	0.9836	0.2023	0.7644	0.9745	0.3483	0.7283	0.9834	0.1945	0.7530	0.9837	0.3489	0.7692

TABLEAU 2. Comparaison des modèles découverts en utilisant le trace clustering et sur l'ensemble des traces

Modèles de processus obtenus	F	P	G
sans utiliser le trace clustering (Hachicha <i>et al.</i> , 2021)	0,9903	0,1596	0,7611
en utilisant le <i>trace clustering</i>	0,9837	0,3489	0,7692

Afin de valider l'apport du *trace clustering*, nous avons effectué une comparaison entre les modèles de processus découverts sans appliquer le clustering et avec. Pour faire la comparaison, nous utilisons les critères de qualité classiques (Fitness, Précision et Généralisation). Table 2 montre que les modèles obtenus en utilisant le *trace clustering* sont de meilleure qualité. En effet, la valeur de Précision passe de 0,1596 à 0,3489 et la valeur de Généralisation augmente légèrement. Même si la valeur de Fitness diminue légèrement, notre expérimentation montre que le *trace clustering* nous permet d'extraire des modèles plus pertinents. Modèles qui nous permettront de produire des recommandations plus précises.

6. Conclusion

Dans le cadre de l'élaboration d'un outil d'aide à la construction d'un parcours personnalisé, nous visons à développer un outil qui recommande en se basant sur la fouille de processus. Nous avons présenté le domaine de la fouille de processus. Puis nous avons proposé un état de l'art des travaux utilisant la fouille de processus pour l'apprentissage. Nous n'avons pas trouvé de travaux utilisant le *trace clustering* pour la recommandation dans le domaine de la formation.

Le *trace clustering* vise à regrouper des traces d'exécutions qui possèdent des dynamiques proches. Nous avons proposé une architecture pour la recommandation qui utilise le *feature based clustering*. Nous avons validé cette architecture à partir de données issues d'un cours d'introduction à la programmation d'IHM comportant 42438 événements. Nous avons montré que nous pouvons déterminer des regroupements pertinents.

L'intérêt de passer par le *trace clustering* est que l'on détermine à quelle catégorie de parcours d'apprentissage un apprenant se rattache. Pour ce parcours nous pouvons

extraire un modèle et ainsi nous pouvons fournir des éléments d'explication de la proposition de recommandation faite.

Les perspectives de ces travaux portent sur la définition d'algorithmes de recommandation qui seront destinés à améliorer l'efficacité du processus de suivi et d'anticipation des actions en lien avec un objectif déclaré ou estimé. Nous visons à développer une plateforme numérique définissant un espace facilitateur et fédérateur d'engagement installant une dynamique nouvelle de l'occupation de l'espace d'apprentissage et d'enseignement et intégrant la gestion personnalisée ou collaborative des ressources d'apprentissage. Nous visons à favoriser l'accessibilité, la continuité et la porosité de l'apprentissage en s'appuyant sur la recommandation des ressources adaptées à un projet personnel ou professionnel d'apprentissage réalisés sur un territoire.

La méthode que nous proposons, qui consiste à considérer le parcours d'un apprenant comme un processus, puis à utiliser les traces laissées par l'apprenant pour regrouper et découvrir les modèles de comportement pour recommander la prochaine action afin d'atteindre un objectif, peut être utilisée dans différents contextes. Par exemple celui des bibliothèques numériques où il est possible de caractériser la navigation de l'utilisateur afin d'adapter le contenu proposé.

Ce besoin de caractériser la navigation de l'utilisateur se retrouve également au niveau des sites web. Nous nous rapprochons alors du *Web Usage Mining*, cependant nous proposons un modèle de navigation qui permet de construire un raisonnement explicable pour la recommandation. Notre approche trouverait tout son intérêt pour des environnements où l'utilisateur déroule son propre processus (banque, assurance, administration...).

Remerciements

Ce travail a été soutenu financièrement par le programme PHC Utique du Ministère français des Affaires étrangères et du Ministère de l'Enseignement supérieur et de la recherche et par le Ministère tunisien de l'Enseignement supérieur et de la recherche scientifique dans le cadre du projet CMCU numéro 22G1403.

Bibliographie

- Aalst W. Van der, Weijters A., Mărușter L. (2004). Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, p. 1128–1142.
- Adriansyah A. (2014). *Aligning observed and modeled behavior*. Thèse de doctorat non publiée, Technische Universiteit Eindhoven.
- Agrawal R., Gunopulos D., Leymann F. (1998). Mining process models from workflow logs. In H.-J. Schek, G. Alonso, F. Saltor, I. Ramos (Eds.), *Advances in database technology — edbt'98*, p. 467–483. Berlin, Heidelberg, Springer Berlin Heidelberg.
- Beemt A. van den, Buijs J., van der Aalst W. (2018). Analysing structured learning behaviour in massive open online courses (moocs): An approach based on process mining and clustering. *International Review of Research in Open and Distributed Learning*, vol. 19, n° 5, p. 37–60.

- Berland M., Baker R., Blikstein P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, vol. 19, p. 205–220.
- Buijs J. C., Van Dongen B. F., van Der Aalst W. M. (2012). On the role of fitness, precision, generalization and simplicity in process discovery. In *Otm confederated international conferences*, p. 305–322.
- Cook J. E., Wolf A. L. (1998). Discovering models of software processes from event-based data. *ACM Trans. Softw. Eng. Methodol.*, vol. 7, p. 215–249.
- Cordier A., Lefevre M., Champin P.-A., Georgeon O., Mille A. (2013). Trace-Based Reasoning - Modeling interaction traces for reasoning on experiences. In P. McCarthy (Ed.), *The 26th International FLAIRS Conference*, p. 1-15. St. Pete Beach, Florida, United States. Consulté sur <https://hal.archives-ouvertes.fr/hal-00830444>
- Diamantini C., Genga L., Potena D. (2016). Behavioral process mining for unstructured processes. *Journal of Intelligent Information Systems*, vol. 47, n° 1, p. 5–32.
- Ester M., Kriegel H.-P., Sander J., Xu X. *et al.* (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Second international conference on knowledge discovery and data mining*, vol. 96, p. 226–231.
- Ghorbel L., Zayani C., Amous I. (2015). Improve the adaptation navigation in educational cross-systems. *Procedia Computer Science*, vol. 60, p. 662-670.
- Hachicha W., Ghorbel L., Champagnat R., Zayani C. A. (2023). Trace clustering based on activity profile for process discovery in education. In *Intelligent systems design and applications: 22th international conference on intelligent systems design and applications (isda 2022) held december 12-14, 2022*.
- Hachicha W., Ghorbel L., Champagnat R., Zayani C. A., Amous I. (2021). Using process mining for learning resource recommendation: A moodle case study. *Procedia Computer Science*, vol. 192, p. 853-862. Consulté sur <https://www.sciencedirect.com/science/article/pii/S1877050921015763> (Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 25th International Conference KES2021)
- Ho H. N., Rabah M., Nowakowski S., Estrailier P. (2016). Toward a Trace-Based PROMETHEE II Method to answer ” What can teachers do? ” in Online Distance Learning Applications. In *13th International Conference on Intelligent Tutoring Systems*, p. 480-484. Zagreb, Croatia. Consulté sur <https://hal.science/hal-01334151>
- Leblay J., Rabah M., Champagnat R., Nowakowski S. (2018). Process-based Assistance Method for Learner Academic Achievement. In *E-learning conference (el'2018)*, p. 89-96. Madrid, Spain. Consulté sur <https://hal.science/hal-01834096>
- Li G., De Carvalho R. M. (2019). Process Mining in Social Media: Applying Object-Centric Behavioral Constraint Models. *IEEE Access*, vol. 7, p. 84360–84373.
- Martinez P., Montañes O., Serralta J. M., Tansini L. (2021). Modelling computer engineering student trajectories with process mining. In *Latin american conference on learning analytics*, p. 48–57.
- Müllner D. (2011). Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*.

- Pika A., Wynn M. T., Budiono S., Hofstede A. H. M. ter, Aalst W. M. P. van der, Reijers H. A. (2019). Towards privacy-preserving process mining in healthcare. In C. Di Francescomarino, R. Dijkman, U. Zdun (Eds.), *Business process management workshops*, p. 483–495. Cham, Springer International Publishing.
- Real E. M., Pimentel E. P., Oliveira L. V. de, Braga J. C., Stiubiener I. (2020). Educational process mining for verifying student learning paths in an introductory programming course. In *2020 IEEE Frontiers in Education Conference (FIE)*, p. 1–9.
- Romero C., Ventura S. (2013). Data mining in education. *Wiley Int. Rev. Data Min. and Knowl. Disc.*, vol. 3, n° 1, p. 12–27. Consulté sur <https://doi.org/10.1002/widm.1075>
- Romero C., Ventura S., García E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, vol. 51, n° 1, p. 368–384. Consulté sur <https://www.sciencedirect.com/science/article/pii/S0360131507000590>
- Sedrakyan G., Snoeck M., De Weerd J. (2014). Process mining analysis of conceptual modeling behavior of novices – empirical study using jmermaid modeling and experimental logging environment. *Computers in Human Behavior*, vol. 41, p. 486–503.
- Song M., Günther C. W., Aalst W. M. Van der. (2008). Trace clustering in process mining. In *International conference on business process management*, p. 109–120.
- Trabelsi M., Suire C., Morcos J., Champagnat R. (2019a). Fouille de processus auto-définis : cas d'étude d'un moteur de recherche d'une bibliothèque numérique. In *Inforsid'2019*, p. 131–146.
- Trabelsi M., Suire C., Morcos J., Champagnat R. (2019b). User's behavior in digital libraries: Process mining exploration. In A. Doucet, A. Isaac, K. Golub, T. Aalberg, A. Jatowt (Eds.), *Digital libraries for open knowledge*, p. 388–392. Cham, Springer International Publishing.
- Trabelsi M., Suire C., Morcos J., Champagnat R. (2021). A new methodology to bring out typical users interactions in digital libraries. In *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, p. 11–20.
- Troudi A., Ghorbel L., Amel Zayani C., Jamoussi S., Amous I. (2020). MDER: Multi-Dimensional Event Recommendation in Social Media Context. *The Computer Journal*, vol. 64, n° 3, p. 369–382. Consulté sur <https://doi.org/10.1093/comjnl/bxaa126>
- Van der Aalst W. (2016). *Process mining: Data science in action*. Springer.
- Weijters A., Ribeiro J. (2011). Flexible heuristics miner (fhm). In *Computational intelligence and data mining*, p. 310–317.
- Zandkarimi F., Rehse J.-R., Soudmand P., Hoehle H. (2020). A generic framework for trace clustering in process mining. In *2020 2nd international conference on process mining*, p. 177–184.

