



Portage Network

Dataverse North Working Group: Year 1 Recommendations

**Prepared by:
Portage Dataverse North Working Group (DVNWG)**

March 2018

Dataverse North Working Group

Overview

Portage established the Dataverse North Working group (DVNWG) in March 2017 with a mandate to “develop a community of practice for libraries using or interested in using the Dataverse repository platform for research data in Canada.” The group currently has 26 active members representing 19 universities from all regions, as well as OCUL, COPPUL, and Scholar’s Portal.

At the May 2017 meeting, members collectively identified three major issues that we felt should be tackled in our first year. We divided ourselves into three groups, each of which was tasked to deliver a report and set of recommendations by March 2018:

- Dataverse North Business Models Group, which considered the question of sustainable, equitable access to Dataverse in Canada;
- Dataverse North Metadata Group, who were asked to recommend a common, standard metadata template for Canadian Dataverse repositories; and
- Dataverse North Training Group, who were to assess training needs and to recommend strategies for Dataverse training and education.

In all cases these groups were directed to work closely with related Portage expert groups, and fortunately quite a few DVNWG members are cross-appointed to a related Portage expert group. In the main body of this report you will find the details of each group’s work, and a full set of recommendations related to each topic.

Recommendations in Brief

Portage, through its Dataverse North Working Group, should work with key stakeholders to establish a national Dataverse North service, based on a sustainable business model and hosted by Scholars Portal at the University of Toronto Libraries. (Business Models)

CARL institutions should adopt a common base metadata template for Dataverse repositories using standard vocabularies and identifiers. (Metadata)

CARL institutions should adopt a common approach to labelling, describing, and organizing data files in Dataverse repositories. (Metadata)

Dataverse North should work closely with TEG to develop multi-modal training materials with an immediate priority on general introductory manuals, videos, and workshops that are aimed at librarians and library staff. (Training)

Full details and rationale for these recommendations are included in the reports that follow.

Groupe de travail Dataverse Nord

Aperçu

Portage a constitué le Groupe de travail Dataverse Nord (GTDN) en mars 2017 et lui a confié le mandat de « développer une communauté de pratique pour les bibliothèques qui utilisent ou qui souhaitent à utiliser la plateforme Dataverse pour le dépôt des données de recherche canadiennes ». Le groupe compte actuellement 26 membres actifs représentant 19 universités de toutes les régions, ainsi que le CBUO, le COPPUL et Scholars Portal.

Lors de la réunion de mai 2017, les membres ont conjointement identifié trois enjeux stratégiques à aborder dans la première année. Les membres se sont divisés en trois groupes, chacun ayant la responsabilité de produire un rapport et des recommandations avant mars 2018 :

- Le Groupe des modèles d'affaires de Dataverse Nord, qui a pris en charge la question de l'accès pérenne et équitable à Dataverse au Canada;
- Le Groupe des métadonnées de Dataverse Nord, à qui on a demandé de recommander un modèle de métadonnées commun et normalisé pour les dépôts Dataverse au Canada; et
- Le Groupe de formation de Dataverse Nord, qui a évalué les besoins en formation et recommandé des stratégies à Dataverse en matière de formation.

Les trois groupes ont reçu le mandat de travailler de près avec les experts des groupes Portage respectifs et heureusement de nombreux membres du GTDN ont été ajoutés aux groupes d'experts de Portage. Dans le rapport, vous trouverez les détails du travail des trois groupes ainsi qu'une liste complète de recommandations pour chaque enjeu.

Recommandations en bref

Portage, par l'intermédiaire du Groupe de travail Dataverse Nord, devrait collaborer avec des intervenants clés pour établir un service Dataverse Nord national, ancré sur un modèle d'affaires pérenne et hébergé par Scholars Portal aux bibliothèques de la University of Toronto. (Modèles d'affaires)

Les membres de l'ABRC devraient adopter un modèle de métadonnées de base commun pour les dépôts Dataverse et opter pour des vocabulaires et des identifiants normalisés. (Métadonnées)

Les membres de l'ABRC devraient adopter une approche commune pour l'étiquetage, la description et l'organisation des fichiers de données dans les dépôts Dataverse. (Métadonnées)

Dataverse Nord devrait travailler de près avec le GEF pour développer du matériel de formation multimodal et prioritairement des manuels d'introduction, des vidéos et des ateliers ciblant les bibliothécaires et le personnel en bibliothèque. (Formation)

Tous les détails et toutes les notes explicatives justifiant ces recommandations se trouvent dans le rapport qui suit.

