

Métadonnées pour la découverte :

Rapport d'étape sur les normes disciplinaires et les tableaux de concordance

Préparé par le Groupe de travail sur les métadonnées du Groupe d'experts sur la découverte des données du réseau Portage au nom de l'Association des bibliothèques de recherche du Canada (ABRC)

Amber Leahey (Scholars Portal, présidente)
Eugene Barsky (University of British Columbia)
John Brosz (University of Calgary)
Alex Garnett (Simon Fraser University)
Vincent Gray (Western University)
Joseph Hafner (McGill University)
Kara Handren (Scholars Portal)
Amanda Harrigan (University of Alberta)
Christian Lacroix (Université Laval)
Julienne Pascoe (Canadiana.org)
Catelynne Sahadath (Université d'Ottawa)
Dany Savard (York University)
Andrew Senior (McGill University)
Barbara Towell (University of British Columbia)
Lee Wilson (Portage/ACENET)

SEPTEMBRE 2017

Réseau Portage
Association des bibliothèques de recherche du Canada
portage@carl-abrc.ca

www.portagenetwork.ca

portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES • ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

Table des matières

Contexte.....	3
Rapport d'étape — phase 1 : 2017.....	4
Liste des normes.....	4
Profil de découverte du DFDR.....	7
Évaluation des normes et mappage/tableau de concordance.....	10
Recommandations pour améliorer la découverte dans le DFDR.....	15
Prochaines étapes — phase 2 : 2017-2018.....	17
Annexe 1.....	18

Contexte

L'objectif du Groupe d'experts sur la découverte des données du réseau Portage est de soutenir les créateurs de données et les responsables de l'intendance des données lors de la planification, de la production et de la gestion des métadonnées descriptives pour permettre une découverte et une réutilisation efficaces des données de recherche dans un large éventail de disciplines. Le groupe facilite les discussions concernant la découverte des données au Canada, en vue de promouvoir l'utilisation de normes des métadonnées pour les données de recherche qui soutiennent à la fois les activités de découverte personne-machine et inter-machines.

Le Groupe de travail sur les métadonnées du Groupe d'experts sur la découverte des données (GTM)¹ s'emploie à définir la portée des normes des métadonnées et les éléments à prendre en considération dans la création du [Dépôt fédéré de données de recherche](#) (DFDR) du Canada², un service national de découverte et de dépôt de données (à l'étape du développement depuis juillet 2017). Le principal objectif du GTM consiste à cerner les normes des métadonnées utilisées dans les dépôts de données canadiens et à élaborer des tableaux de concordance détaillés pour permettre la découverte des données de recherche. Le travail comprend la collecte d'une liste de normes de métadonnées descriptives, l'évaluation détaillée des éléments de métadonnées utilisés dans un échantillon significatif des disciplines et des dépôts, et la formulation de recommandations pour un ensemble fondamental d'éléments pour la découverte dans le DFDR. La portée du travail de ce groupe ne se limite pas au seul projet du DFDR, mais profitera aussi de manière générale à toute personne qui participe à la gestion des métadonnées pour les systèmes de découvertes et les dépôts de données de recherche. Certaines discussions étaient aussi centrées sur le rayonnement auprès d'autres initiatives et organisations nationales comme Données de recherche Canada³, Linked Open Data in Libraries, Archives, and Museums⁴, et le projet

¹ Groupe d'experts sur la découverte des données du réseau Portage — Groupe de travail sur les métadonnées — <https://portagenetwork.ca/fr/collaborer-avec-portage/reseau-dexperts/decouverte/appel/gt-metadonnees>

² On trouvera un complément d'information sur ce projet d'infrastructure nationale pour appuyer la découverte de la recherche en ligne dans le réseau Portage de l'ABRC (<https://portagenetwork.ca/frdr-dfdr/francais>)

³ Données de recherche Canada <https://www.rdc-drc.ca/fr>

⁴ LODLAM <http://lodlam.net>

SHARE de l'Association of Research Libraries (ARL)⁵, pour acquérir une compréhension plus approfondie et pour mieux partager les connaissances concernant les systèmes et les outils de découverte de données de recherche en libre accès. Dans l'avenir, des liens plus étroits avec les communautés nationales et les initiatives de données liées peuvent être nécessaires.