COVID-19 et (dés)information : l'open data face à ses limitations

Robert Viseur¹

*1. Service TIC, FWEG, UMONS
17 place Warocqué, B-7000 Mons, Belgique
robert.viseur@umons.ac.be*

RÉSUMÉ. La COVID-19 a mis en évidence l'importance du numérique dans la lutte contre une pandémie. En particulier, la France s'est distinguée par la diversité des jeux de données publiés en open data. Cependant, elle n'a pas été épargnée par les polémiques relatives à leur qualité et à leur interprétation. Sur base d'un ensemble d'études de cas réutilisant les données du SPF (Santé publique France) et de la DREES (Direction de la Recherche, des Études, de l'Évaluation et des Statistiques), nous analysons la qualité des données, en nous basant sur le concept de pertinence des représentations de Reix et al. (2011), puis les obstacles à leur réutilisation. Nous identifions des problèmes classiques en analyse de données (p. ex. difficultés d'interprétation liées au volume ou à la sémantique des données). Nous analysons ensuite les conséquences sur la réutilisation des open data de deux dimensions de ces systèmes d'information : (1) le développement itératif des systèmes d'information de santé utilisés dans le cadre de la pandémie et (2) le fonctionnement en silos de ces systèmes d'information. Dès lors, nous identifions comment les défauts de pertinence peuvent entraîner un défiance ou un mésusage de ces données.

ABSTRACT. COVID-19 highlighted the importance of digital technology in the fight against a pandemic. In particular, France stood out for the diversity of the datasets published in open data. However, it has not been spared from controversies regarding their quality and interpretation. Based on a set of case studies reusing SPF and DREES data, we analyse the quality of the data, based on the concept of relevance of representations by Reix et al. (2011), and then the obstacles to their reuse. We identify classic problems in data analysis (e.g., interpretation difficulties related to data volume or semantics). Then we analyze the consequences for open data reuse of two dimensions of these information systems: (1) the iterative development of health information systems used in the pandemic and (2) the siloed operation of these information systems. Therefore we identify how relevance flaws can lead to distrust or misuse of these data.

Mots-clés : open data, COVID-19, pertinence des représentations.

KEYWORDS: open data, COVID-19, data quality.

1. Contexte

En décembre 2019 démarrait en Chine une épidémie de pneumonie attribuée à un nouveau coronavirus baptisé SARS-CoV-2. Parmi les moyens numériques, les *open data*, c'est-à-dire des informations qui ont été rendues techniquement et légalement disponibles à des fins de réutilisation (Lindman & Tammisto, 2011), se sont rapidement dégagées comme un instrument facilitant le suivi de la pandémie. Cependant, si Lobre et Lebraty (2012) reconnaissent que « *la discussion offre une chance de réinterpréter les données de manière efficace et fiable* », ils soulignent aussi les risques liés à la « *manipulation de l'information* ». Un élément nous semble par ailleurs sous-étudié dans le cadre de cette pandémie : les conséquences, sur la qualité de l'information, de la politique de publication des *open data*. Même si la France est apparue comme un bon élève en matière d'*open data* COVID-19 (Viseur, 2021 ; Nikiforova, 2020), des voix se sont aussi élevées pour dénoncer la lenteur de publication des données relatives à la pandémie (Chignard, 2021). La politique en matière d'*open data* a été stimulée par l'activité de collectifs (p. ex. OpenCOVID et « *Data against Covid* ») procédant méthodiquement à l'encodage de nouveaux jeux de données sur base des rapports officiels, publiant des tableaux de bord permettant un suivi de la pandémie et interpellant régulièrement le gouvernement pour qu'il publie davantage de données (Goëta, 2022 ; Chignard, 2021 ; Bothorel et al., 2020). Une fois publiées, ces *open data* peuvent donc faciliter l'information du public. En France, des sites comme [CovidTracker](https://covidtracker.fr/)¹ sont progressivement devenus incontournables pour le suivi médiatique de la pandémie (Goëta, 2022 ; Bothorel et al., 2020). Malgré ces efforts de transparence, la pandémie de la COVID-19 a été associée à une « *prolifération de théories du complot* » (Bottemanne, 2022). Cette recherche s'attaque donc à la compréhension des facteurs liés à la qualité de ces *open data* expliquant pourquoi, dans un pays ayant joué la carte de la transparence, le suivi de la pandémie n'a pas échappé aux polémiques constatées ailleurs.

Les organisations prennent leurs décisions sur la base de représentations de la réalité (p. ex. tableau de bord de gestion). Dès lors, en tant que représentation d'une réalité, ces *open data* COVID-19 sont-elles « *pertinentes* » ? La « *pertinence des représentations* », analysée par Reix et al. (2011), renvoie à leur utilisation. Est pertinent « *ce qui convient, ce qui est approprié à l'action* » (Reix et al., 2011). Huit critères de qualité permettent, selon Reix et al. (2011), de juger la pertinence des représentations : (1) l'accessibilité (modalités de recherche et d'accès) ; (2) l'exactitude, qui concerne l'absence de bruit (les représentations devraient idéalement échapper au risque de deuxième espèce relatif à la prise en compte d'évènements provoqués par des variations aléatoires dues à des imperfections de la fonction d'information) ; (3) l'exhaustivité, qui fait référence à la complétude de la représentation (tous les changements d'état significatifs du réel sont couverts par les représentations ; cf. risque de première espèce) ; (4) la fiabilité (confiance en la source ; dépendance aux risques de 1^{ère} et de 2^{ème} espèce) ; (5) le degré de finesse, qui concerne la précision de la représentation (niveau de détail ou intervalle de variation) ; (6) l'actualité (fraîcheur), (7) la ponctualité (respect des échéances) ; et (8) la forme (données, dessins, images fixes ou animées...). Cette question de la pertinence des représentations s'inscrit dans un questionnement plus général sur le poids occupé par les chiffres dans notre société. « *On peut débattre de tout sauf des*

¹ Cf. <https://covidtracker.fr/>.

chiffres » affirmait ainsi le gouvernement français lors d'une campagne en faveur de la vaccination, usant ainsi « *des chiffres comme argument d'autorité* » (Goëta, 2022). Dans une série de cours au [Collège de France](#) donnés entre 2012 et 2014, Alain Supiot (2015) critique notamment le mésusage des indicateurs, devant être interprétés, « *substituant la carte au territoire et la réaction à l'action* », congédiant « *le vocabulaire de la démocratie politique au profit de celui de la gestion* », ayant pour effet de se couper de la complexité du réel.

2. Données et analyse des résultats

Les *open data* fournies par l'état français le sont par Santé Publique France (SPF) via le portail [data.gouv.fr](#). Ces données incluent les « *Données hospitalières relatives à l'épidémie de COVID-19* » (cf. [\[url\]](#)), les « *Données relatives aux résultats des tests virologiques COVID-19* » (cf. [\[url\]](#)) et les « *Données relatives aux personnes vaccinées contre la Covid-19* » (cf. [\[url\]](#)). La DREES (Direction de la Recherche, des Études, de l'Évaluation et des Statistiques) a par la suite publié des fichiers constitués des appariements entre ces trois sources de données (cf. [\[url\]](#)). Compte tenu de leur mention fréquente sur les réseaux sociaux francophones, les *open data* américaines relatives aux effets secondaires présumés des vaccins, publiées au sein du [VAERS](#) (*Vaccine Adverse Reporting System*), ont également été exploitées. Plusieurs expérimentations sur base de ces données ont été réalisées entre août 2021 et février 2022. Elles suivaient en pratique les polémiques médiatiques, notamment relatives à la comptabilisation du nombre total de décès, aux causes réelles des décès, à l'efficacité des vaccins, à leur innocuité et à l'ampleur des variations du nombre d'hospitalisations (notamment chez les enfants). À l'aide des critères de « *pertinence des représentations* » de Reix et al. (2011), nous avons procédé à l'analyse des données mobilisées dans les cas de réutilisation. Le Tableau 1 distingue, pour chacun des huit critères, les données relatives aux tests, aux hospitalisations, aux décès et aux vaccinations.

Tableau 1. Données COVID et pertinence des représentations.

	(a) Tests	(b) Hospitalisations	(c) Morts	(d) Vaccinations
(1) Accessibilité	Pas d'accès (01/2022) au pourcentage de tests positifs à Omicron.	Pas d' <i>open data</i> sur le statut vaccinal (S1 2021) ou sur le statut de comorbidité.	Plus d'accès aux statistiques de décès dans les EHPAD (cf. [url]).	Pas d'accès en France, au contraire des USA (VAERS), aux données brutes relatives aux signalements d'effets secondaires présumés.
(2) Exactitude	Positivité des tests influencée par la capacité ou par la politique de tests.	Pas de distinction entre les hospitalisations « pour COVID » et « avec COVID ».	Pas de distinction entre les morts « pour COVID » et « avec COVID ».	Positivité (tests) des populations vaccinés influencées par les politiques de tests différenciées entre vaccinés et non vaccinés (principe du pass sanitaire).
(3) Fiabilité	Problèmes liés à la sémantique des fichiers <i>open data</i> .	Problèmes liés à la sémantique des fichiers <i>open data</i> (total des hospitalisations, hospitalisations conventionnelles, soins critiques, réanimations...).	Difficulté à comptabiliser, et distinguer, les morts en hôpital, en EHPAD et à domicile.	Régularité des remontées d'effets secondaires présumés liés aux vaccins (complexité du formulaire de signalement de l'ANSM, actes militants liés aux mouvements antivax...).
(4) Exhaustivité	Influence du choix de technologies de tests comptabilisés (auto-tests, salivaires, antigéniques, RT-PCR).	Indisponibilité, réparée au fil du temps, de l'âge, du statut vaccinal... et du statut de comorbidité.	Comptabilisation séparée des morts en hôpital ou des morts en EHPAD.	Absence d'indicateur lié à l'immunité naturelle (primoinfections).
(5)	Publication d'indicateurs	Publication d'indicateurs	Publication d'indicateurs	Publication d'indicateurs

	(a) Tests	(b) Hospitalisations	(c) Morts	(d) Vaccinations
Finesse	globaux ou d'indicateurs par critère (p. ex. tranche d'âge et statut vaccinal).	globaux ou d'indicateurs par critère (p. ex. tranche d'âge et statut vaccinal).	globaux ou d'indicateurs par critère (p. ex. tranche d'âge et statut vaccinal).	globaux ou d'indicateurs par critère (p. ex. tranche d'âge).
(6) Actualité	Irrégularités dans les remontées de cas (p. ex. <i>weekends</i> et reprise des erreurs).	Problème du décalage temporel entre les cas et les hospitalisations (prévisions) et de la variation des délais au fil du temps (variants).	Problème du décalage temporel entre les cas et les décès (prévisions) et de la variation des délais au fil du temps (variants).	Accès aux informations sur les effets secondaires présumés ou confirmés après traitement (ANSM).
(7) Ponctualité	Retards sur le criblage pour le suivi du variant Omicron.		Rattrapage sur les morts en EHPAD (cf. [url1] et [url2]).	
(8) Forme	Focalisation sur des indicateurs globaux, graphiques et spectaculaires (p. ex. exponentielle du nombre de cas).	Utilisation d'échelles permettant d'exagérer les phénomènes (p. ex. hospitalisations pédiatriques).	Focalisation sur des indicateurs globaux.	Focalisation sur des indicateurs globaux (p. ex. graphique de l'évolution de la couverture vaccinale globale).

