

Une classification interdisciplinaire pour l'échange et la médiation des données ouvertes de la recherche

Marcin Trzmielewski, Claudio Gnoli

▶ **To cite this version:**

Marcin Trzmielewski, Claudio Gnoli. Une classification interdisciplinaire pour l'échange et la médiation des données ouvertes de la recherche. 12ème Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ?, Oct 2019, Montpellier, France. hal-02411446v2

HAL Id: hal-02411446
https://hal.science/hal-02411446v2

Submitted on 7 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une classification interdisciplinaire pour l'échange et la médiation des données ouvertes de la recherche

Marcin Trzmielewski

Doctorant en Sciences de l'information et de la communication

Lerass-Ceric, Université Paul-Valéry Montpellier 3

marcin.trzmielewski@etu.univ-montp3.fr

Claudio Gnoli

Bibliothécaire

Bibliothèque des Sciences et Techniques, Université de Pavie

claudio.gnoli@unipv.it

Résumé

Cette réflexion propose une évaluation de l'*Integrative Levels Classification* (ILC) en vue d'organiser des données ouvertes de la recherche en sciences humaines et sociales. Elle s'appuie sur une analyse des propriétés de l'ILC par rapport aux pratiques, usages et contexte de partage et de médiation des données. Les résultats obtenus sont satisfaisants et présentent l'ILC comme prometteuse pour la construction d'une science ouverte, dynamique et novatrice.

Mots clés

organisation des connaissances, classifications interdisciplinaires, données ouvertes de la recherche, Big Data

Title

An Interdisciplinary Classification for Exchanging and Mediation of Open Research Data

Abstract

The present paper evaluates the Integrative Levels Classification (ILC), aiming to organize research open data, in human and social sciences. We based our assessment on the analysis of the properties of the ILC in relation to the practices, uses and context of data sharing and mediation. The results are satisfactory and show that the ILC has good potentials for the construction of an open, dynamic and innovative science.

Keywords

knowledge organization, interdisciplinary classifications, open research data, Big Data

INTRODUCTION

Les données ouvertes de la recherche sont une partie des Big Data qui apparaît à la suite d'un changement de la manière de communiquer les résultats de recherches financées par les fonds publics. Ainsi, les chercheurs bénéficiant des aides à la recherche sont obligés de publier leurs données sur le Web. Il existe des infrastructures spéciales pour déposer et échanger les données telles que les banques, les archives, les portails, les entrepôts et les plateformes de données. L'objectif qu'elles portent est de rendre la science plus collaborative et démocratique pour accélérer l'innovation et développer une croissance économique (Dillaerts, 2014 ; Leonelli, 2019).

En consultant d'une manière aléatoire des infrastructures médiatisant les données en sciences humaines et sociales (SHS), répertoriées dans le registre d'entrepôts *r3data*, nous avons été frappés par le nombre de sites qui proposent une simple navigation par des mots-clés ou qui ne proposent pas d'indexation sujet. Nous avons également observé que pour organiser les données scientifiques sont encore proposées des classifications disciplinaires, alors que dans nos travaux précédents nous avons démontré que ces systèmes mettent des barrières à l'interdisciplinarité (Gnoli, 2016 ; Gnoli *et al*., 2018, Trzmielewski *et* Gnoli, 2018).

La notion d'*interdisciplinarité* concerne le transfert de méthodes et de résultats entre disciplines (Nicolescu, 1996) et, est à l'heure actuelle l'un des défis majeurs à la fois pour l'organisation des connaissances et pour la recherche académique (León Manifesto, 2007 ; Szostak *et al.*, 2016). Elle promet de relier les connaissances de différentes disciplines entre elles, en faisant découvrir une meilleure information dans l'environnement numérique, et de bâtir des collaborations entre les chercheurs de différents champs.

Le défi de recherche interdisciplinaire engage une réflexion sur des systèmes d'organisation des connaissances (SOC) plus innovants que jusqu'à présent, avec lesquels nous pourrions lier les connaissances entre elles sur le Web. Dans ce souci naissent les classifications interdisciplinaires qui reposent sur des unités élémentaires de la connaissance et qui sont fondées sur une approche synthétique, permettant de combiner infiniment les phénomènes, relations et propriétés. Elles ont déjà été tentées dans le passé avec la *Subject Classification* de James Duff Brown (Beghtol, 2004) ou avec le projet du *Classification Research Group* (1969) financé par l'OTAN dans les années 1960. Ce dernier a plus récemment été la source d'inspiration pour le projet de recherche intitulé *Integrative Levels Classification* (ILC), auquel participent des chercheurs de différents pays. Ce projet vise précisément à produire un nouveau schéma de classification interdisciplinaire. Une première édition a été publiée en 2011. La deuxième édition, comprenant plus de 10 000 classes et facettes, est en voie d'achèvement (Park *et al.*, 2019).