Rapport d'étape — phase 1 : 2017

Liste des normes

Le GTM a été créé en février 2017 et a commencé à évaluer les normes des métadonnées utilisées par divers dépôts de données de recherche au Canada, notamment Données ouvertes Canada, Ocean Networks Canada, et les instances canadiennes de Dataverse⁶. Puisque ce type de travail n'est pas unique, le groupe a évalué les mappages de normes de métadonnées disciplinaires, les correspondances, les modèles de données, les ontologies, *et cetera* qui existent pour les données de recherche. Les résultats de l'évaluation ont été partagés au moyen de nos téléconférences collectives et par une liste de diffusion, et ils ont fait l'objet de discussions. Puisque le travail du DFDR avait déjà commencé plusieurs mois avant la création du GTM, il a été déterminé qu'une recommandation de modèle de données dépassait le mandat du groupe (au moins aux stades initiaux de notre travail). Le groupe s'est concentré sur la collecte de normes, notamment l'examen des dépôts figurant sur la liste du registre de la Passerelle vers les données de recherche de l'Institut canadien de l'information scientifique et technique (ICIST)⁷ et du registre des dépôts mondiaux Re3data.org⁸, pour mieux cerner les normes des métadonnées utilisées dans l'ensemble des dépôts des diverses disciplines. En fin de compte, le fait de faire appel à l'expertise riche et variée au sein du groupe (allant de l'expertise en métadonnées dans les secteurs des sciences sociales, des sciences de la Terre, des sciences de la santé, du patrimoine culturel et des bibliothèques, etc.) a permis de produire une liste de plus de vingt normes de métadonnées disciplinaires pour une évaluation plus approfondie (voir la figure 1).

⁵ Projet SHARE de l'ARL <http://www.share-research.org>

⁶ Le projet Dataverse, carte des installations : <https://dataverse.org>

⁷ <https://dr-dn.cisti-icist.nrc->

[cnrc.gc.ca/fra/accueil/collection/Passerelle%20de%20donn%C3%A9es%20de%20recherche](http://www.re3data.org)

⁸ <http://www.re3data.org>

Figure 1 - Liste des normes de métadonnées générales et disciplinaires (considérées)

Norme	Général/Discipline	Organisation
Audio-MD		Library of Congress https://www.loc.gov/standards/amdvmd
CIDOC		ICOM International Committee for Documentation http://network.icom.museum/cidoc
Darwin Core	Information sur la biodiversité	Biodiversity Information Standards (TDWG) http://rs.tdwg.org/dwc/terms/simple/#simpledwcaxml
Data Documentation Initiative (Codebook and Lifecycle versions)	Général / Sciences sociales	DDI Alliance https://www.ddialliance.org
Data Tag Suite (DATS)	Général / Sciences de la santé	BioCADDIE https://biocaddie.org
DataCite	Général, attribution des DOIs	DataCite https://www.datacite.org
EAD3	« Norme internationale de transmission des métadonnées pour les descriptions hiérarchiques des enregistrements d'archives ». [traduction]	Society of American Archivists https://www2.archivists.org/sites/all/files/TagLibrary-VersionEAD3.pdf
Ecological Markup Language (EML)	Écologie	The Knowledge Network for Biocomplexity https://knb.ecoinformatics.org/#external/emlparser/docs/index.html
FGDC	Géographique	Federal Geographic Data Committee https://www.fgdc.gov/metadata/csdgm-standard
ISAAR/ EAC-CPF	« Règles générales pour la normalisation des descriptions d'archives des créateurs d'enregistrements et le contexte de la création des enregistrements ». [traduction]	International Council on Archives http://www.ica.org/sites/default/files/CBPS_Guidelines_ISAAR_Second-edition_EN.pdf
ISAD(G)	Conseils généraux pour la préparation des descriptions d'archives.	International Council on Archives http://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition
ISO 19115	Géographique	Ressources naturelles Canada http://nap.geogratis.gc.ca/metadata/napMetadata-eng.html
MARC 21	Description des enregistrements bibliographiques.	Library of Congress https://www.loc.gov/marc/marcdocz.html