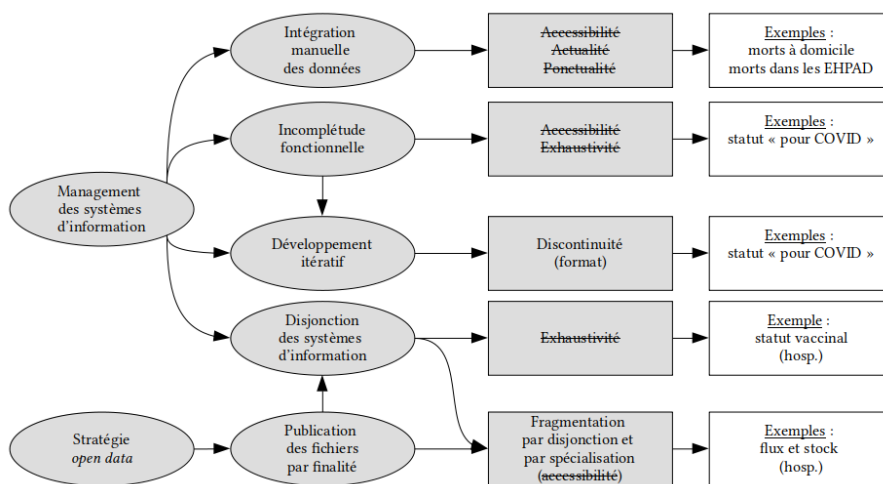


Figure 1. Construction des SI et pertinence des représentations.

Les lenteurs de publication s'expliquent en partie par la construction du système d'information de santé en France. Ronai (2021) distingue ainsi système d'information orienté statistique, avec une filière administrative (INSERM) et une autre sanitaire (INSEE), et système d'information en temps réel. Les *open data* publiées par Santé Publique France (SPF) dépendent essentiellement du second, constitué de trois systèmes d'information : SI-VIC, SI-DEP et VAC-SI (Guillotini et al., 2021 ; Ronai, 2021 ; Bothorel et al., 2020). Créé en 2016, déployé initialement pour répertorier les victimes des attentats de Paris de novembre 2015, SI-VIC sert au départ de système d'information pour le suivi des victimes d'attentats et de situations sanitaires exceptionnelles. Ses fonctionnalités ont été étendues, de manière itérative, pour les besoins liés à la pandémie. Cependant, SI-VIC ne comptabilisait que les décès survenus à l'hôpital et excluait les morts à domicile ou en maison de retraite (soit 44 % des morts de la COVID-19 ; cf. [url]). Ces derniers ont finalement pu être comptabilisés via un quatrième système (Voozanoo) dédié aux EHPAD (Bothorel et al., 2020). SI-DEP, développé par l'AP-HP, disponible depuis juin 2020, contient les résultats des tests PCR et des tests antigéniques. Il est

complété par Contact-COVID qui est dédié au suivi des cas positifs et à la conduite des enquêtes sanitaires permettant la remontée des chaînes de contamination. VAC-SI, développé par le CNAM, disponible à partir du 4 janvier 2021, contient les données relatives à l'identité du patient et à la vaccination.

Si l'existence des systèmes d'information offre globalement les bénéfices d'une intégration rapide des données collectées sur le terrain, ce que la Belgique a par exemple peine à réaliser (Viseur, 2021), leur développement progressif, itératif, conduit à des défauts d'exhaustivité et à des discontinuités dans les formats de données (cf. Figure 1). De plus, le manque d'intégration entre systèmes d'information conduit à la publication de jeux de données reflétant les silos dont ils sont issus. Les données pour un même fichier sont donc relatives, soit aux hospitalisations et aux décès (dans les hôpitaux), soit aux tests, soit aux statuts vaccinaux. Premièrement, il en résulte des angles morts. Les décès en EHPAD ne sont ainsi pas comptabilisés dans les décès repris dans SI-VIC. Deuxièmement, les croisements entre statut vaccinal et hospitalisation sont difficiles. Au niveau des *open data*, le croisement entre les tests, les hospitalisations et les décès n'a été permis que tardivement (c'est-à-dire début août 2021) suite à la publication par la DREES de fichier d'appariement (cf. [url]) permettant par exemple d'analyser le risque d'hospitalisation ou de décès en fonction du statut vaccinal. Cette fragmentation est renforcée par la stratégie de publication en *open data*, où chaque fichier reflète une finalité particulière. L'analyse d'un phénomène (p. ex. hospitalisations pédiatriques) va donc nécessiter le croisement de plusieurs fichiers, ce qui peut décourager certains médiateurs (*gatekeepers*), en renforcer d'autres, sans que leur professionnalisme ne soit nécessairement au rendez-vous.

À ces difficultés s'ajoutent des problèmes, classiques en analyse de données (Bronner, 2013 ; Baillargeon, 2006), relatives à trois formes de complexité liées (1) au volume de données (p. ex. taille des jeux de données), (2) à la compréhension des données (p. ex. sémantique des données) et (3) à l'interprétation des données (p. ex. signification des indicateurs tels que la létalité, méconnaissance des processus de validation des données brutes et confusion entre corrélation et causalité ; cf. effets secondaires présumés des vaccins par exemple). En particulier, certains problèmes de fiabilité se sont révélés imputables à des problèmes de sémantique des fichiers, handicapant, d'une part, la compréhension du sens associé aux données réutilisées, d'autre part, le croisement des données entre fichiers complémentaires. Ces incohérences ponctuelles dans la sémantique des fichiers portent notamment sur les capacités en réanimation. La Cour des comptes a ainsi publié en juillet 2021 un rapport intitulé « *Les soins critiques* » dans lequel le problème a été décrypté (cf. [url]). Le rapport note la confusion opérée par Santé Publique France (SPF) entre les réanimations et les soins critiques ainsi qu'une incertitude quant, d'une part, aux capacités maximales prises en référence, d'autre part, à la part des lits affectés aux patients COVID (pp. 57-60).

3. Conclusion

Au delà des problèmes classiques en analyse de données, cette recherche a permis d'en dégager de plus spécifiques liés, d'une part, à la temporalité du phénomène, d'autre part, à l'intégration des systèmes d'information de santé. La rapidité et la nouveauté du phénomène ont conduit, d'une part, à un développement dans l'urgence des systèmes d'information, d'autre part, à une mise à mal du rythme

de validation des connaissances. Si l'amélioration continue de la qualité est favorisée par les pratiques collaboratives qui accompagnent souvent l'*open data*, elle peut aussi donner l'impression à un public moins informé d'une piètre fiabilité des données publiées, voire d'une volonté de tromper. La complexité d'interprétation des données se trouve accrue par la nécessité d'identifier, puis de croiser, lorsque cela est possible, les différents jeux de données indispensables à l'analyse.

4. Références

- Baillargeon, N. (2006). *Petit cours d'autodéfense intellectuelle*. Lux.
- Bothorel, É., Combes, S., Vedel, R. (2020). *Mission BOTHOREL : pour une politique publique de la donnée*. Conseil général de l'économie.
- Bottemanne, H. (2022). *Théories du complot et COVID-19: comment naissent les croyances complotistes?*. L'encephale.
- Bronner, G. (2013). *La démocratie des crédules*. PUF.
- Chignard, S. (2021). *L'open data de crise : entre mobilisation citoyenne et communication gouvernementale*. *Enjeux numériques*, n°14, juin 2021, pp. 73-77.
- Cour des comptes (2021). *Les soins critiques*. Communication à la commission des affaires sociales du Sénat. Cour des comptes, juillet 2021.
- Goëta, S. (2022). *La pandémie de Covid-19: un moment charnière pour l'ouverture des données en France*. *Statistique et Société*.
- Guillot, V., Lavarde, C., Savary R.-P. (2021). *Crises sanitaires et outils numériques : répondre avec efficacité pour retrouver nos libertés*. Rapport d'information au Sénat, n°673, 03 juin 2021.
- Lindman, J., & Tammisto, Y. (2011). *Open Source and Open Data: Business perspectives from the frontline*. In : *IFIP International Conference on Open Source Systems* (pp. 330-333). Springer, Berlin, Heidelberg.
- Lobre, K., & Lebraty, J. F. (2012). *L'open Data: nouvelle pratique managériale risquée?*. *Gestion 2000*, 29(4), 103-116.
- Nikiforova, A. (2020). *Timeliness of Open Data in Open Government Data Portals Through Pandemic-related Data: a long data way from the publisher to the user*. In *2020 Fourth International Conference on Multimedia Computing, Networking and Applications (MCNA)*, pp. 131-138. IEEE.
- Reix, R., Fallery, B., Kalika, M., & Rowe, F. (2011). *Systèmes d'information et management des organisations*. Vuibert.
- Ronai, M. (2021). *La construction d'un système d'information épidémiologique*. *Enjeux numériques*, n°14, juin 2021, pp. 62-72.
- Supiot, A. (2015). *La gouvernance par les nombres*. Fayard.
- Viseur, R. (2021). *Open data in digital strategies against COVID-19: the case of Belgium*. 17th International Symposium on Open Collaboration. September 2021. Pages 1-10.

ChouBERT : Pré-entraînement d'un modèle de langue française pour le Crowdsensing avec des Tweets dans un contexte phytosanitaire

Shufan Jiang^{1,2}, **Rafael Angarita**^{3,4}, **Stéphane Cormier**²,
Julien Orensanz⁵, **Francis Rousseaux**²

1. Valda, DIENS, École normale supérieure, Paris, France

jiang.chou.fan@gmail.com

2. CRESTIC, Université de Reims Champagne Ardenne, Reims, France

3. Université Paris Nanterre, Nanterre, France

4. LIP6, Sorbonne Université, Paris, France

5. Cap2020, Gradignan, France

RÉFÉRENCE DE L'ARTICLE INTERNATIONAL. ChouBERT: Pre-training French Language Model for Crowdsensing with Tweets in Phytosanitary Context. RCIS 2022: 653-661.

MOTS-CLÉS : apprentissage par transfert, détection des foules, surveillance de la santé végétale, twitter

KEYWORDS: transfer learning, crowd-sensing, plant health monitoring, twitter

Crowdsensing ou la *détection des foules* est un paradigme de détection qui permet aux personnes ordinaires de contribuer avec des données détectées ou générées par leurs appareils mobiles équipés de capteurs (Boubiche *et al.*, 2019). Il introduit un nouveau changement dans la manière dont nous collectons les données en permettant d'acquérir des connaissances locales par le biais de dispositifs intelligents portés par les gens, tels que les smartphones, les tablettes, les montres intelligentes, entre autres. Cela permet d'exploiter les capteurs améliorés des smartphones de manière rapide et économique, contrairement aux méthodes traditionnelles plus coûteuses. Poussés par la reconnaissance croissante de l'importance de l'agriculture pour le maintien de l'humanité et le rôle central des agriculteurs dans la transformation numérique de l'agriculture, nous avons témoigné de l'émergence d'applications de crowdsensing pour l'agriculture (Mendes *et al.*, 2020).