L'ILC est utilisée actuellement dans des bases et catalogues d'intérêt disciplinaire et multidisciplinaire. Dans le contexte d'un libre accès aux connaissances, nous nous sommes interrogés pour savoir si elle pourrait être envisagée aussi pour échanger et médiatiser les données de la recherche en SHS. Notre méthodologie consiste à confronter des fonctionnalités de l'ILC avec les pratiques, usages et contexte d'échange et de médiation des données dans ces sciences. Nous nous appuyons sur des travaux en sciences de l'information et de la communication (SIC), en organisation des connaissances (OC) et en philosophie des sciences et nous interrogeons le registre d'entrepôts *r3data*. Nous faisons aussi des simulations d'indexation, de classification et de description. Toute traduction de passages de la littérature anglophone est faite d'une manière libre.

La communication est composée en 2 parties. Le premier volet aborde les pratiques, les usages et le contexte d'échange et de médiation des données. Le deuxième volet confronte les informations recueillies avec les propriétés de l'ILC. Il est suivi par la présentation de

résultats et une discussion. La réponse à la problématique et les pistes pour le futur développement de l'ILC clôturent la réflexion.

1 – PRATIQUES, USAGES ET CONTEXTE D'ÉCHANGE ET DE MÉDIATION DES DONNÉES DE LA RECHERCHE EN SHS

Les bénéfices pour la communauté scientifique découlant du partage des données scientifiques sont nombreux. Les institutions finançant la recherche et les initiatives pour la science ouverte incitent les chercheurs à partager leurs données et tendent à faire converger les pratiques. Les chercheurs tentent en même temps de définir explicitement la notion de *données* pour établir des politiques de médiation.

1.1 Echange des données entre les chercheurs

Selon le modèle de *Data-Information-Knowledge-Wisdom* les données sont des unités élémentaires de connaissance (Frické, 2019). Dans le contexte académique, elles constituent une matière principale pour les activités quotidiennes de chercheurs, qui les exploitent pour produire et valider de nouvelles connaissances (Leonelli, 2019).

Le partage libre des données de la recherche est encouragé par les grandes institutions telles que l'OCDE (2007) et la Commission européenne (European Commission, 2016). Dans cette optique, les chercheurs qui bénéficient des aides à la recherche sont obligés de rendre leurs données disponibles sur le Web à toute personne qui souhaiterait les réutiliser. Les chercheurs en SIC, comme Dillaerts (2014), Délémontez *et* Boukacem-Zeghmouri (2015), Lyon (2016) mettent en avant deux motivations majeures pour échanger les données. Premièrement, le partage et la réutilisation des données permettent aux chercheurs d'avancer dans leur carrière, notamment d'augmenter l'impact et la visibilité de leur travail, ainsi que de trouver le matériel scientifique et de créer de nouvelles collaborations. Deuxièmement, les institutions qui financent la recherche s'attendent à ce que l'échange de données augmente la reproductibilité de la science, permette de restreindre les coûts liés à la génération des données, contribue à la croissance économique, et encourage la collaboration interdisciplinaire et internationale pour affronter de grands problèmes sociétaux.

En SHS, l'échange des données scientifiques s'inscrit dans le contexte des humanités numériques (HN), qui utilisent le potentiel fourni par des données numériques et l'informatique pour produire de nouvelles connaissances. Malgré les avantages découlant du partage des données, selon l'enquête menée par Wiley en 2014 (Ferguson, 2014), seulement 36% des chercheurs en SHS le font, ce qui donne le score bien au-dessous de la moyenne mondiale (52%) prenant en compte toutes les sciences. La moitié d'entre eux joint les données comme un matériel supplémentaire dans une revue et les dépose sur des sites personnels ou institutionnels, alors qu'un quart partage les données dans des entrepôts institutionnels. Peu de chercheurs partagent en 2014 leurs données dans des entrepôts disciplinaires (2%) et interdisciplinaires (3%), mais ils déclarent être prêts à le faire dans le futur, si cela est requis par les financeurs et augmente l'impact et la visibilité de leur travail.