MODS	Ensemble d'éléments bibliographiques.	Library of Congress http://www.loc.gov/standards/mods
NetCDF CF Metadata Conventions 1.6	Climat et prévisions / Géospatial	Climate and Forecast Conventions http://cfconventions.org
Open Data Canada Profile (DCAT)	Général	W3C https://www.w3.org/TR/vocab-dcat
Protocol Data Element Definitions	Sciences de la santé. Utilisé pour décrire les études d'intervention (essais cliniques) et les études d'observation	U.S. National Institutes of Health https://clinicaltrials.gov
RAD	Général. Description des archives.	Canadian Committee on Archival Description http://www.cdncouncilarchives.ca/archdesrules.html
Sensor Model Language (SensorML)	Multidisciplinaire / Géoscience	Open Geospatial Consortium http://www.opengeospatial.org/standards/sensorml
Video-MD		Library of Congress https://www.loc.gov/standards/amdvmd
VRA-CORE	Général, utilisé pour la « description des œuvres de la culture visuelle ainsi que des images qui les documentent ». [traduction]	Library of Congress https://www.loc.gov/standards/vracore

À partir de la définition des données de recherche⁹ fournies par le Groupe de travail sur le développement des collections du Groupe d'experts sur la découverte des données (GTDC)¹⁰, un sous-groupe connexe du GEDD du réseau Portage, le GTM a décidé de se concentrer uniquement sur les normes qui décrivent les données de recherche numériques. Un travail de configuration a été effectué pour les normes des métadonnées supplémentaires susmentionnées, toutefois le rapport est centré sur les normes de données de recherche. Les membres du groupe de travail ont adhéré aux normes et ont fourni des notes détaillées concernant les champs et une analyse dans un tableau commun. Puisque les normes portaient sur une gamme de disciplines et de types de données, comme les sciences sociales, les sciences de la santé et de la vie, et les sciences de la Terre, il était important que le groupe prenne le temps d'examiner

⁹ Définition : [traduction] *Données employées comme sources principales de soutien aux enquêtes techniques ou scientifiques, à la recherche, aux études et aux activités artistiques; elles servent à étayer le processus de recherche et sont couramment considérées comme nécessaires pour valider les constatations et les résultats des recherches.* (fondée sur la définition du CASRAI http://dictionary.casrai.org/Research_data)

¹⁰ <https://portagenetwork.ca/wp-content/uploads/2017/04/GEDD-CollectionsGT-Mandat-FR.pdf>

les normes disciplinaires et de s'informer à leur sujet avant que commence le mappage des métadonnées ou l'établissement de tableaux de concordance pour ces données. Dans de nombreuses situations, les membres avaient une connaissance approfondie d'une norme précise, ou de l'expérience connexe, ce qui donnait lieu à de très intéressantes discussions et à l'échange de connaissances entre les domaines.

Dans certains cas, il était simple de recueillir des métadonnées complètes concernant les champs pour les diverses normes, tandis que dans d'autres cas, il fallait faire des recherches supplémentaires, y compris l'extraction des métadonnées de l'échantillon provenant de dépôts connus, de sites Web d'organismes de normalisation et même de documents conceptuels mentionnant les éléments de métadonnées. En fin de compte, nous nous sommes retrouvés avec treize ensembles complets de normes de métadonnées descriptives pour l'évaluation et les tableaux de concordance. On a choisi un simple ensemble de Dublin Core (modèle) pour le modèle de données du DFDR¹¹ avant la création du groupe de travail initial; il était donc la norme commune utilisée pour le mappage.

Profil de découverte du DFDR

Le profil de découverte de référence du DFDR est fondé sur le Dublin Core non qualifié et est défini au moyen de la norme Dublin Core de l'Open Archives Initiative (initiative pour des archives ouvertes) (OAI_DC) qui est largement utilisée pour recueillir des métadonnées dans les dépôts numériques. Certains éléments du schéma DataCite sont aussi utilisés pour définir des renseignements géographiques liés aux données. Ces éléments réunis constituent le profil de métadonnées du DFDR.

¹¹ Bien que ce choix a été fait sachant que des métadonnées de domaine qui peuvent être recueillies nativement (ou fournies par des dépositaires) seraient toujours préservées et indexées en arrière-plan du DFDR avec leur espace de nommage original, afin que la découverte ne se limite pas au Dublin Core — il est plutôt utilisé à titre de référence minimale.