Les agriculteurs sont également de plus en plus présents dans les médias sociaux tels que Facebook, WhatsApp et Twitter, où ils partagent et discutent volontairement

leurs observations sur l'environnement. En particulier, Twitter permet aux agriculteurs de publier librement de courts messages appelés "tweets" pour partager leurs observations. Pour tirer parti de ces observations, il faut garder la trace des sources de données pertinentes parmi le bruit, extraire et organiser les informations qu'elles contiennent et les partager avec d'autres utilisateurs intéressés. Cela n'est possible qu'au prix d'un effort humain important, en inspectant, filtrant et nettoyant manuellement toutes les données et en reliant les entités et les contextes connexes.

Les applications récentes de modèle de langage semblent prometteuses pour résoudre les problèmes d'extraction d'informations. Poussés par la connectivité croissante des agriculteurs et l'émergence de communautés agricoles en ligne, notre objectif est d'explorer l'application émergente des observations à la ferme via les réseaux sociaux -en particulier Twitter- et de proposer une approche pour la classification des tweets. Nous cherchons à répondre aux questions de recherche suivantes : *RQ1. comment des modèles de langage (LM) pré-entraînés peuvent aider à l'exploration des tweets sur la snta végétale ?*; et *RQ2. comment adapter les LM génériques pour la classification de textes dans un domaine spécifique ?*

Nous avons construit ChouBERT¹ en pré-entraînant CamemBERT (Martin *et al.*, 2020) sur les bulletins de santé des végétaux (BSV) et les tweets pour augmenter l'intégration contextuelle des tweets pour la détection des observations. Nos résultats expérimentaux mettent en évidence la généralisation de la représentation de ChouBERT sur des aléas non vus pour la tâche de classification. Nous pouvons généraliser cette approche pour améliorer le crowdsensing basé sur le contenu textuel des tweets : en collectant un jeu initial de tweets à l'aide de mots-clés; en étiquetant manuellement un petit jeu de tweets; en pré-entraînant davantage les modèles de langage à l'aide de documents du domaine et de tweets; et en construisant des applications NLP avec le jeu étiqueté et le modèle de langage adapté au domaine. Enfin, notre expérience montre que l'observation des foules sur Twitter ne remplace pas les autres paradigmes de surveillance, mais constitue une source d'information complémentaire. L'objectif du crowdsensing sur Twitter est de détecter les signaux faibles plutôt que de quantifier la gravité d'un problème par la fréquence des mentions. Il peut être intéressant de croiser ces informations avec d'autres sources de données.

Bibliographie

- Boubiche D. E. *et al.* (2019). Mobile crowd sensing—taxonomy, applications, challenges, and solutions. *Computers in Human Behavior*, vol. 101, p. 352–370.
- Martin L. *et al.* (2020). Camembert: a tasty french language model. In *Proc. of the 58th annual meeting of the association for computational linguistics*, p. 7203–7219. Online, Association for Computational Linguistics.
- Mendes J. *et al.* (2020). Smartphone applications targeting precision agriculture practices—a systematic review. *Agronomy*, vol. 10, n° 6, p. 855.

1. <https://huggingface.co/ChouBERT>

Évaluer la polarisation des réseaux sociaux numériques par l'étude des frontières de communautés – Approche ERIS

Alexis Guyot, Annabelle Gillet, Éric Leclercq, Nadine Cullot

Laboratoire d'Informatique de Bourgogne, Université de Bourgogne
9 avenue Alain Savary, F-21078 Dijon, France
<mailto:{prenom}.{nom}@u-bourgogne.fr>

REFERENCE DE L'ARTICLE INTERNATIONAL. Cet article est un résumé de l'article :
Alexis Guyot, Annabelle Gillet, Éric Leclercq, Nadine Cullot:

ERIS: An Approach Based on Community Boundaries to Assess Polarization in Online Social Networks. RCIS 2022: 88-104

MOTS-CLES : réseaux sociaux, polarisation, frontière de communauté, structure de communauté, fouille de graphe

KEYWORDS : social networks, polarization, community boundaries, community structure, graph mining

La polarisation d'un réseau social se caractérise par la division et l'isolation des membres du réseau à l'intérieur de deux communautés qui interagissent de manière antagoniste (Isenberg, 1986). Détecter ce phénomène peut avoir de nombreuses applications : modération (réduire les conflits et le partage de fausses informations), marketing, etc.

Nous avons proposé une approche non-supervisée basée sur les graphes pour pouvoir évaluer la polarisation d'un réseau social en nous basant uniquement sur sa topologie. Cette approche ne repose pas sur une analyse du contenu des tweets pour éviter les difficultés liées à la taille des messages (< 280 caractères), à la qualité de la langue et à la présence potentielle de sarcasme. Elle ne nécessite pas non plus d'étape d'initialisation basée sur une connaissance a priori du réseau et des utilisateurs. Nous étudions pour cela les frontières des communautés d'utilisateurs. Les frontières sont des individus qui interagissent à la fois avec l'extérieur et avec l'intérieur de la communauté. En tant qu'intermédiaires, leur comportement a un impact important sur la polarisation de leur communauté.

L'approche ERIS (Guyot *et al.*, 2022) consiste à identifier les frontières des communautés d'utilisateurs du réseau puis à étudier leurs interactions avec

l'intérieur et l'extérieur de la communauté pour calculer deux indicateurs. Le premier indicateur, l'antagonisme des communautés, estime le degré d'opposition entre les communautés. Le second, la porosité des frontières, estime le degré d'ouverture des communautés au reste du réseau. En plus des définitions formelles des formules de nos indicateurs, nous avons proposé un algorithme et une implémentation efficace et libre d'accès en R pour pouvoir les calculer. Cette implémentation est basée sur des calculs matriciels. Elle améliore de manière significative les temps d'exécution des implémentations existantes. Nous avons montré son applicabilité sur de grands graphes similaires aux données réelles extraites des réseaux sociaux numériques.

Pour accompagner et valider expérimentalement notre proposition, nous avons mené une étude de cas avec des collègues chercheurs en sciences humaines et sociales dans le cadre du projet Cocktail. Nous avons collecté plus de 18 millions de tweets faisant mention de vaccination et de la Covid-19 publiés entre Décembre 2020 et mars 2021. De ce corpus, nous avons extrait un graphe de citations, une fonctionnalité de partage avec ajout de commentaire adaptée pour la détection d'antagonisme. Dans ce graphe, nous avons identifié 8 communautés d'utilisateurs à l'aide de l'algorithme non-supervisé Louvain. Nous avons ensuite appliqué notre implémentation de l'approche ERIS. Une analyse des hashtags nous a permis d'identifier que les deux communautés principales du réseau avaient des positions explicitement pro ou anti-vaccins. Comme attendu, nous avons observé un fort antagonisme mutuel entre ces deux communautés. Nous avons également observé une faible porosité de leurs frontières, traduisant un certain renfermement. Nos indicateurs nous ont également permis d'identifier quelques différences de comportements entre les frontières pro et anti-vaccins qui traduisent différentes stratégies de contrôle du débat et de l'opinion portée par la communauté. L'analyse de ces différentes informations nous a permis de conclure sur la polarisation du réseau de notre cas d'étude.

Remerciements :

Ce travail est soutenu par le projet ISITE-BFC Cocktail (ANR-15-IDEX-0003), piloté scientifiquement par Gilles Brachotte, laboratoire CIMEOS EA-4177, Université de Bourgogne.

Bibliographie

- Guyot A., Gillet A., Leclercq E., Cullot N. (2022). ERIS: an approach based on community boundaries to assess polarization in online social networks. *Research Challenges in Information Science: 16th International Conference, RCIS 2022 (pp. 88-104)*, Springer International Publishing, Barcelona.
- Isenberg D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of personality and social psychology*, vol. 50, n° 6, p. 1141.

Interrogation des graphes temporels de propriétés

**Landy Andriamampianina^{1,2}, Franck Ravat¹, Jiefu Song¹,
Nathalie Vallès-Parlangeau³**

1. IRIT-CNRS (UMR 5505), Université Toulouse 1 Capitole

2 Rue du Doyen-Gabriel-Marty 31042 Toulouse, France

prenom.nom@irit.fr

2. Activus Group,

1 Chemin du Pigeonnier de la Cépière 31100 Toulouse, France

prenom.nom@activus-group.fr

3. LIUPPA– T2i, Université de Pau et des Pays de l'Adour,

2 Allée du Parc Montaury, 64600 Anglet,

prenom.nom@iutbayonne.univ-pau.fr

Cet article est un résumé de l'article : Landy Andriamampianina, Franck Ravat, Jiefu Song, and Nathalie Vallès-Parlangeau. 2022. Querying Temporal Property Graphs. In Advanced Information Systems Engineering: 34th International Conference, CAiSE 2022, Leuven, Belgium, June 6–10, 2022, Proceedings. Springer-Verlag, Berlin, Heidelberg, 355–370. https://doi.org/10.1007/978-3-031-07472-1_21

MOTS-CLES : Graphe de propriétés, Opérateurs, Evolution temporelle, Bases de données graphes.

KEYWORDS: Property graph, Operators, Temporal evolution, Graph data stores.

De nombreuses activités impliquent des personnes (ou des objets) qui interagissent de différentes manières au fil du temps (Debrouvier et al., 2021). Par exemple, une maladie infectieuse se propage dans le temps à travers les contacts des personnes. Développer des applications pour gérer ces activités nécessite de prendre en compte à la fois l'interconnexion et l'évolution temporelle de ces dernières. L'interconnectivité peut être gérée par un graphe de propriétés, qui relie les entités entre elles par des relations et les décrit par des attributs (Angles, 2018).

L'évolution des données interconnectées dans le temps (ajout, suppression et mise à jour des entités, des relations ainsi que de leurs attributs) n'est pas représentée et donc inexploitable dans les systèmes existants. En effet, les travaux actuels se centrent sur la modélisation statique des graphes de propriétés. Les bases de données basées sur le modèle de graphe de propriétés ne permettent pas non plus de gérer la temporalité des données. Certains travaux proposent différentes extensions des

modèles orientés graphe ou développent leurs propres systèmes de gestion de données graphes pour intégrer la dimension temporelle. Le constat général est qu'il n'existe pas de cadre unifié à la fois pour la modélisation et l'interrogation de données interconnectées temporelles (Sharma, Sinha, et Johnson, 2021). Afin de répondre à ces limites, nous proposons une solution conceptuelle (indépendante de tout aspect d'implantation) pour les concepteurs et les développeurs afin de modéliser et d'interroger des données interconnectées temporelles.