En effet, comme met en avant l'étude internationale de Van den Eynden *et al*. (2016), les obligations posées par les organismes fondateurs sont le facteur le plus important pour les chercheurs en SHS pour échanger les données. L'étude confirme aussi que les chercheurs en SHS partagent et réutilisent les données moins souvent que les chercheurs d'autres sciences. Il y a plusieurs barrières freinant cette activité : la peur de mésusage et de mauvaise

interprétation, la crainte de manque des ressources financières pour préparer les données, l'affection par rapport aux données, la nature confidentielle des données sensibles et les droits de tiers.

Pour compléter cet état des lieux, nous avons collecté quelques informations auprès de *re3data*. Nous avons lancé une requête avec le terme « *Humanities and Social Sciences* » provenant de la DFG Classification of Subject Area, qui est utilisée pour le classement d'entrepôts par « sujet » dans le registre. Sur 2385 infrastructures Open Data de toutes sciences confondues, nous en avons identifié 781 qui sont dédiées aux données en SHS. Le résultat de l'interrogation de la base est surprenant, car en le comparant avec le chiffre donné par Gómez *et al*. en 2016 (397), le nombre des dispositifs a augmenté de 96,7% en 3 ans (781). En analysant ces données, il faut cependant prendre en compte qu'un entrepôt multidisciplinaire peut figurer dans plus d'une catégorie. La somme des dispositifs est donc surestimée dans les deux études, mais le pourcentage de croissance est un indicateur fiable et confirme un intérêt grandissant des institutions pour le partage des données relatives aux SHS ; l'intérêt qui ne se traduit pas à priori au nombre d'usages. Toutefois, à cause du manque de données quantitatives récentes sur le nombre de chercheurs qui déposent leurs données, nous devons rester prudents devant cette hypothèse.

1.2 Caractéristiques des données échangées

Les données scientifiques en SHS sont problématiques, à cause du manque d'accord quant à leur définition. Ce problème vient probablement de la richesse des ressources utilisées, plus hétérogènes qu'en sciences exactes et appliquées (Gómez *et al*., 2016), et car « *presque toute trace de l'activité humaine peut être considérée comme des données* » (Borgman, 2008). La confusion peut également provenir du nombre de champs en SHS. La DFG Classification of Subject Area en répertorie 52, ce qui est égal à au moins 52 communautés de pratiques diverses. A cause de la pluralité de manières de faire, un accord sur la notion de *données* semble compliqué, ce qui peut empêcher la mise en place d'une politique raisonnée de médiation.

Dans les positionnements de Borgman (2008)), Schöch (2013), Frank *et al*. (2015) et Gómez *et al*., (2016), portant sur les données en SHS, les frontières entre les données et les documents semblent s'estomper. Pour Buckland (2014), *données* n'est qu'un nouveau mot pour indiquer des documents. Hjørland (2018) souligne pourtant une différence entre les documents, les informations et les données : les documents contiennent des données sur un objet, et celles-ci délivrent des informations lorsqu'elles servent à la réalisation d'une tâche spécifique. Kaase (2001) distingue aussi les données de documents. Selon lui « *les données sont des informations sur des propriétés d'unités d'analyse* ». Nous trouvons cette définition particulièrement pertinente pour les SHS où nous enregistrons des informations sur des œuvres d'art, documents (notices bibliographiques), artistes, positions théoriques, phénomènes sociaux, etc.

Selon les résultats de l'enquête de Van den Eynden *et al*. (2016), les données les plus fréquemment générées par les chercheurs dans l'ensemble des SHS sont qualitatives, alors que les données à caractère personnel et longitudinales restent exploitées majoritairement en sciences sociales. Les auteurs soulignent aussi que les méthodes de recherche en SHS s'appuient principalement sur des enquêtes, des analyses secondaires et des observations.

Dans le registre *re3data*, nous avons observé qu'en effet les documents partagés dans les entrepôts, contenant des données relatives aux SHS, sont hétérogènes. Ce sont le plus souvent des documents bureautiques, des formats statistiques, des textes, des images, des données brutes, des données audiovisuelles, des textes structurés, des données archivées, des

graphiques structurés, des bases données, des applications logicielles, et moins souvent des codes source, des données en réseau et des données de configuration.

1.3 Stratégies de médiation

La gestion des données de la recherche est devenue de plus en plus importante dans tous les champs scientifiques. Cette évolution a été incitée par le développement des technologies permettant une recherche data-centrée et par des initiatives politico-économiques, favorisant les mouvements pour l'accès ouvert aux publications et à la science ouverte (Dillaerts, 2014 ; Gómez *et al*., 2016). Après les préoccupations liées à la gestion, le stockage et la conservation des données de la recherche, l'attention se concentre désormais sur leur libre partage et réutilisation pour créer une science plus ouverte.