Dataverse North Business Models Group Activity Report & Recommendations

Membership

Eugene Barsky (University of British Columbia)
Corey Davis (COPPUL)
Alan Darnell (Scholars Portal)
Jason Flynn (Dalhousie University)
Lisa Goddard (University of Victoria)
Meghan Goodchild (Queen's University/Scholars Portal)
Amber Leahey (Scholars Portal)
Erin MacPherson (Dalhousie University)
Pierre Roberge (UQAM)
Brienne Selman (University of Winnipeg)
Lee Wilson (Portage/ACENET)

Background

The Portage Network's Dataverse North Working Group¹ is developing a community of practice for Dataverse in Canada. As part of its work, the group has been looking at opportunities that could be addressed by nationally coordinated strategies, including hosting services provided by regional Dataverse providers. Specifically, the Working Group was tasked to:

- Coordinate and develop a framework for Dataverse hosting and support services for designated libraries or other special interests that do not currently have a place to deposit research data.
- Explore a common business model to level the access to universities across Canada.²

Over the past six months, the Business Models Subgroup has undertaken an evaluation of the current Dataverse landscape, including information gathering and an assessment of hosted Dataverse services and library users across Canada. The Subgroup completed two surveys in the Fall of 2017 of both institutional Dataverse users and Dataverse hosting providers in Canada. Using the results from the surveys and a subsequent environmental scan of existing repository service models, the Subgroup developed several potential models for how the Dataverse community could proceed. In evaluating these models, it became clear that a single national Dataverse service hosted by an experienced service provider would work best in the Canadian context, where funding is limited and expertise is widely dispersed. While the

¹ Portage's Dataverse North Working Group - <https://portagenetwork.ca/wp-content/uploads/2017/09/DataverseNorthWG.pdf>

² <https://portagenetwork.ca/wp-content/uploads/2017/09/DataverseNorthWG.pdf>

Subgroup members recognize that any national effort should adequately recognize and support those institutions and regions that choose to operate their own repository infrastructure, we believe that significant benefits would accrue to the Canadian research community through a unified national service.³

Recommendation

Portage, through its Dataverse North Working Group, should work with key stakeholders⁴ to establish a national Dataverse North service, based on a sustainable business model⁵ and hosted by Scholars Portal at the University of Toronto Libraries. This will enable all Canadian researchers and the entire academic library community to effectively utilize a robust, scalable, affordable, and open research data repository platform that aligns with the suite of national Research Data Management (RDM) Services currently under development by Portage through its Networks of Expertise. This service would engage with other Dataverse providers in Canada to support and strengthen regional and institutional efforts through the Dataverse North Working Group.

Key benefits of a national Dataverse North service

While the centralization of repository infrastructure at a national level does not necessarily make sense in every domain, this group believes that a national service will significantly benefit the Canadian research community in a number of important ways:

- *Realizing true economies of scale.* While recognizing that some institutions or regions will choose to operate their own research data repositories for a variety of sound reasons, it is also important to understand that the greatest threat to digital materials over the long term is often economic.⁶ The economies of scale realized through a national service could provide the kind of cost optimizations needed to effectively and sustainably manage digital assets over time: “collaboration and federation can help to manage, share, and reduce costs.”⁷ A national service will have a significant impact on the ability of many small and medium-sized institutions to develop institutional RDM repository services.

³ This is of particular moment, as the Tri-Agency Digital Data Management Mandate approaches in 2018.

⁴ University of Toronto Libraries, Scholars Portal, OCUL, the other regional academic library consortia (COPPUL, BCI, and CAUL), and existing Canadian Dataverse providers (e.g. University of British Columbia, University of Alberta, University of Manitoba, Dalhousie, University of New Brunswick, and others)

⁵ This will likely be a mix of structural funding provided through Portage from the federal government, and from the academic library community through annual service fees, and from project or grant funding in areas such as feature development and other enhancements.