Le profil complet contient 18 éléments consultables dans le DFDR :

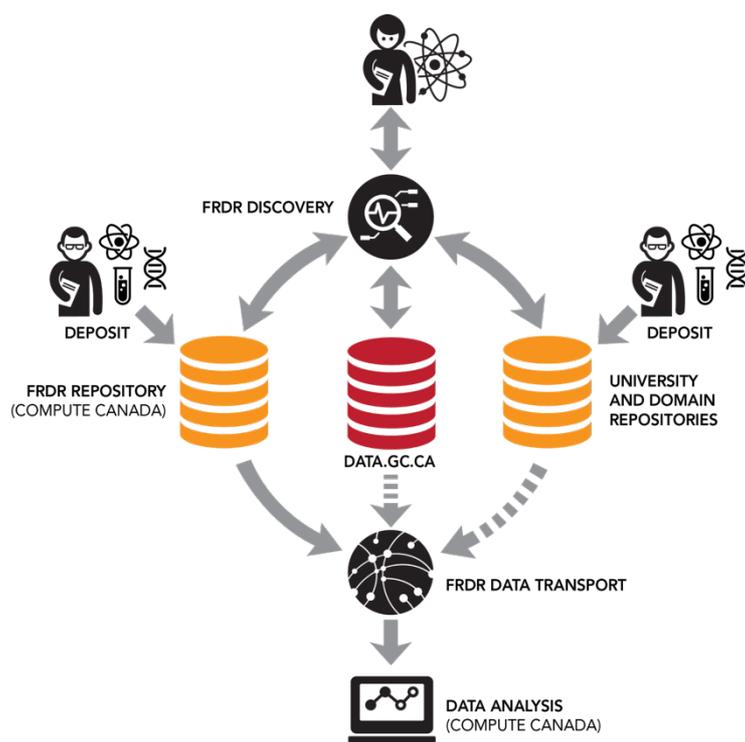
dc:title (titre)	dc:identifiant (identifiant de la ressource)
dc:creator (créateur)	dc:source
dc:subject (sujet)	dc:language (langue)
dc:description	dc:relation
dc:publisher (éditeur)	dc:coverage (couverture)
dc:contributor (contributeur)	dc:rights (droits)
dc:date	DataCite_geolocationPlace
dc:type	DataCite_geolocationPoint
dc:format	DataCite_geolocationBox

Le DFDR comprend un système de collecte de métadonnées (voir la figure 2) qui fonctionne en arrière-plan du système pour recueillir des métadonnées dans les dépôts de données de recherche, lesquels utilisent diverses normes de métadonnées disciplinaires. Cette étape est principalement effectuée au moyen d'une interface de programmation d'application (API) de métadonnées comme CKAN et OAI-PMH (Open Archives Institute Protocol pour la collecte de métadonnées)¹². Les outils de moissonnage personnalisés dépendent du tableau de concordance conceptuel des métadonnées du profilé de découverte du DFDR et du mappage schématique JSON¹³ pour afficher des métadonnées uniformes pour la recherche effectuée dans la plateforme de découverte du DFDR.

¹² <https://www.openarchives.org/pmh>

¹³ Mise en correspondance du schéma OAI JSON dans GitHub https://github.com/axfelix/globus_oai

Figure 2 - La récolte de données par le DFDR¹⁴



Dans les discussions initiales portant sur la mise en correspondance du profil de découverte du DFDR, le groupe a exprimé plusieurs préoccupations liées à la façon dont chaque membre ferait la mise en correspondance des normes disciplinaires avec un simple ensemble d'éléments du Dublin Core. Des préoccupations concernant l'uniformité du processus de mise en correspondance intellectuelle pour l'ensemble des normes ont été soulevées, puisque les membres individuels du groupe effectuaient des concordances différentes. La deuxième préoccupation portait sur l'interprétation des champs Dublin Core et sur la façon dont nous allons les utiliser au chapitre de l'interface de découverte dans le DFDR. Nous avons notamment discuté des liens de consultation vers le fournisseur de données d'origine/dépôt, des définitions et des rôles des contributeurs, et des rapports et des couplages entre les données de recherche et d'autres ressources connexes. La figure 3 fournissent des renseignements sur la façon dont certains éléments Dublin Core ont été interprétés et utilisés par les membres du groupe de travail.

¹⁴ La figure est reproduite ici avec la permission de Calcul Canada et est disponible seulement en anglais.