Selon Van den Eynden (2018), partager les données de la recherche qui inciterait leur réutilisation nécessite une bonne gestion au niveau de la documentation, de la sécurité, de l'accessibilité, et de l'organisation. Comme souligne Odeh (2017), cette activité nécessite également de grosses infrastructures pour un traitement approprié des données tout au long de leur cycle de vie et implique la création de métiers de la curation.

Dans le souci d'une médiation incitant à la réutilisation des données, la Commission européenne (2016) a publié les *Lignes directrices pour la gestion des données FAIR*. L'objectif de ce document est d'aider « *les bénéficiaires d'Horizon 2020 à rendre leurs données de recherche trouvables, accessibles, interopérables et réutilisables pour s'assurer qu'elles sont bien gérées* ». Ceci est censé mener les chercheurs à « *la découverte des connaissances et à l'innovation, aux données ultérieures ainsi qu'à l'intégration et à la réutilisation des connaissances* ».

Pour faciliter l'échange et l'interprétation des données par l'homme et par la machine, les principes FAIR (disponibles sur : https://www.go-fair.org/fair-principles/) recommandent :

- de choisir des SOC validés par les communautés scientifiques ;
- d'utiliser des SOC représentés en formats du Web sémantique pour décrire et lier les données entre elles d'une façon interdisciplinaire. Leonelli (2019) appelle d'ailleurs cette tâche *décontextualisation* – une perte de « signification qui est attribuée [aux données] dans leur contexte [disciplinaire] de recherche initial pour les rendre compatibles avec des ensembles de données provenant d'autres sources, de sorte qu'elles soient faciles à analyser par les chercheurs qui les voient pour la première fois » ;
- de fournir une riche description des données via les métadonnées pour faciliter leur *recontextualisation* (Leonelli, 2019), c'est-à-dire le jugement des utilisateurs pour formuler leur propre interprétation de la signification des données et de leur crédibilité ».

Par ailleurs, comme souligne Szostak (2014), l'histoire de la science est fondée sur des exemples où les meilleures découvertes ont nécessité d'abord de combiner des éléments séparés de connaissance, ce qui obligeait parfois à sortir au-delà des champs disciplinaires. En transposant son idée dans le contexte de médiation des données en SHS, il s'agirait de guider les chercheurs à trouver les données qu'ils cherchent, mais aussi à découvrir des informations sur des thématiques connexes qu'ils n'auraient pas pensé à chercher. La délibération récente sur ce problème lors du Colloque NumeRev MSH-Sud met en relief un besoin accru de trouver des « connecteurs interdisciplinaires » capables de soutenir la construction d'une science « *plus ouverte et novatrice et de relier les données scientifiques entre elles* ». En face de la pluralité des communautés de pratiques en SHS qui se caractérisent, selon Wenger (1998), par des dimensions sociales (activités, pratiques) et

cognitives diverses (vocables, ressources, outils), pouvons-nous trouver ces connecteurs dans l'ILC ?

2 - EVALUATION DE L'ILC

Grâce à l'étude des pratiques, usages et contexte liés à l'échange et la médiation des données de la recherche en SHS, nous avons pu dégager trois défis majeurs. Dans ce volet, nous les confronterons avec les propriétés de l'ILC.

2.1 Décrire un vaste spectre des connaissances

Vu le nombre de champs disciplinaires qui font partie des SHS, les SOC devraient porter sur des connaissances vastes. L'ILC, comme la majorité des classifications bibliographiques, est un système général ayant pour vocation de classer tout le savoir humain et plus particulièrement, tous les phénomènes qui en font partie. Les classes de cette classification représentent justement des *phénomènes*, qui sont des unités basiques de la connaissance (Szostak, 2011). Ils sont représentés par les lettres minuscules (indices) :

a forms
b spacetime
c branes
d energy
e atoms
f molecules
g continuum bodies
h celestial bodies
i rocks
j land
k genes
l bacteria
m organisms
n populations
o instincts
p consciousness
q language
r rituals
s communities
t polities
u enterprises
v technologies
w artifacts
x artworks
y knowledge

Dans l'ILC, nous pouvons représenter des combinaisons de concepts touchant à tout champ scientifique grâce à la possibilité d'exprimer des relations entre n'importe quelle paire de classes. Cette fonctionnalité s'appuie sur un schéma hiérarchique des phénomènes, basé sur la *Théorie des niveaux d'intégrations* introduite par Needham (1937) et formalisée par Feibleman (1954). Selon cette théorie, lorsque les éléments d'un niveau inférieur (p. ex. : molécules, gènes) sont combinés, ils peuvent donner lieu à une nouvelle intégration (p. ex. : organismes), avec des propriétés et natures différentes de celles d'avant. Dans l'ILC, les *niveaux* aident à établir une séquence de phénomènes par des relations de dépendance basées sur des critères phylogénétiques. Ils sont complétés par des relations d'inclusion arrangeant les phénomènes selon l'augmentation de spécificité morphologique (Gnoli, 2017a). Cet exemple de zoologie nous semble bien explicite :

m organismes
mq animaux
mqv cordés
mqvt mammifères
mqvtu biongulés
mqvtuv moutons
mqvtuvsd moutons domestiques