⁶ <http://blog.dshr.org/2013/10/the-major-threat-is-economic.html>

⁷ OECD (2017), "Business models for sustainable research data repositories", OECD Science, Technology and Industry Policy Papers, No. 47, OECD Publishing, Paris. Available at <http://dx.doi.org/10.1787/302b12bb-en>

- *Commitment to responsible data management.* A national Dataverse instance demonstrates a commitment to data management, including the Tri-Agency mandates. The Tri-Agency Statement of Principles on Digital Data Management outlines the roles and responsibilities of researchers, institutions, research communities and funders. A national Dataverse instance will assist in meeting some of these roles and responsibilities by providing a place to securely deposit, promote and share data.
- *Provides equal access to data services for all institutions.* A national Dataverse will provide access to RDM platforms and services to small and medium sized institutions that may be under-resourced in this area.
- *Help researchers and journal publishers navigate a complex environment.* Researchers looking to increase visibility and discoverability of their data, and to fulfill deposit mandates, and journals looking to manage the submission, review, and publication of data associated with published articles, must navigate an increasingly complex environment associated with RDM. A well-publicized national service will help declutter this environment by providing a well-articulated suite of services available to all researchers and university publishers in Canada.
- *Strengthening technical staffing and other support services.* Based on our recent survey of Canadian Dataverse users, local support is limited and most institutions report that librarians and others undertake Dataverse-related activities as a part of broader responsibilities. A national service could provide centralized technical support, and at the same time encourage the collective creation of materials to support, for example, advocacy and awareness, outreach activities, the creation of metadata templates, and discipline specific guidance, with everyone benefiting from working with the same version of software system.
- *Leveraging existing expertise and infrastructure.* Scholars Portal runs one of the largest Dataverses in Canada, in addition to other established technical library services on infrastructure at the University of Toronto Libraries. UTL/Scholars Portal Data Centre is a secure environment that conforms to industry best practices for maintaining data integrity and longevity, and UTL/Scholars Portal staff are able to upgrade software on a regular basis to enable the latest features and fixes, and to address the latest security vulnerabilities.
- *Pooling resources for new feature development.* Canadian Dataverse users have indicated a number of new features they would like to see, including better visualization tools, data curation support, file organization, media streaming, digital preservation, and support for large files, to name a few. New feature development would more efficiently take place on a single platform where resources are pooled nationally; and, where applicable, the code-base for enhancements and new features could be made openly

available via Github (or a similar service) so that other Canadian Dataverse providers might take advantage of these developments.

- *Improving funding opportunities.* A national service with broad participation across the country would also be a more attractive recipient for grant funding and research dollars, such as that made available through CANARIE⁸ and other CFI and even Tri-Agency funding sources. Developments to a national Dataverse North system would also improve outcomes for future integration with other RDM systems and tools in support of RDM workflows.
- *Supporting local and regional efforts.* Some institutions and regions might have compelling reasons to run their own Dataverse or other research data repositories (local expertise and capacity with different platforms, privacy impacts and related local or provincial policies, policies and practices in support of local or regional culture, etc.). A national service would work constructively with these institutions or regions running their own Dataverses and help them build capacity by invigorating the Canadian research data management community of practice for the benefit of all. We are seeking to build and foster relationships across Canada, while working in good-faith to develop a shared national service where it may be beneficial to all.
- *Alignment with national RDM efforts.* A national Dataverse instance will be better aligned with, and poised to take advantage of, other national RDM services currently in development. For example, a single instance of Dataverse can more easily work with Preservation Service Providers to meet the preservation processing specifications for research data outlined by the Preservation Expert Group (PEG),⁹ receive curation staffing support from the regionally distributed, nationally coordinated (and technology agnostic) curation network being developed by the Portage Curation Expert Group (CEG).
- *The ability to influence international efforts.* A national Dataverse instance will be able to influence international efforts that intersect with research data repository developments, such as the work being done under the auspices of the Digital Curation Centre (DCC) and the Research Data Alliance (RDA).

Guiding principles for establishing and operating a national service

Research data repositories are an increasingly important component of the digital research infrastructure and open science landscape (contributing to its economic and social benefits). Moreover, research policy makers and funders increasingly mandate open data for publicly funded research. A national Dataverse North service will be guided by principles that will ensure ongoing, sustainable, accountable, and responsive operations.