Notre proposition intègre un modèle conceptuel de graphe temporel de propriétés composé d'*entités* et de *relations temporelles*. Nous étendons le modèle classique en ajoutant la notion de l'état aussi bien au niveau des entités qu'au niveau des relations pour traduire leur évolution temporelle. Chaque état contient une liste d'attributs ainsi que leurs valeurs présentes pendant une période de validité. De plus, nous proposons deux opérateurs d'interrogation associés. Dans un premier temps, l'opérateur *matching_{predicate}* permet l'extraction d'un sous-graphe satisfaisant des critères de sélection et des critères temporels. Dans un deuxième temps, l'opérateur *matching_{pattern}* permet l'extraction d'un sous-graphe satisfaisant un motif donné. Contrairement aux opérateurs d'interrogation de graphes statiques, nos opérateurs permettent d'obtenir des informations sur l'évolution des données interconnectées pour découvrir des événements et/ou de comprendre leurs impacts. Enfin, nous définissons des règles de traduction directes de notre modèle vers le graphe de propriétés statique ainsi que de nos opérateurs vers les opérations de graphe de propriétés statique. Cela permet une implémentation directe dans les systèmes de gestion de graphes de propriétés.

Notre proposition a été évaluée à travers une série d'expérimentations afin de valider notre modèle et nos opérateurs. Pour ce faire, nous avons utilisé un jeu de données d'un benchmark de référence et un jeu de données réelles contenant des données interconnectées évoluant au cours du temps. Nous les avons implantés à l'aide des systèmes de gestion de données graphe Neo4j et OrientDB. Nous avons formulé des analyses temporelles incluant différents niveaux du graphe et différentes granularités temporelles basées sur nos opérateurs. Puis, nous les avons traduits dans les langages de requête de Neo4j et OrientDB. Nous avons ensuite lancé nos requêtes dans les deux systèmes de gestion de données. Nous avons ainsi démontré que nos opérateurs peuvent être utilisés pour exprimer des analyses de cas réels. De plus, ils sont implantables sans développements spécifiques sous Neo4j et OrientDB. La prochaine étape de nos travaux se concentre sur la découverte de connaissances dans le graphe pour permettre une analyse explicative des évolutions dans les données interconnectées.

Bibliographie

Angles R. « The Property Graph Database Model ». In *Proceedings of the 12th Alberto Mendelzon International Workshop on Foundations of Data Management, Cali, Colombia, May 21-25, 2018*.

Debrouvier A., Parodi E., Perazzo M., Soliani V., Vaisman A. « A model and query language for temporal graph databases ». *The VLDB Journal*. septembre 2021. Vol. 30, n°5, p. 825-858.

Sharma C., Sinha R., Johnson K. « Practical and comprehensive formalisms for modeling contemporary graph query languages ». *Information Systems*. 2021. p. 101816.

Alignement incrémental des données dans des environnements dynamiques

Oumaima El Haddadi^{1,2,3}, Max Chevalier¹, Bernard Dousset¹, Ahmad El Allaoui², Anass El Haddadi², Olivier Teste¹

1. IRIT, SIG, Université de Toulouse, CNRS, France

2. LSA, SDIC, Université Abdelmalek Essaadi Tetouan, Maroc

3. Oumaima.ElHaddadi@irit.fr

RESUME. De plus en plus de Systèmes d'Informations (SI) sont appliqués dans des environnements dynamiques. Dans de tels environnements, au-delà de la problématique de l'hétérogénéité des données, s'ajoute le problème de la dynamique des données voire des besoins en informations. Une des réponses apportées pour gérer l'hétérogénéité des données repose sur la notion d'alignement des données. L'objectif de l'alignement est d'identifier des correspondances entre des ensembles de données. Pour considérer ces environnements dynamiques dans leur globalité, une réflexion doit être menée autour des ressources nécessaires pour mettre à jour les correspondances entre les sources de données existantes au fil du temps. Dans cet article, nous nous intéressons particulièrement à la dynamique des sources afin d'identifier une démarche incrémentale d'alignement des données visant à limiter ces ressources.

ABSTRACT. Today, information systems (IS) increasingly applied in dynamic environments. Within such environments, the problem of data heterogeneity is compounded by the problem of dynamic data and information needs. One of the answers to manage data heterogeneity is based on the notion of data alignment. The objective of alignment is to identify correspondences between data sets. To consider these dynamic environments in their entirety, a reflection must be made around the resources needed to update the correspondences between existing data sources over time. In this paper, we focus on the dynamics of the sources in order to identify an incremental approach to data alignment aimed at limiting these resources.

Mots-clés : Alignement incrémental des données, Graphe, Plongement, Environnements Dynamiques

KEYWORDS: Incremental Data Alignment, Graph, Embedding, Dynamic Environment

1. Introduction

Dans des environnements dynamiques, les sources de données sont multiples, variées (les sources de données sont hétérogènes, chaque source a un format et une structure différente) et elles-mêmes dynamiques (les sources de données évoluent avec le temps). D'après (Velegarakis et al., 2003), dans ces environnements « data

sources may change not only their data but also their schemas, their semantics, and their query capabilities ». Ceci implique que les SI doivent considérer les caractéristiques de ces sources de données pour pouvoir répondre au mieux aux besoins en informations qui eux aussi peuvent être dynamiques. Cela signifie que, dans certains contextes (tels que l'aide à la décision, la recherche ou la découverte/exploration de données), les besoins peuvent évoluer au gré des jours et/ou des utilisateurs limitant ainsi l'usage de méthodes manuelles ou semi-automatiques. Ces caractéristiques ont d'autant plus d'impact que les besoins en informations sont critiques comme c'est le cas dans le domaine de l'aide à la décision. Cependant, dans cet article, par soucis de place, nous limitons notre discussion à la dimension dynamique des sources de données. Pour répondre au problème de l'hétérogénéité des données, l'alignement des données (ou « schema matching » en anglais) est essentiel. En effet, l'alignement des données apparaît comme un « passage obligé » pour tout processus visant à exploiter et manipuler les données dans de tels environnements. Le résultat de l'alignement est ainsi exploité à la fois dans les approches de stockage centrées besoins – i.e. basées sur l'intégration de données (ou « schema mapping » en anglais) (Miller et al., 2000), mais également dans les approches non centrées besoins telles que les lacs de données (Alserafi et al. 2020) ou encore dans le domaine des ontologies (Aumuller et al., 2005).

De nombreuses méthodes d'alignement des données existent dans la littérature. Cependant, dans le cadre des environnements dynamiques, ces alignements doivent être mis à jour au fur et à mesure de l'évolution des sources. Or, maintenir ces alignements à jour nécessite de nombreux calculs. C'est dans cette optique que nous souhaitons définir une démarche incrémentale d'alignement des données visant à réduire autant que possible les calculs nécessaires. La structure de cet article est la suivante : la section 2 présente l'alignement des données et les évolutions basées sur des graphes dans lesquels la représentation des éléments de ces graphes est sous la forme de plongements (« embedding » en anglais). Sur cette base, nous proposons enfin, en section 3, une démarche générale d'alignement incrémental des données avant de conclure.

2. L'alignement des données : une réponse à l'hétérogénéité

Les données hétérogènes sont toutes les données présentant une variabilité par exemple de types et/ou de formats. Les données peuvent être des tables de base de données, des fichiers XML, des graphes, des ontologies, des séquences de caractères ou tout autre type de données. Au regard de ces multiples exemples, on peut souligner l'hétérogénéité structurelle ou hétérogénéité des schémas. On peut également souligner l'hétérogénéité au niveau de la donnée elle-même (le type, la langue, la valeur, etc.) – via les instances – qui est qualifiée d'hétérogénéité des données. Une classification des approches d'alignement ainsi que la dimension que ces approches considèrent dans le calcul de correspondance est présentée dans (Bernstein et al., 2011). Les approches d'alignement peuvent être notamment linguistiques, syntaxiques, à base de connaissances externes (ontologies,

dictionnaires, ...), sémantiques ou encore hybrides. Ces approches peuvent s'appliquer soit au niveau des éléments des schémas (e.g. de la structure) soit au niveau des instances (e.g. des valeurs). Plus récemment, les méthodes d'alignement des données, afin d'améliorer notamment les alignements sémantiques, intègrent les travaux autour de l'apprentissage automatique et notamment le traitement du langage naturel. On peut citer par exemple les travaux de (Zhang et al. 2021, Merieme et al., 2022) qui reposent sur des réseaux de neurones profonds.

On peut également citer les travaux intégrant des approches d'apprentissage de représentation (Bengio et al., 2013) et notamment les plongements (e.g. « embedding »). Un plongement est une représentation vectorielle d'éléments (attributs d'un schéma, instance, ...). Ces plongements simplifient le calcul de correspondances puisqu'ils permettent d'utiliser des méthodes classiques de distances car ce sont des vecteurs numériques « classiques ». Dans le cadre de l'alignement des données, ils sont aujourd'hui utilisés pour représenter, souvent dans des espaces réduits et à différents niveaux de granularité, des bases de données ou encore des graphes (Hättasch et al., 2020, Zhao et Castro Fernandez, 2022). Au-delà de l'intégration de ces approches basées sur l'apprentissage automatique, nous pouvons souligner une orientation de certains travaux qui, pour calculer les alignements, transforment, dans un premier temps, les données dans un graphe pour ensuite appliquer des mesures de correspondance sur ce dernier. Dans ce cadre, (Melnik et al., 2002) proposent une approche basée sur les graphes qui intègre à la fois la structure et les instances pour identifier les correspondances.

De manière combinée, des travaux tels que REMA (Koutras et al., 2020) proposent également la transformation des sources de données (structure et instances) sous forme d'un graphe ainsi que l'usage des plongements pour calculer les alignements. Il est à noter que ce type d'approche (graphe + plongements) trouve également des applications dans l'alignement de données via des graphes de connaissances (Ayala et al., 2022). L'intérêt des approches d'alignement des données basées à la fois sur les graphes et les plongements, de notre point de vue, réside tout d'abord dans le fait que le graphe permet d'homogénéiser et peut rendre dynamique la représentation des éléments des différentes sources de données (une base de données relationnelles ou un fichier CSV par exemple). Dans le même temps les plongements permettent d'uniformiser à la fois l'espace de représentation des éléments des schémas et/ou des instances à aligner et le calcul des alignements. Cependant, les méthodes d'alignement des données présentées dans cette section, à notre connaissance, ne prennent pas en compte la notion de dynamique des sources, c'est-à-dire l'évolutivité des schémas et des instances de manière incrémentale pour optimiser les ressources nécessaires. De ce fait, lors de l'évolution d'une source de données, il sera nécessaire de recalculer les plongements ou stocker les informations qui ont permis de les calculer (les « random walk » par exemple). Cela peut ainsi amener à une utilisation très importantes de ressources (stockage/calcul) surtout lorsque le nombre de sources augmente. Notre objectif est donc de réduire cette utilisation de ressources en mettant à jour autant que possible les plongements de façon incrémentale.