En suivant cette logique, les concepts relatifs aux SHS, comme *langues slaves* et *proverbes* ont des cheminements des niveaux d'intégration comme suit :

q langue
qv langues
qve langues indo-européennes
qves langues slaves

x art
xl littérature
xlb proverbes

Grâce à ces propriétés, l'ILC peut à priori classer tout phénomène relevant de SHS.

2.2 Décrire les données hétérogènes d'une manière nuancée

Les principes FAIR recommandent de décrire les données par des métadonnées riches afin de faciliter leur recontextualisation. Les SOC en question devraient donc être capables de générer des descriptions très nuancées des phénomènes et des propriétés documentaires de données.

L'ILC donne la possibilité de combiner infiniment les phénomènes, relations et propriétés, grâce à ce que Szostak (2011) appelle une *approche synthétique*. Dans cette classification, une combinaison entre des phénomènes est spécifiée grâce à une syntaxe à facettes, permettant

d'établir des relations libres (Gnoli, 2017b). Voici les principales catégories de facettes de l'ILC :

0 under *perspective*
1 at *time*
2 in *place*
3 by *agent*
4 affected by *disorder*
5 with *transformation*
6 having *property*
7 with *part*
8 as *form*
9 of *kind*

et voici comment nous pouvons combiner les phénomènes en établissant une relation entre eux :

xlb « proverbes »	73 « en »	*qves* « langues slaves »
phénomène	**facette**	**phénomène**

ce qui nous donne : *xlb73qves* « proverbes, en langues slaves ».

Grâce à une combinaison libre de phénomènes via des facettes, nous pouvons décrire facilement les propriétés documentaires de données qui auraient pu être échangées en SHS (identifiées en partie dans la première partie de la réflexion), comme leur type, le format d'enregistrement, la langue, le type de licence, ou les méthodes d'étude typiques pour cet ensemble des disciplines :

r « traditions »	*03* « étudié par »	*yamk* « enquête »
phénomène	**facette dimensionnelle**	**perspective** [*cfr*. Gnoli, 2016]

ce qui nous donne : *r03yamk* « traditions, étudié[es] par enquête »
ou : *r2ttbzw00ytm* « traditions, en Pologne, enregistré[es] en vidéo ».

Avec les facettes de l'ILC, il est également possible de filtrer une collection en choisissant un ou plusieurs aspects de documents (sujet, format, etc.). Cette fonctionnalité semble importante pour faire des présélections de données et pourrait trouver sa place dans des entrepôts en SHS.

Comme nous pouvons combiner infiniment les phénomènes et facettes, l'ILC permet de décrire les données d'une manière nuancée et riche. Mais, compte tenu de sa création récente, certains concepts, surtout ceux qui relèvent du monde digital, comme les formats

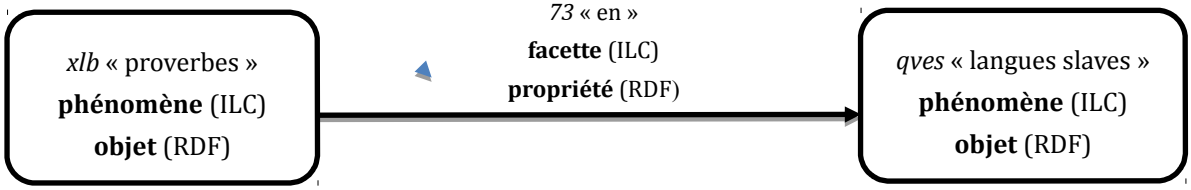
d'enregistrement (HTML, XML), les types de licence (Creative Commons) ou de nouvelles méthodes d'étude (machine learning, clustering) n'existent pas encore. Toutefois, ces concepts peuvent être facilement rajoutés.

2.3 Valoriser et faire découvrir l'information au-delà des frontières disciplinaires

Dans la partie précédente, nous avons noté le besoin de « connecteurs interdisciplinaires » pour décontextualiser les données, améliorer la RI et valoriser les données, vu les enjeux scientifico-économiques et les taux bas d'usages des données ouvertes en SHS. Grâce à la traduction des éléments de l'ILC vers le langage du Web sémantique, nous pouvons nous projeter vers la valorisation des données dans cet environnement.