⁸ CANARIE software funding call for 2018 - <https://www.canarie.ca/software/funding/>

⁹ Can footnote PEG White Paper once it is published.

- *Developing a sustainable and equitable business model.* A sustainable business model will take into account how cost drivers (e.g. data volume, frequency of deposit, mix of users, levels of curation) and available revenue sources (e.g. structural funding from the federal government, service fees from libraries, support from participating institution, and grant funding) will adequately scale to meet future demand. Stakeholder identification and engagement will be key to articulating this model and demonstrating value over time. At the same time, in order to improve equity across Canada, we anticipate there being significant in-kind support offered to offset traditional cost models which may be prohibitive to some institutions and regions in Canada. Different cost models are being explored and we are anticipating a mix of funding from institutions, regions, and government.
- *Cultivating a community of practice.* Repository services will not be effectively utilized without encouraging a community of practice to support capacity building across organizations. A national service will support the Dataverse North Working Group in its efforts to develop a community of practice for libraries using or interested in using the Dataverse repository platform for research data in Canada.
- *Community-led, community-owned.* A national Dataverse North service will be offered by the academic library community¹⁰ in order to provide Canadian researchers with effective mechanisms to share, digitally preserve, and get credit for their data.
- *Collaboration.* A national Dataverse North service will work with other service providers and platforms like the Federated Research Data Repository (FRDR), institutional and regional repositories, and Canadian-based disciplinary repositories, in order to support research data management in Canada generally, and to look for collaborative opportunities around feature development, enhanced functionality, and cost-sharing where possible.
- *Transparency.* “Transparency of digital curation costs will help data repositories identify greater efficiencies and pinpoint potential optimisations. Insight into how and why peers target their investments can lead to the better use of resources, help identify weaknesses and drivers in current practices, and inspire innovations.”¹¹
- *Consultation.* A national service will be created in consultation with researchers, librarians, developers, and other stakeholders within the RDM community, as well as

¹⁰ “Canadian university libraries have a long history of the kinds of collaborations required in the multi-stakeholder RDM environment, deep experience in developing programs to advance research, and critical expertise in preservation.” https://portagenetwork.ca/wp-content/uploads/2016/06/IATUL2016_Multi_Stakeholder_Engagement_in_RDM.pdf

¹¹ OECD (2017), "Business models for sustainable research data repositories", OECD Science, Technology and Industry Policy Papers, No. 47, OECD Publishing, Paris. Available at <http://dx.doi.org/10.1787/302b12bb-en>

with existing Dataverse service providers in Canada, to ensure that the work done at the national level is as supportive as possible of institutional or regional efforts.

- *Digital preservation.* A national service will be developed with long-term preservation in mind, including the development of integrations with preservation management systems like Archivematica, and preservation storage services provided by institutions or regions (e.g. university-based storage services, OCUL's Ontario Libraries Research Cloud¹², and COPPUL's WestVault¹³)

Relation to other Portage Services and Platforms

Portage, through its Network of Expertise and in partnership with library consortia, institutions, and other infrastructure partners, is coordinating the development of a suite of national services for RDM that will better enable academic libraries to serve their researcher communities. At this stage, the groundwork is being laid for services related to data management planning (DMPEG), curation (CEG), preservation (PEG), discovery (DDEG), and repositories for the storage, description, and publication of research data (DVN & FRDR), in addition to training modules for a wide range of RDM-related topics (TEG). As a part of this broader landscape, DVN will work in conjunction with existing and planned initiatives to support the vision of providing seamless and equitable access to RDM services for Canadian researchers and institutions.

Summary

Research repositories are an essential part of the infrastructure for open scholarship and open science. Research data repositories provide for the long-term stewardship of research data, thus enabling verification of findings and the re-use of data. They bring considerable economic, scientific, and social benefits. Hence, it is important to ensure the sustainability of research data repositories, especially in Canada, where we do not have central funding for open scholarship in the country. Many research data repositories are largely dependent on public funding. The key policy question to be addressed is how this funding is most effectively provided - by what mechanism and from what source? There are advantages and disadvantages of various business models in different circumstances that can greatly affect data repository operations.

The key recommendation put forward by this paper is the development of a Portage Dataverse North Service, hosted by Scholar's Portal, that will enable all Canadian researchers and the academic library community to effectively utilize a robust, scalable, and affordable research data repository platform.