3. Démarche d'alignement incrémental des données

L'alignement incrémental des données vise à intégrer l'évolution des données (schéma/instances) dans le processus d'alignement des données. On doit ainsi considérer l'ajout, la suppression et la mise à jour des éléments décrivant les données (schéma/instances) ainsi que leurs relations. Nous n'avons pas identifié dans la littérature de travaux intégrant la dimension incrémentale pour l'alignement des données tel que défini ci-dessus. À l'inverse, indépendamment de l'alignement des données, nous avons identifié des travaux relatifs au calcul de plongements dans des graphes dynamiques. Deux types d'approches sont identifiées dans (Bielak et al., 2020) : les approches « online » qui mettent à jour les représentations des éléments à chaque modification et les approches « incrémentales » qui font la mise à jour par « lot ». Notre définition d'alignement incrémental des données englobe, à ce niveau de discussion, ces deux temporalités car, souhaitant répondre aux enjeux de mise à jour incrémental des représentations, notre proposition pourrait être utilisée dans ces deux cas. Parmi l'ensemble des approches proposées la plus proche de nos problématiques est sans doute la méthode Online-Node2Vec proposée dans (Béres et al., 2019). Cette méthode vise à mettre à jour le plongement correspondant à chaque nœud d'un graphe au fur et à mesure des modifications. Cependant, la mise à jour des plongements repose notamment sur une fenêtre temporelle excluant les relations existantes dans le graphe les plus « anciennes » ce qui n'est pas pertinent dans le cadre de l'alignement de données. Enfin, Online-Node2Vec vise à faire évoluer la représentation d'un nœud suite à l'ajout d'un nœud et d'une nouvelle relation. Les auteurs ne traitent pas de l'ensemble des mises à jour possibles contrairement à notre objectif.

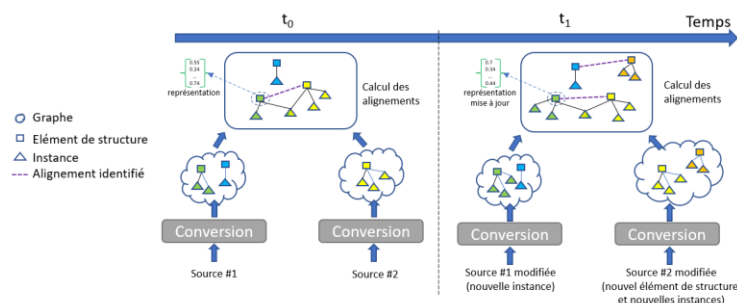


Figure 1 – Illustration générale d'alignement incrémental des données. À t_1 , une nouvelle instance est ajoutée dans la source #1 impliquant une mise à jour de la représentation de l'élément de structure correspondant et un recalcul des alignements. Par ailleurs, l'ajout d'un nouvel élément de structure et de nouvelles instances dans la source #2 implique le calcul de la représentation correspondante ainsi que le calcul de l'ensemble des alignements possibles (un nouvel alignement été identifié avec un élément de structure de la source #1).

Sur la base de ces travaux, l'illustration d'alignement incrémental des données que nous proposons (Figure 1) repose sur des plongements dans un graphe traduisant

la structure et/ou les instances des sources à aligner. Dans un premier temps, la méthode nécessite la transformation de chaque source de données en un sous-graphe via une méthode de transformation adaptée au format original de la source. Des méthodes de conversion entre différents types de données sont disponibles (https://bit.ly/M2_GM (visité le 03/03/2023)) et peuvent être adaptées. Tout comme dans REMA, nous souhaitons intégrer à la fois la structure et les instances pour identifier les alignements car ces deux niveaux permettent d'augmenter le nombre de nœuds communs potentiels entre les deux sources et ainsi obtenir des représentations plus précises. Avant de calculer les alignements, pour chaque élément de structure, nous calculons leur représentation (i.e. un plongement). Cependant, le défi de notre méthode qu'il nous faut remporter est d'identifier la méthode de plongement la plus adaptée à la dimension incrémentale. Nous recherchons donc actuellement une méthode de représentation des nœuds supportant nativement ou après adaptation l'ensemble des opérations de mise à jour que ce soit en ajout comme c'est le cas dans Online-Node2Vec mais également en suppression/modification. A titre d'illustration, la Figure 1 ne présente qu'un exemple de modifications des sources existantes car l'ajout de nouvelles sources nécessite le calcul exhaustif des représentations correspondantes à cette nouvelle source. Nous avons donc aujourd'hui certaines bases qui malheureusement ne traitent pas, sauf erreur de notre part, l'ensemble du problème. Après avoir calculé ou mis à jour les représentations concernées par les modifications, les alignements sont calculés sur la base de ces dernières. Pour optimiser ce calcul nous privilégions l'usage d'une méthode de hachage, e.g. MinHash (Broder et al., 2000) ou d'une matrice de similarités.

4. Conclusion et Perspectives

L'exploitation des données dans des environnements dynamiques nécessite à la fois la prise en compte de l'hétérogénéité mais également la dynamique des données et des besoins. Pour optimiser l'alignement des données nous proposons les premières bases d'une démarche incrémentale reposant sur un graphe traduisant les sources de données et dont chaque nœud est représenté par un plongement. Ces plongements doivent supporter une mise à jour incrémentale et/ou ne pas nécessiter trop de ressources pour leur calcul (stockage/calcul). Les objectifs de notre démarche visent à limiter le coût des différentes mises à jour nécessaires pour maintenir les alignements à jour. Notre travail actuel a pour objectif d'affiner cette démarche en identifiant la bonne méthode de représentation des nœuds du graphe afin de la rendre opérationnelle et d'évaluer celle-ci. Concernant la validation, nous souhaitons réaliser des expérimentations sur des jeux de données de différents domaines, ayant différentes caractéristiques, afin d'évaluer les différents cas liés à la dynamique des données. A moyen terme, vu que l'orientation actuelle de nos travaux s'intéresse particulièrement sur les plongements pour représenter les données, nous n'excluons pas d'étudier d'autres approches potentiellement intéressantes et optimisables pour identifier les alignements et ce de façon incrémentale. A plus long terme, nous souhaitons considérer la problématique de la dynamique des besoins.

Bibliographic

- Alserafi, A., Abelló, A., Romero, O., Calders T., 2020. Keeping the Data Lake in Form: Proximity Mining for Pre-Filtering Schema Matching. *ACM Trans. Inf. Syst.* 38, 3.
- Aumueller, D., Do, H-H., Massmann, S., Rahm, E., 2005. Schema and ontology matching with COMA++. *ACM international conference on Management of data (SIGMOD '05)*. Association for Computing Machinery, New York, NY, USA, 906–908.
- Ayala, D., Hernández, I., Ruiz, D., Rahm, E., 2022. LEAPME: Learning-based Property Matching with Embeddings. *Data & Knowledge Engineering*, Volume 137.
- Bengio, Y., Courville, A., Pascal, V., 2013. Representation Learning: A Review and New Perspectives. *IEEE trans. on pattern analysis and machine intelligence.* 35. 1798-1828.
- Béres, F., Kelen, D.M., Pálovics, R., Benczúr, A.A., 2019. Node embeddings in dynamic graphs. *Applied Network Science* 4, 64. 10.1007/s41109-019-0169-5
- Bernstein, P., Jayant, M., Rahm, E., 2011. Generic Schema Matching, Ten Years Later. *PVLDB.* 4. 695-701. 10.14778/3402707.3402710.
- Bielak, P., Tagowski, K., Falkiewicz, M., Kajdanowicz, T., Chawla, N.V., 2020. FILDNE: A Framework for Incremental Learning of Dynamic Networks Embeddings. *Knowledge-Based Systems*, 236. doi.org/10.1016/j.knosys.2021.107453
- Broder, AZ, Charikar, M, Frieze, AM, Mitzenmacher, M, 2000. Min-wise independent permutations. *Journal of Computer System Science* 60(3):630–659.
- Hättasch, B., Truong-Ngoc, M., Schmidt, A., Binnig, C., 2020. It's AI Match: A Two-Step Approach for Schema Matching Using Embeddings. *2nd International Workshop on Applied AI for Database Systems and Applications (AIDB'20)*.
- Koutras, C., Fragkoulis, M., Katsifodimos, A., Lofi, C., 2020. REMA: Graph Embeddings-based Relational Schema Matching. *EDBT/ICDT Workshops*.
- Melnik, S., Garcia-Molina, H., Rahm, E., 2002. Similarity flooding: a versatile graph matching algorithm and its application to schema matching. *Proceedings 18th International Conference on Data Engineering*, San Jose, CA, USA, 2002, pp. 117-128.
- Merieme, E.A., Mohamed, A., Ali, C., Fakhri, Y., Noredine, G., 2022. Schema Matching Based On Deep Learning Using LSTM Model, *3rd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, pp. 1–5.
- Miller, RJ, Haas, LM, Hernandez, MA, 2000. Schema Mapping as Query Discovery. *Very Large DataBase conference (VLDB)*, pp 77–88.
- Velegarakis, Y., Miller, R. J., Popa, L., 2003. Mapping adaptation under evolving schemas. In *Proceedings of the 29th VLDB'03*, Vol. 29. VLDB Endowment, 584–595.
- Zhang, J., Shin, B., Choi, J.D., Ho, J.C., 2021. SMAT: An Attention-Based Deep Learning Solution to the Automation of Schema Matching. *Advances in Databases and Information Systems*, LNCS. Springer International Publishing, pp. 260–274.
- Zhao, Z., Castro Fernandez, R., 2022. Leva: Boosting Machine Learning Performance with Relational Embedding Data Augmentation, in: *Proceedings of the 2022 International Conference on Management of Data*. ACM, Philadelphia PA USA, pp. 1504–1517.

Un moteur de recherche d'événements pour explorer la presse numérique ou historique

Guillaume Bernard¹, Thomas Blot²

1. Laboratoire L3i

Université de La Rochelle, France

guillaume.bernard@univ-lr.fr

2. Université de Bordeaux, France

thomas.blot@u-bordeaux.fr

RÉSUMÉ. La presse écrite rapporte tous les jours des événements. Faciliter l'exploration de cette presse est particulièrement pertinent pour les scientifiques en humanités numériques qui en font leur matériau principal. Alors que les approches par clustering sont le sujet majeur de la littérature scientifique, nous présenterons à la conférence un moteur de recherche d'événements contraint par des critères de sobriété énergétique et à la latence faible.

ABSTRACT. The written press reports events on a daily basis. Facilitating the exploration of this press is particularly relevant for digital humanities scientists who use it as their primary material. While text-based approaches are the main focus of the scientific literature, we will present at the conference a low-latency, energy-efficient event search engine.

MOTS-CLÉS : Ingénierie des données liées, Événements, Fouille dans les données massives

KEYWORDS: Linked Data, Events, Data mining

1. Introduction

La presse rapporte au quotidien les événements qui surviennent dans le monde. Les articles, une fois publiés, ancrent les événements dans le temps et l'espace pour former une chronologie qui s'achève à la disparition médiatique de l'événement. Établir cette chronologie *a posteriori*, même en se focalisant uniquement sur la presse écrite, est un problème ouvert. Nous ajoutons la contrainte d'exiger une réponse rapide de quelques secondes, qui ne nécessite pas d'apprentissage. Ces contraintes s'inscrivent dans le cadre de la sobriété numérique qui est devenue un enjeu sociétal majeur. Nous émettons l'hypothèse que se passer d'apprentissage réduit les coûts de calcul et les temps de traitement. Pourtant, répondre à la question *comment regrouper toutes les connaissances, articles et documents relatifs à un événement passé?* est d'un grand intérêt pour la recherche en Sciences Humaines et Sociales ou le journalisme. Rechercher les

articles liés à un événement A pour établir une chronologie des éléments rapportés par les médias à ce sujet est en effet une tâche cruciale pour ces disciplines.