Etant donné que les *phénomènes* sont probablement les termes pour lesquels il existe le plus grand accord sur le sens entre groupes, individus et disciplines (Szostak, 2014), ils peuvent devenir des connecteurs interdisciplinaires. Comme nous venons de montrer, l'*approche ontologique* de l'ILC s'appuie sur les phénomènes et permet de décrire les données d'une façon très nuancée et précise. Elle est souvent opposée à une *approche épistémologique* sur laquelle sont basées les classifications disciplinaires (p. ex. la Classification décimale de Dewey ou la Classification décimale universelle), qui enferment les concepts dans des classes disciplinaires et rendent difficile (voire impossible) leur combinaison (Trzmielewski *et* Gnoli, 2018).

L'approche synthétique s'inscrit dans la logique du Web sémantique. Ce dernier est une extension du Web permettant de connecter les données entres elles, grâce à une libre combinaison de concepts, de relations et de propriétés. Pour coder l'information dans cet environnement, les éléments basiques de l'ILC *phénomène – facette – phénomène* se traduisent en triplets RDF (Resource Description Framework) *objet – propriété – objet*. Pour illustrer cette traduction dans le contexte de médiation des données en SHS, revenons sur l'exemple de proverbes et de langues slaves :



Pour que les données sur les *proverbes en langues slaves* (décrites via l'ILC) puissent être liées sur le Web à d'autres données sur cette thématique, les concepts de l'ILC devraient être publiés en formats proposés par le WC3 (https://www.w3.org) et mis en correspondance avec les concepts des autres SOC.

L'ILC sera bientôt publiée en SKOS, format de représentation des SOC sur le Web sémantique. Cette mission confiée aux chercheurs de l'Hypermedia Research Unit de l'Université du Pays de Galles du Sud est en cours d'achèvement. Une fois que l'ILC sera disponible dans ce format, son interopérabilité avec d'autres SOC faisant partie du Web sémantique sera possible, notamment grâce à l'établissement de passerelles terminologiques dans des grilles de correspondance. Cela permettra par la suite de créer automatiquement des liens sémantiques entre les données décrites par l'ILC et celles décrites par d'autres systèmes. L'ILC est donc à priori capable d'améliorer la visibilité des données publiées par les chercheurs sur le Web et de rendre possible une recherche interdisciplinaire.

3 – RÉSULTATS ET DISCUSSION

A l'issue de la confrontation des fonctionnalités de l'ILC avec les pratiques, usages et contexte d'échange et de médiation des données ouvertes en SHS, nous pouvons tirer des premières conclusions.

Premièrement, une libre combinaison de *phénomènes* et *facettes* dans l'ILC permet à priori d'organiser les connaissances vastes sur lesquelles portent les SHS et de classer les propriétés des données hétérogènes qui sont échangées au sein de cet ensemble des disciplines. Les métadonnées peuvent donc être décrites d'une manière riche et nuancée.

Deuxièmement, la navigation par facettes de l'ILC pourrait faciliter le repérage des données au sein des infrastructures Web et permettre leur meilleure interprétation par les chercheurs. Cette fonctionnalité réduirait probablement le temps passé à la présélection des données.

Troisièmement, la possibilité de générer des connecteurs interdisciplinaires dans la syntaxe RDF incite à envisager l'ILC comme un système qui pourrait être utilisé pour valoriser les données sur le Web et améliorer leur recherche. Ceci permettrait de répondre aux taux bas d'usages des données ouvertes et de s'inscrire dans les pratiques de médiation.

Quatrièmement, il semble en revanche manquer certains concepts à l'ILC dû à la conception récente de la classification. Toutefois, elle est en train de croitre progressivement, car en 2017 elle comptait 9 000 concepts (Gnoli, 2017a) alors qu'en 2019 elle en représente déjà 10 000 (Park *et al*., 2019). Cette limite nous incite quand même à être prudents dans notre conclusion finale.

La classification est aujourd'hui appliquée au sein de trois bases de connaissances d'intérêt disciplinaire et multidisciplinaire : *BARTOC*, *BioAcoustic Reference Database* et *Where the Apennine Begins*. Malheureusement, l'on ne dispose pas encore de retour d'expérience de la part des utilisateurs et il est donc pour l'instant difficile de dire si l'ILC permet d'accompagner les pratiques informationnelles d'une manière efficace. Une évaluation qualitative auprès des utilisateurs nous semble indispensable et fait partie de nos projets futurs. Nous sommes aussi en train d'établir une méthodologie pertinente pour indexer via l'ILC un corpus important de données relatives aux SHS afin de mieux l'évaluer.