Figure 3 – Interprétation et application des éléments Dublin Core pour l'exercice du mappage

Éléments Dublin Core	Interprétation et application
dc:contributor (contributeur)	Un contributeur désigne une personne ou un organisme qui a contribué, soit directement au contenu intellectuel, ou, à assurer la diffusion de la ressource ou l'accès à la ressource (p. ex., rôle de distributeur, ou autre, selon la définition dans certaines normes). Répétable.
dc:coverage (couverture)	La couverture permet de définir la couverture temporelle ou la période couverte ou visée par l'ensemble de données, et peut inclure d'autres types de couvertures qui ne sont pas définis ici, à l'exception de la couverture spatiale qui est gérée au moyen d'un ensemble d'éléments de géolocalisation DataCite, conjointement au profil de découverte du DFDR.
dc:source	<p>La source désigne les métadonnées sources d'origine qui ont été présentées pour la découverte et la réutilisation dans le Dépôt fédéré de données de recherche. Il s'agit d'un champ obligatoire qui peut être dérivé et aussi autoalimenté dans le cadre du processus de collecte ou d'échange de métadonnées. Les conditions de ce champ sont les suivantes : il doit s'agir d'une adresse URL, c'est-à-dire un emplacement repérable unique de ressources, qui permet au système d'aiguiller les utilisateurs pour qu'ils accèdent aux métadonnées source d'origine.</p> <p>Remarque supplémentaire : Dans certains cas, ce champ pourrait être le même que le champ dc:identifier; toutefois, dc:source doit être repérable sur le Web.</p>
dc:relation	Les relations comprennent les références, les citations, sous forme structurée ou non, relatives à du matériel, à des publications, à des documents, à des études ou autres, qui peuvent être liées à l'élément en question.

Évaluation des normes et mappage / tableau de concordance

Pour commencer, le groupe a établi des critères pour codifier les éléments en vue d'aider au processus de mappage. Un code de couleur a servi pour l'évaluation des éléments et a été entré directement dans le tableau de métadonnées. Il comprenait la détermination d'éléments qu'on estimait étroitement liés au Dublin Core, d'identificateurs uniques ou de liens vers des ressources apparentées, en plus d'éléments considérés comme étant propres à une discipline.

Figure 4 – Mappage des métadonnées DATS à Dublin Core (la sélection exemplaire est présentée en anglais)

FRDR-MD (OAI_Dublin Core) http://www.openarchives.org/OAI/2.0/oai_dc.xsd	Refinement	DATS	
dc:title		entity:Dataset property:title	"title" : "Recurrent somatic mutations in POLR2A define a distinct subset of meningiomas [RNA-seq]"
dc:title	Alternative	entity:Dataset property:alternateIdentifiers	
dc:creator		entity:Dataset property:creators	
dc:subject		entity:Dataset property:keywords	"keywords" : "functional genomics"
dc:description	Abstract	entity:Dataset property:description	"description" : "RNA polymerase II mediates the transcription of all protein-coding genes in eukaryotic cells, a process that is fundamental to life..."
dc:publisher		entity:Person property:fullName or entity:Organization property:name	
dc:contributor	Distributor	entity: DataRepository property:name	
dc:contributor	Contact	entity:Person property:fullName	
dc:date		entity:Dataset property:dates (The type of date is specified in the dateType field, following the DataCite practice.)	"dateModified" : "09-12-2016"
dc:type		entity:Dataset property:type	"types" : "gene expression"

Mappage proche à Dublin Core	Potentiel de liaison	Discipline spécifique
------------------------------	----------------------	-----------------------

L'évaluation des treize normes des données a nécessité des discussions parmi les membres du groupe afin de comprendre les différences inhérentes entre les

normes, comme la granularité et le langage propre à une discipline utilisé pour définir les éléments. Un problème défini pour le groupe consistait à soupeser les avantages d'inclure des parties des diverses normes qui ne se prêtaient pas bien à la mise en correspondance avec le Dublin Core, avec une certaine souplesse. À titre d'exemple, si un dépôt utilise la norme de métadonnées DDI (qui fournit des champs descriptifs jusqu'à la variable d'ensemble de données et le niveau de valeur), il serait dommage de perdre cette information dans l'interface de découverte. Parallèlement, il est difficile d'établir un modèle de norme pour la découverte qui s'adapte à toutes les disciplines et de présumer que ces métadonnées seront toujours incluses. Des variations dans les métadonnées reçues produiraient des résultats de recherche incomplets et potentiellement trompeurs.