Afin de répondre à cette problématique et permettre la création de corpus fiables pour ces publics variés, nous modélisons des *histoires* sur les événements passés. En section 2, nous introduirons la modélisation des événements puis, en section 3, nous présenterons notre moteur de recherche d'événements.

2. Modéliser les chronologies d'événements

Le suivi d'événements est exploré par divers projets depuis les années 2000. Ils utilisent des algorithmes de *machine learning* qui reposent sur un pré-traitement du texte, généralement un encodage *TF-IDF* (Miranda *et al.*, 2018) ou dense (Linger, Hajaiej, 2020). La date et le texte des articles sont pris en compte pour créer des *clusters* d'articles qui décrivent le même événement.

Dans l'analyse d'événements passés, certaines propriétés sont connues d'avance : date, lieu de l'événement ainsi que les entités impliquées. Ces connaissances *a priori* existent dans des bases ouvertes comme Wikidata ou Wikipédia. (Bernard *et al.*, 2021) ont développé un outil, *wikivents*, qui analyse ces bases pour extraire ces informations : date, lieu et entités nommées concernées (personnes, lieux, organismes). (Wang *et al.*, 2022) ont proposé une architecture d'indexation pour requêter des articles de presse et extraire des informations pour une tâche de *question answering*. Nous suggérons d'associer ces deux stratégies pour : 1. extraire les connaissances sur les événements ; 2. utiliser ces connaissances pour créer des requêtes.

3. Architecture du système

Notre architecture se base sur trois éléments : les données de presse à explorer, un outil de description d'événements, ici *wikivents* et une infrastructure pour indexer les documents. Pour l'indexation, nous déployons *ElasticSearch*, comme (Wang *et al.*, 2022). Nous extrayons, pour chaque document, son titre et son texte ainsi que les lemmes, les termes et les entités nommées dudit texte. Ces éléments sont utilisés en complément du texte brut pour effectuer les recherches. En entrée du système, un identifiant d'événement sur Wikidata est fourni par l'utilisateur, comme montré en figure 1. Il ou elle peut choisir un événement prédéterminé ou indiquer l'identifiant d'un événement qu'il a repéré sur Wikidata. Les articles étant rédigés dans une langue unique, il convient de sélectionner également une langue cible. *wikivents* extrait les entités, dates ou noms de lieux relatifs à cet événement en explorant à la fois Wikidata et les résumés des pages Wikipédia associées aux événements. La représentation ainsi obtenue est considérée comme agnostique à la langue, car elle ne se base que sur des identifiants de concepts dans la base Wikidata. Par exemple, le lieu « Saint Petersbourg » est identifié par *Q656*. Une seconde étape permet de convertir cette représentation dans une langue cible. Nous utilisons l'ensemble des noms dans une langue donnée utilisés pour décrire les entités sur Wikidata. Une fois la description

des événements ainsi obtenue, elle est transformée par un second outil qui forge une requête pour *ElasticSearch*. La requête est adaptée à la recherche de documents dans une unique langue. Le travail soumis dans cette démonstration se positionne donc à l'interface entre *wikivents* et *ElasticSearch*, tel que présenté en figure 1.

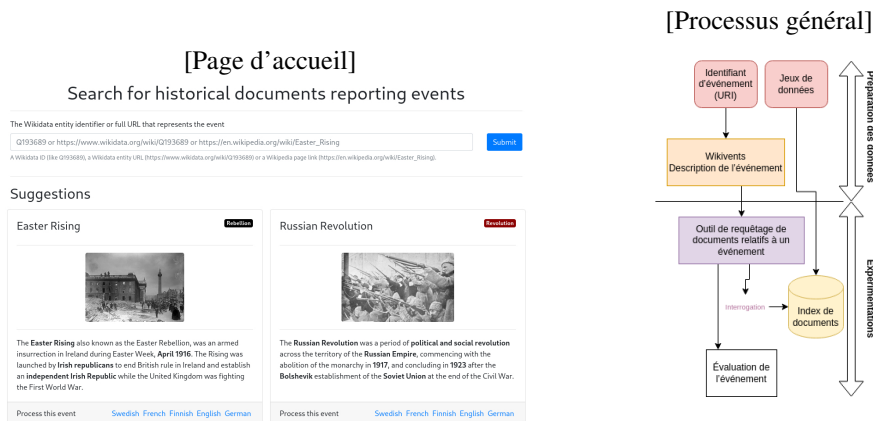


FIGURE 1. Page d'accueil et schéma décrivant le fonctionnement du système.

Une fois l'événement analysé et la requête construite, celle-ci est envoyée à *ElasticSearch* qui à son tour recherche les documents et les affiche. Ce sont les articles de presse qui font mention de l'événement. Une fois ces résultats obtenus, l'utilisateur dispose des articles correspondants (figure 2). Pour chaque événement, un court résumé est affiché et l'ensemble des entités détectées - qui ont contribué à la recherche - sont listées. Elles sont dites participantes : elles décrivent l'événement. Les documents connus et rapportant l'événement sont listés et l'image source, si disponible, est affichée. Ici, le texte est extrait d'images par reconnaissance optique de caractères. Les données présentées sont des articles de presse en français publiés par la Bibliothèque Nationale de France et distribués *via* la plate-forme du projet NewsEye¹.

4. Résultats et conclusion


Nous avons comparé notre moteur de recherche à des techniques de *clustering* qui remplissent les mêmes objectifs (Miranda *et al.*, 2018). Nous avons utilisé le jeu de données de ces derniers, composé d'articles en anglais, allemand et espagnol (). Chaque article est associé à un identifiant d'événement qui nous permet d'identifier tous les articles de presse qui évoquent un événement commun. Nous avons comparé la *F1* du *clustering* par *machine learning* avec la *F1* de notre moteur de recherche. Les

1. <https://www.newseye.eu>

assassinat de Raspoutine

Wikidata ID [Q2882749](#)
Rendering in [fr](#)
Date [1916-12-30](#)
Processed languages: [\(es, fr\)](#)

L'assassinat de Raspoutine aurait été perpétré par le prince Félix Ioussouпов, le grand-duc Dimitri Pavlovitch, le député Vladimir Pourichkevitch, le lieutenant Sergueï Soukhotine et le docteur Stanislas Lazover, à Petrograd dans la nuit du 16 décembre 1916 (29 décembre 1916 dans le calendrier grégorien) au 17 décembre 1916 (30 décembre 1916 dans le calendrier grégorien). Le récit du prince Ioussouпов à propos des mobiles de l'assassinat, variable au cours de sa longue vie, semble aujourd'hui inexact. Les dernières recherches sur ce sujet s'orientent vers une liquidation voulue par les services secrets des Alliés pour éviter que le tsar Nicolas II renonce à son engagement dans le conflit de la Première Guerre mondiale....



- The event took place in [Saint-Petersbourg](#); [palais Ioussouпов](#); [Russie](#);
- The event is a [assassinat politique](#);

Entities found in processed summaries 4 entities found in lead sections 2 processed languages

<p>Geo-Political entities</p> <ol style="list-style-type: none"> Saint-Petersbourg (*2) 	<p>People</p> <ol style="list-style-type: none"> Grigori Raspoutine (*2) Felix Ioussouпов (*2) Vladimir Pourichkevitch (*2) 	<p>Organizations</p>
--	--	----------------------

Associated documents 239 documents found

<p>Article n° Leseure_12148-bpt6q46148438_article-242 issued on 1917-02-16</p> <p>Petrograd, 15 février. — Selon des renseignements parvenus à Petrograd, les tendances antiallemandes continuent à se répandre avec une nouvelle intensité en Pologne. On</p>	<p>Petrograd, 15 février. — Selon des renseignements parvenus à Petrograd, les tendances antiallemandes continuent à se répandre avec une nouvelle intensité en Pologne. On</p>
--	---

FIGURE 2. Recherche des documents évoquant l'événement. Un résumé de l'événement est présenté et les articles trouvés sont listés plus bas.

données et métriques utilisées ont été décrites par (Bernard, 2022) et les références vers les jeux de données y sont partagées.

Nous avons analysé le jeu de données Miranda *et al.* à la recherche des événements qui sont décrits par un nombre élevé d'articles. Nous avons retenu 81 événements, soit plus de 7 000 documents sur les plus de trente mille qui composent le corpus. Les événements sont de différents types : judiciaires, conflits, sportifs ou politiques, pour ne citer que les plus importants. Pour chacun de ces 81 événements, nous avons recherché leur identifiant Wikidata manuellement en parcourant à la fois les articles et l'encyclopédie. Cette association faite, il nous est possible d'évaluer le processus inverse et de vérifier que le moteur de recherche associe adéquatement des identifiants d'événements aux documents qui les décrivent.

Nous évaluons la capacité des algorithmes de *clustering* ou le moteur de recherche à grouper correctement les documents selon qu'ils rapportent les mêmes événements. Nous obtenons des métriques de *clustering* de précision, de rappel et de F-mesure, et

dans chaque langue, puisque le moteur de recherche fonctionne sur un principe monolingue. Puisque la valeur de précision et de rappel dépend du nombre de documents renvoyés par le moteur de recherche (plus le nombre de documents récupéré augmente, plus la précision décroît), nous avons, par analogie, considéré que la F-mesure maximale représente l'expertise du scientifique en humanités numériques. Par exemple, il ou elle sait déterminer grâce à ses connaissances à partir de quel moment les documents renvoyés ne sont plus pertinents au regard de la requête initiale. C'est cette valeur de F-Mesure maximale que nous utilisons pour évaluer un événement unique. La F-Mesure pour une requête (évaluée donc sur 81 événements) est la médiane de cet ensemble de F-Mesures par événement.

À contexte comparable (mêmes données, même langue, mêmes événements), nous obtenons pour le moteur de recherche une F-Mesure de 0,67 en anglais. Elle est de 0,91 pour les méthodes basées sur du *machine learning*. La différence de 0,24 point peut s'expliquer par l'absence de données suffisamment descriptives sur Wikidata et Wikipédia. Cela entraîne une mauvaise représentation des événements (dates, lieux ou participants manquants). Une autre cause possible est la redondance de certains événements ayant lieu annuellement (ex: Roland Garros) et présents dans le jeu de données testé pour lesquels les données Wikidata sont peu fiables ou incomplètes, du fait de cette redondance. Compte tenu de la sobriété de la méthode, les résultats sont cependant encourageants et peuvent être obtenus rapidement, sans investissement important. À ce titre, cette méthode et ces outils peuvent être appliqués à de nombreux types de documents et de nombreuses situations de recherche impliquant la création de corpus.

Remerciements

Les auteurs souhaitent remercier particulièrement Cyrille Suire pour sa relecture attentive ainsi que Cyril Faucher et Antoine Doucet pour les conseils scientifiques qu'ils ont apporté durant l'élaboration de ce projet.