CONCLUSION

L'échange et la réutilisation des données scientifiques est bénéfique pour la science, l'économie et la carrière des chercheurs. Pourtant, les chercheurs en SHS les font moins souvent que leurs collègues des autres sciences. L'enjeu majeur est donc de fournir des SOC qui permettraient d'organiser les données très hétérogènes portant sur des thématiques vastes, afin d'augmenter le taux de leur consultation et de leur réutilisation. Par ailleurs, selon les politiques actuelles de médiation, les SOC en question sont amenés à valoriser les données sur des infrastructures Web et à faciliter leur interprétation par les chercheurs. C'est en vue de ces tâches que nous avons décidé d'évaluer l'ILC.

Les résultats de notre analyse nous amènent à constater que l'ILC pourrait être envisagée dans le futur pour échanger et médiatiser les données de la recherche en SHS. Elle pourrait par exemple intégrer des entrepôts qui actuellement ne sont dotés d'aucun SOC, ou bien qui proposent uniquement des systèmes disciplinaires pour organiser les données. Néanmoins, la classification devra d'abord être évaluée avec des méthodologies plus élaborées, et le rythme de son développement maintenu.

Dans l'article *The Importance of Knowledge Organization*, Szostak (2014) souligne des avantages découlant de l'intégration des classifications interdisciplinaires au sein des

systèmes d'information. Il en énumère trois : l'épargne du temps passé à la recherche d'information pertinente, l'orientation vers la découverte d'une information de qualité et l'extension des horizons documentaires. Ces avantages conduisent selon lui à améliorer la productivité des agents au travail et à augmenter le taux des découvertes scientifiques et des innovations technologiques. Au vu des bénéfices découlant de l'utilisation des classifications interdisciplinaires et des résultats de cette étude, nous proposons l'ILC comme artefact qui pourrait participer à la création d'une science plus ouverte, plus dynamique et novatrice. Nous invitons ainsi les chercheurs en SHS à prendre connaissance de cette classification et à l'enrichir via : http://www.iskoi.org/ilc.

RÉFÉRENCES BIBLIOGRAPHIQUES

BEGHTOL Clare (2004). Exploring new approaches to the organization of knowledge: James Duff Brown. *Library trends*, 2004, n° 52, p.702-718.

BORGMAN Christine L. (2008). Data, disciplines, and scholarly publishing. *Learned publishing*, 2008, n° 21, p.29-38.

BUCKLAND Michael (2014). Knowledge Organization and the Technology of Intellectual Work. In : BABIK Wiesław (dir.). *Knowledge Organization in the 21st Century: Between Historical Patterns and Future Prospects: Proc. Thirteenth International ISKO Conference, 19-22 May 2014, Kraków, Poland*. Würzburg : Ergon, 2014, p.14-21.

CLASSIFICATION RESEARCH GROUP (1969). *Classification and information control*. London : Library Association, 1969, 130p.

COMMISSION EUROPEENNE, DIRECTION GENERALE DE LA RECHERCHE ET DE L'INNOVATION (2016). *Lignes directrices pour la gestion des données FAIR dans Horizon 2020 Version 3.0 26 Juillet 2016*.

DÉLÉMONTEZ Renaud, BOUKACEM-ZEGHMOURI Chérifa (2015). Données de la recherche : entre discours, réalités et valeur. *I2D – Information, Données & Documents*, 2015, n° 53, p.56-57.

DILLAERTS Hans (2014). Le libre accès et le financement de projets de recherche transversaux : des vecteurs d'interdisciplinarité dans l'économie de la connaissance ?. *Revue européenne des sciences sociales*, 2014, no° 52, p.111-135.

EUROPEAN COMMISSION (2016). *Communication: European Cloud Initiative – Building a competitive data and knowledge economy in Europe (COM(2016) 178 final). Brussels, 19.4.2016.*

FEIBLEMAN James K. (1954). Theory of Integrative Levels. *The British Journal for the Philosophy of Science*, 1954, no° 5, p.59-66.

FERGUSON Liz (2014). How and Why Researchers Share Data (And Why They don't). *Wiley*, 3 November 2014. Disponible sur : https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-researchers-share-data-and-why-they-dont (01.09.2019)

FRANK Rebecca D., YAKEL Elizabeth, FANIEL Ixchel M. (2015). Destruction/reconstruction: preservation of archaeological and zoological research data. *Archival science*, 2015, n° 15, p.141-167.