Ensuite, après l'achèvement du mappage initial pour chacune des normes, le groupe a combiné ces mappages dans un super-mappage qui a produit un tableau de concordance de haut niveau de toutes les normes de métadonnées disciplinaires et générales, pour le profil de découverte du DFDR (voir la figure 5; consultez l'Annexe 1 pour une version plus grande). Ce tableau de concordance de haut niveau fournit un aperçu comparatif du mappage conceptuel de l'ensemble des normes disciplinaires. Il est utile pour déterminer les éléments de base pour l'ensemble des normes, et la façon dont nous pouvons combler les lacunes ou envisager l'élaboration d'une approche souple pour afficher des métadonnées multidisciplinaires dans une seule interface de découverte.

Figure 6 – Liste courte des dépôts avec sujet et norme

Dépôt de donnée de recherche	Sujet de haut niveau (utilisant re3data.org)	Norme(s)	Inclut dans le DFDR beta Oui/Non
Canadian Opinion Research Archive (http://www.queensu.ca/cora/)	Humanités et Sciences sociales	OAI/DDI-Codebook	Non
Ocean Networks Canada (http://www.oceannetworks.ca/)	Sciences de la vie - Sciences de la nature - Génie	N/A - en cours d'investigation	Non
Polar Data Catalogue (https://www.polardata.ca/)	Humanités et Sciences sociales - Sciences de la vie	OAI/FGDC	Oui
Canadian Dataverses (http://dataverse.scholarsportal.info , https://dataverse.library.ualberta.ca/ , http://dvn.library.ubc.ca/dvn/)	Humanités et Sciences sociales - Sciences de la vie - Sciences de la nature - Génie	OAI/DC/DDI	Oui
Mouse Atlas of Gene Expression (http://www.mouseatlas.org/mouseatlas_index.html)	Sciences de la vie	N/A	Non
World Ozone and Ultraviolet Radiation Data Centre (http://woudc.org/)	Sciences de la nature	OAI/ISO 19115	Non
Ocean Tracking Network (OTN) (http://oceantrackingnetwork.org/)	Sciences de la vie - Sciences de la nature	CSW	Non
Hakai Institute (https://www.hakai.org/)	Sciences de la nature	N/A	Non
BC Conservation Data Centre (http://www2.gov.bc.ca/gov/content/environment/plants-animals-ecosystems/conservation-data-centre)	Sciences de la vie	OAI/CKAN API	Oui - à travers de « Open Data B.C. »

Le mappage des nouveaux dépôts se poursuivra jusqu'à l'automne de 2017, puisque nous devons peut-être déterminer si le dépôt utilise un schéma sur mesure. Dans l'affirmative, un mappage supplémentaire avec le profil du DFDR sera nécessaire.

Recommandations pour améliorer la découverte dans le DFDR

Pour appuyer la granularité et les éléments propres aux disciplines nécessaires pour la richesse de la découverte dans l'interface de découverte du DFDR, le GTM recommande d'élaborer un soutien supplémentaire dans le système du DFDR.

Recommandations pour améliorer la découverte des données dans le DFDR :

- Soutien pour la recherche à facettes de sujets ou la navigation des résultats de recherche à facettes
- Soutien de métadonnées disciplinaires souple et granulaire (p. ex., au niveau des variables d'ensemble de données)
- Amélioration des liens vers des ressources connexes et des ensembles de données sources

Il peut y avoir plusieurs façons de concrétiser ces recommandations pour améliorer la découverte dans le DFDR, dont certaines pourraient aussi nécessiter que le GTM et d'autres groupes effectuent des enquêtes et des recherches supplémentaires, en étroite collaboration avec l'équipe de développement du DFDR. Jusqu'à maintenant, le GTM a évalué des démarches qui pourraient fournir des solutions pour améliorer la découverte dans le DFDR, notamment l'utilisation d'une classification normalisée par sujets comme les sujets de re3data.org accessibles au moyen d'une API de métadonnées ouverte pour attribuer des sujets à toutes les métadonnées de dépôt (pour ceux qui figurent dans re3data.org)¹⁶, ou l'utilisation de services de données comme le service FAST d'OCLC¹⁷ pour l'accès programmatique aux vedettes-matières de la Library of Congress. Néanmoins, il y a d'importantes difficultés associées à la tenue à jour de métadonnées perfectionnées supplémentaires pour les ressources externes de métadonnées. La pérennité des métadonnées perfectionnées serait peu probable, puisque chaque dépôt peut utiliser des normes différentes pour attribuer des sujets aux métadonnées et aux sources de données, ce qui nécessite un entretien supplémentaire et exigeant en fait de ressources des mappages de sujet avec les sujets de base du DFDR (s'il est adopté).