Bibliographie

- Bernard G. (2022). *Détection et suivi d'événements dans des documents de presse historique*. Thèse de doctorat non publiée, Université de La Rochelle (ULR), La Rochelle.
- Bernard G., Suire C., Faucher C. *et al.* (2021). A Comprehensive Extraction of Relevant Real-World-Event Qualifiers for Semantic Search Engines. In *Proceedings of the 25th International Conference on Theory and Practice of Digital Libraries*, p. 153–164.
- Linger M., Hajaiej M. (2020). Batch Clustering for Multilingual News Streaming. In *Proceedings of Text2Story - ECIR*.
- Miranda S. a., Znotiņš A., Cohen S. B. *et al.* (2018). Multilingual Clustering of Streaming News. In *2018 Conference on EMNLP*, p. 4535–4544.
- Wang J., Jatowt A., Yoshikawa M. (2022). *ArchivalQA: A Large-scale Benchmark Dataset for Open Domain Question Answering over Historical News Collections*.

Exploration des Modèles de Processus Contextuels à l'Aide des Données de Capteurs : un Cas d'Activités Quotidiennes dans une Maison Intelligente

Ramona Elali¹, Elena Kornyshova², Rébecca Deneckère¹, Camille Salinesi¹

1. Paris 1 Panthéon Sorbonne, Paris, France
{ramona.elali, rebecca.deneckere, camille.salinesi}@univ-paris1.fr

2. Conservatoire National des Arts et Métiers, Paris, France
elena.kornyshova@cnam.fr

*REFERENCE DE L'ARTICLE INTERNATIONAL. Cet article est un résumé de l'article :
Ramona Elali, Elena Kornyshova, Rebecca Deneckere, Camille Salinesi. Mining Contextual Process Models Using Sensors Data: a Case of Daily Activities in Smart Home. 19th European Mediterranean & Middle Eastern Conference on Information Systems (EMCIS 2022), Dec 2022, Online, Dubai, United Arab Emirates. (hal-03884131)*

MOTS-CLES : Fouille de processus, Modèle de processus, Contexte, Modèle de processus contextuel, Log de capteurs, Log d'évènements, Maison intelligente

KEYWORDS : Process Mining, Process Model, Context, Contextual Process Model, Sensors Log, Events Log, Smart Home

1. Introduction

La transformation rapide des systèmes et l'augmentation du volume des données disponibles constituent une source de nouvelles informations. La fouille des processus (Aalst, 2016) permet de découvrir des modèles de processus en s'appuyant sur les logs d'activités. Néanmoins, elle ne prend pas en considération le contexte qui englobe l'activité effectuée par l'utilisateur et ne permet de découvrir un modèle de processus qu'en se basant uniquement sur les logs d'activités, alors qu'il existe de nombreux facteurs externes tels que le temps, l'emplacement, le profil, etc. qui peuvent affecter le choix de l'activité, alors qu'aujourd'hui, il est facile d'obtenir des informations sur l'environnement contextuel d'une activité en se basant sur des données de capteurs.

2. Proposition

Notre objectif est de faire évoluer les modèles de processus par la sémantique issue des données de capteurs afin de construire un modèle de processus contextuel. Nous travaillons sur un ensemble de données d'une maison intelligente pour vérifier comment les informations contextuelles peuvent affecter les activités des utilisateurs pour construire un modèle de processus contextuel amélioré. Les logs des données viennent du défi BP-Meets-IoT (Koschmider *et al.*, 2021) et contiennent des données sur la vie de tous les jours et les activités domestiques de deux personnes. Il est composé de deux logs de données : un log de données qui se compose d'un log d'activité d'un seul habitant à l'intérieur d'un logement avec le log de capteurs correspondant, et un second qui se compose d'un log d'activité de deux habitants également avec leurs logs de capteurs. Tous les logs ont été fournis au format XES.

Nous supposons qu'une personne a tendance à avoir des activités variables, sans se tenir forcément à un horaire ou à une routine fixe et que la personne agira ou ajustera ses activités et tâches quotidiennes en fonction de son environnement. Par exemple, une personne pendant le week-end effectuera des activités différentes de celles des jours de la semaine. De plus, lorsque c'est un jour de pluie, la personne fera de l'exercice sportif à l'intérieur de la maison tandis que lorsqu'il fait beau, elle pourra sortir se promener ou courir dehors. Nous constatons donc qu'il est possible d'identifier les liens entre les données contextuelles et les activités des utilisateurs. Par conséquent, nos questions de recherche sont : (1) L'environnement contextuel affecte-t-il les activités des utilisateurs ? (2) Les liens entre les données des capteurs et les logs d'activités peuvent-ils être automatiquement identifiés ?

Une fois les données étudiées, nous avons pu identifier automatiquement et manuellement les liens entre les capteurs et les activités. Nous avons aussi illustré la différence entre les modèles de processus produits chaque fois que les informations contextuelles changent : (jours de la semaine Vs week-ends) ou le profil de l'utilisateur (ressource 1 Vs ressource 2). La diversité des différents modèles de processus exploités, chacun lié à un contexte spécifique, suggère l'importance de construire des modèles de processus contextuels. De nombreuses autres informations contextuelles peuvent également fournir une meilleure connaissance telles que l'heure, le lieu, le pays, la culture, la météo, etc. L'utilisation des modèles de processus contextuels nous permet d'offrir de meilleures recommandations aux utilisateurs en suggérant la meilleure activité la plus adaptée à un moment et à un endroit précis selon le contexte en cours.

Bibliographie

- Aalst W.M.P. (2016). *Process Mining: Data Science in Action*. Springer, Heidelberg ISBN : 978-3-662-49850-7.
- Koschmider A., Leotta F., Serral E., Torres V. (2021). BP-Meets-IoT 2021 Challenge Dataset.

Du Concept de Vie Intelligente (Smart Life) à celui de l'Ingénierie de la Vie Intelligente (Smart Life Engineering)

Elena Kornyshova¹, Rebecca Deneckère², Eric Gressier-Soudan¹,
John Murray³, Sjaak Brinkkemper⁴

¹ Conservatoire National des Arts et Métiers, Paris, France
elena.kornyshova@cnam.fr, eric.gressier_soudan@cnam.fr

² Université Paris 1 - Pantheon-Sorbonne, Paris, France
rebecca.deneckere@univ-paris1.fr

³ San José State University, California, USA
john.murray@sjsu.edu

⁴ Utrecht University, Utrecht, Netherlands
S.Brinkkemper@uu.nl

*RÉFÉRENCE DE L'ARTICLE INTERNATIONAL. Cet article est un résumé de l'article :
Elena Kornyshova, Rebecca Deneckere, Eric Gressier-Soudan, John Murray, Sjaak
Brinkkemper. From Smart Life to Smart Life Engineering: a Systematic Mapping Study and
Research Agenda, 15th International Baltic Conference on Digital Business and Intelligent
Systems (DB&IS), Jul 2022, Riga, Latvia*

*MOTS CLES : Vie intelligente, Ingénierie Smart Life, Technologies Smart, SMS, Agenda
KEYWORDS : Smart Life, Smart Life Engineering, Smart Technologies, Systematic Mapping
Study, Research Agenda*

1. Introduction

La montée des applications et technologies Smart dans les différents domaines scientifiques exprime un besoin fort d'une compréhension plus globale de ce que l'on appelle *Smart* en tant que concept théorique fondamental : de quoi s'agit-il exactement ? Récemment, tous ces termes semblent être regroupés sous le thème général de *Smart Life*, la vie intelligente. Cependant, il n'y a pas de définition précise de ce que cela implique. De nombreux articles existent et contribuent à élargir les connaissances sur certaines parties de ce domaine, mais il n'y a toujours pas d'explication claire de ce que l'on appelle la vie intelligente et, par extension, ce qu'on peut alors appeler l'ingénierie de la vie intelligente. L'objectif principal de l'article résumé ici est de donner des définitions et de caractériser ce domaine.

2. Principales Définitions autour de la Vie intelligente (Smart Life)

Nous étudions la vie intelligente en utilisant l'approche systémique. L'élément central est ce qu'on appelle un *Smart Artefact* (terme choisi pour éviter d'être mêlé à des termes déjà établis comme Smart Object ou Smart Thing). Il correspond à tout appareil doté d'un programme et d'une capacité de stockage et de traitement des données autonome (eg. Smart Window). Les artefacts appartenant au même domaine d'application forment une *Smart Application* (eg. Smart Home). Les Smart Artefacts peuvent appartenir à deux ou plusieurs Smart Applications. La *Smart life* est composée de diverses Smart Applications.

Degré d'intelligence. Nous le définissons selon le type d'intégration et de traitement des données. Le premier axe reflète les niveaux de prise en compte des données venant de l'environnement et des utilisateurs. La complexité des données va des données des capteurs intégrés aux données de l'environnement éloigné et, enfin, aux données sur le bien-être et l'état des utilisateurs. Le deuxième axe montre l'intégration progressive de l'IA dans les processeurs. On distingue trois degrés : *smart* (capacité à traiter les données et adapter son fonctionnement), *intelligent* (capacité à traiter les données et adapter son fonctionnement à l'utilisation de l'IA), et *sentient* (capacité à traiter des données venant des utilisateurs sur leurs sentiments, son état physique, etc.).

Smart Artefact : tout artefact autonome fourni avec un processeur de données lorsque les données sont obtenues, stockées, traitées et utilisées pour déclencher des actions de l'artefact dans l'environnement. Un objet Smart peut être connecté ou non.

Smart Application : ensemble d'artefacts et de Smart Applications avec un but commun. Par exemple, une application Smart city peut être composée d'une application Smart parking, d'une application Smart energy, etc. Une application Smart doit contenir au moins deux Smart Artefacts.

Smart Life (Vie intelligente) : domaine englobant les applications et artefacts intelligents pour enrichir l'expérience personnelle, sociétale, environnementale et d'entreprise.

Smart Life Engineering : Ingénierie permettant l'utilisation de principes scientifiques pour concevoir, construire, mettre en œuvre et faire évoluer les applications Smart Life. Elle peut donc être considérée comme chevauchant le domaine de l'ingénierie des systèmes d'information, mais elle a également un large éventail de liens avec les domaines de recherche en technologie des réseaux sans fil, IA, sciences de la vie, etc.

3. Conclusion

Nous avons étudié la littérature existante sur la vie intelligente en utilisant une étude cartographique systématique (Systematic Mapping Study) sur ce domaine afin de systématiser les connaissances existantes. Il s'agit d'un très grand, et en constante croissance, ensemble de sujets Smart allant de très établis à émergents et de très pratiques à plus conceptuels. Nous croyons qu'un énorme effort scientifique devrait être fait pour optimiser ces domaines afin de rendre la vie plus intelligente. Pour nous, cet effort devrait se faire par la définition de méthodes, de technologies et d'outils communs, donc par le développement d'un nouveau domaine, Smart Life Engineering. Les principaux enjeux de recherche sont explicités dans l'article initial.