FRICKÉ Martin (2019). The knowledge pyramid: the DIKW hierarchy. *Knowledge Organization*, 2019, n° 49, p.33-46.

GNOLI Claudio (2016). Classifying phenomena, part 1: Dimensions. *Knowledge Organization*, 2016, n° 43, p.403-415.

GNOLI Claudio (2017a). Classifying phenomena, part 2: Types and Levels. *Knowledge organization*, 2017, n° 44, p. 37-54.

GNOLI Claudio (2017b). Classifying phenomena, part 3: Facets. In : SMIRAGLIA Richard, LEE Hur-Li (dir.). *Dimensions of Knowledge*. Würzburg : Ergon Verlag, 2017, p.55-67.

GNOLI Claudio, LEDL Andreas, PARK Ziyoung, TRZMIELEWSKI Marcin (2018). Phenomenon-based vs. disciplinary classification: possibilities for evaluating and for mapping. *Challenges and opportunities for knowledge organization in the digital age: proceedings of the Fifteenth International ISKO Conference 2018*, n° 16. Würzburg : Ergon, 2018, p.653-662.

GÓMEZ Nancy-Diana, MÉNDEZ Eva, HERNÁNDEZ-PÉREZ Tony (2016). Social Sciences and Humanities Research Data and Metadata: A Perspective from Thematic Data Repositories. *El profesional de la información*, 2016, n° 25, p.545-555.

HJØRLAND Birger (2018). Data (with big data and database semantics). *Knowledge Organization*, 2018, n° 45, p.685-708.

KAASE Max (2001). Databases, Core: Political Science and Political Behavior. In : SMELSER Neil J., BALTES Paul B. (dir.). *International Encyclopedia of the Social and Behavioral Sciences* n° 5. Amsterdam : Elsevier, 2001, p.3251-3255.

The León Manifesto (2007). *Knowledge Organization*, 2007, n° 34, p.6-8.

LEONELLI Sabina (2019). *La recherche scientifique à l'ère des Big Data. Cinq façons dont les Big Data nuisent à la science et comment la sauver*. Sesto S. Giovanni : Éditions MIMESIS, 2019, 122p.

LYON Liz (2016). Transparency: the emerging third dimension of open science and open data. *Liber quarterly*, 2016, n° 25, p.153-171.

NEEDHAM Joseph (1937). *Integrative Levels: A Revaluation of the Idea of Progress*. Oxford : Clarendon Press, 1937, 59p.

NICOLESCU Basarab (1996). *La transdisciplinarité – Manifeste*. Monaco : Éditions du Rocher, 1996, 98p.

OCDE (2007). *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*. Paris : Editions de l'OCDE, 2007.

ODEH Souad (2017). Les données de la recherche : transformation ou transmission du métier de documentaliste. *I2D – Information, données & documents*, 2017, n° 54, p.4-7.

PARK Ziyoung, GNOLI Claudio, MORELLI Daniele P. (2019). The second edition of the Integrative Levels Classification. *NKOS Workshop at DCMI 2019, Seoul, South Korea, 25 September 2019.*

SCHÖCH Christof (2013). Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities*, 2013, n° 3, p.2-13.

SZOSTAK Rick (2011). Complex concepts into basic concepts. *Journal of the American Society for Information Society and Technology*, 2011, n° 62, p.2247-2265.

SZOSTAK Rick (2014). The importance of Knowledge Organization. *Bulletin of the Association for Information Science and Technology*, 2014, n° 40, p.37-42.

SZOSTAK Rick, GNOLI Claudio, LÓPEZ-HUERTAS Maria (2016). *Interdisciplinary Knowledge Organization*. Cham : Springer, 2016.

TRZMIELEWSKI Marcin, GNOLI Claudio (2018). Un défi de recherche interdisciplinaire : comparaison d'une classification disciplinaire avec une classification basée sur les phénomènes. *Digital Libraries & Information Sciences*. Disponible sur : https://dlis.hypotheses.org/1895 (01.09.2019)

VAN DEN EYNDEN Veerle (2016). *Towards Open Research: practices, experiences, barriers and opportunities*. London : Wellcome Trust, 2016.

VAN DEN EYNDEN Veerle (2018). What Motivates Researchers to Manage and Share Research Data. In : THESTRUP Jesper Boserup, KRUSE Filip (dir.). *Research Data Management - A European Perspective*. Berlin, Boston : De Gruyter, 2018, p.43–52.

WENGER Etienne (1998). *Communities of practice: Learning, meaning and identity*. New York : Cambridge University, 1998, 318p.