¹⁶ API Re3data.org <http://www.re3data.org/api/doc>

¹⁷ Recherche FAST - OCLC <http://fast.oclc.org/searchfast>

Le groupe a discuté de la nécessité pour le DFDR d'appuyer les « affichages disciplinaires » pour les diverses normes en vue de faire preuve de souplesse et d'offrir des éléments de métadonnées à l'interface de recherche, au besoin. À titre d'exemple, certains dépôts renferment des métadonnées riches jusqu'à la variable d'ensemble de données ou le niveau d'élément, ce qui peut être très utile pour les chercheurs qui recherchent des éléments de données précis. Le stockage de l'intégralité des métadonnées sources d'origine et l'élaboration de visionneuses de métadonnées souples accessibles à partir de l'interface de découverte du DFDR seront nécessaires pour appuyer ce genre de découverte dans le système.

Le couplage des données avec les publications et les ressources connexes sur le Web soulève de plus en plus de préoccupations au sein des bibliothèques, des archives et du milieu des chercheurs. Il est important de présenter les données de recherche dans le contexte d'autres résultats de recherche analogues accessibles dans les dépôts, et avec des citations et des liens vers des ressources sur le Web. La découverte de données de recherche dans divers systèmes sera nécessaire pour comprendre l'ensemble du cycle de vie de la recherche et pour faire le suivi des extraits de recherche pour l'évaluation de l'ancienneté et des accomplissements universitaires. Pour appuyer un modèle où les données de recherche sont liées à des extraits de recherche associés, y compris la publication, il doit y avoir des liens structurés entre les ressources sur le Web. Cela comprend des approches structurées pour inclure des liens vers des publications, des études apparentées, et des données sources, au sein des modèles de données et des cadres de découverte adoptés. À l'heure actuelle, le DFDR ne supporte pas un modèle de données liées comme le schéma du modèle RDF (Resource Description Framework) Dublin Core¹⁸, lequel décrit les relations sémantiques entre les éléments définis dans le Dublin Core, et avec d'autres ressources connexes et ontologies qu'on trouve sur le Web.

Le groupe a aussi discuté de l'utilisation d'autres modèles de données liées bien documentés comme le DCATS¹⁹ et le Portland Common Data Model²⁰. Même s'il faut peut-être des ressources considérables pour réaliser un modèle détaillé de données liées pour les données de recherche dans le DFDR, nous recommandons

¹⁸ DC-RDF <http://dublincore.org/documents/dc-rdf>

¹⁹ DCATS <https://www.w3.org/TR/vocab-dcat>

²⁰ PCDM <https://github.com/duraspace/pcdm/wiki>

d'envisager des partenariats utiles avec d'autres organismes qui visent l'atteinte d'objectifs similaires. Le projet SHARE ARL-OSF est ce genre d'organisme qui a montré qu'il souhaitait travailler en étroite collaboration avec le réseau Portage et, à compter de 2017, il examinera les données liées dans le cadre de son examen actuel de modèle de données.

Prochaines étapes — phase 2 : 2017-2018

Au cours des prochains mois et jusqu'au début de 2018, le GTM travaillera en étroite collaboration avec l'équipe de développement du DFDR pour assurer le soutien du mappage des métadonnées et de l'établissement des tableaux de concordance pour le reste des dépôts afin de les intégrer. Cela comprendra la détermination des normes des métadonnées utilisées là où il n'y a aucune norme définie à l'heure actuelle. Nous travaillerons aussi à développer les meilleures approches pour offrir la recherche par facettes des sujets et la navigation dans le DFDR, sans oublier la découverte granulaire et souple des métadonnées disciplinaires dans le système. Ultiment, il faudra envisager des approches du domaine des données liées, à mesure que ces dernières se développent dans la vaste communauté des producteurs et gestionnaires de métadonnées.