Related documents

¹² Ontario Libraries Research Cloud - <https://cloud.scholarsportal.info/>

¹³ COPPUL's WestVault - COPPUL's [WestVault](#)

- [Summary of the Dataverse North Business Models Group Dataverse Providers and Institutional Dataverse Users/Clients Survey](#)
- [Dataverse business model evaluation report](#)

Dataverse North Metadata Group Activity Report and Recommendations

Membership

Alexandra Cooper, Queen's University
Amber Leahey, Scholars Portal
Laure Perrier, University of Toronto
Michael Steeleworthy, Wilfrid Laurier University
Sally Taylor, University of British Columbia
Dee Wallace, University of Winnipeg

Background

The Portage Network's Dataverse North Working Group (DVNWG) is a community of practice that aims to coordinate national strategies and services for Dataverse, a robust, mature research data management platform. Its Metadata group (hereafter referred to as DVN Metadata group) is a team of librarians working to develop metadata best practices for Canadian users of Dataverse, provide guidance on project-level and file-level metadata creation and organization, and promote interoperability with and an understanding of DDI, the Data Documentation Metadata Standard, which is the metadata backbone of Dataverse.

In 2017, the DVN Metadata group developed a two-year work plan that includes the following objectives.

Focus for first year

Objective 1: Recommendations for default CARL metadata template, including required/optional fields and guidance on how to interpret fields

Objective 2: Recommendations for labeling, describing and organizing files (including formats)

Focus for second year

Objective 3: Training

Objective 4: Recommendations for DOIs and other identifiers (e.g. ORCID)

Objective 5: Recommendations for default CARL metadata template(s) for specific disciplines (i.e. Social Sciences, Earth and Environmental Sciences, Health and Life Sciences, others as needed)

These objectives, and the order in which they are to be met, were developed through the group's consideration of needs within the current Canadian RDM ecosystem, the existing skill sets of both the librarians whose work already involves research data services and those who

are new to the field, as well as resource availability. Our deliberations produced Year One objectives to develop recommendations for a default CARL metadata template (O1) and recommendations for labeling, describing, and organizing files (O2). Our rationale for working toward these goals in the first year is strong: O1 can offer both new and experienced Dataverse users guidance on developing project-level and file-level metadata according to established DDI best practices; O2 can create nation-wide best practices for file description and organization. Both objectives create metadata and file organizational supports for data managers and data stewards, which will improve access, use, and preservation in either local or federated Dataverse models.

Recommendation for default CARL metadata template (O1)

One of Dataverse's key benefits is the flexibility in the project-level metadata it affords to data stewards and users of the system. Using the DDI data description standard at its core, Dataverse provides an expansive metadata schema that is well-suited to the social, behavioral, economic, and health sciences, and is also easily adapted to the humanities, pure, applied, and environmental sciences, as well as other disciplines. However, this flexibility can sometimes come at a cost as data managers and self-deposit users, especially those who are new to Dataverse, DDI, or RDM, may not know which fields to use or how to interpret their content. Stub records, incorrect metadata, and unclear content can stifle the discovery and use of data.

The DVNWG Metadata group recommends:

- Developing a common CARL metadata template to be deployed on Canadian instances of Dataverse.

This template and its supporting materials will provide guidance to administrators, managers, and users in the creation and cleaning of their project-level metadata. It distinguishes between required, recommended, and optional fields, explains how a field is interpreted, and provides examples of use.

Process

In order to solicit input and take advantage of existing documentation, we ran a survey of Dataverse users and compiled a list of resources from other institutions. A [brief survey](#) was sent to the CanLib-Data listserv and the Dataverse North listserv on June 28, 2017. The questions asked about metadata templates (have any been created, required fields, are there ones for specific disciplines), if self-publishing is allowed, and best practices for describing research data.

We used the [Dataverse Metadata v4.x Citation sheet](#) (from Harvard's Institute for Quantitative Social Science) and edited it to recommend required, recommended, and optional fields, provided guidance on how to interpret the fields, and provided examples on how to use the

fields. In addition, we used the [DataCite Schema](#), [<odesi> Best Practices Document](#), and [FRDR metadata template](#) to ensure consistency with existing metadata practices.

Issues

During discussions, the group identified the following issues with respect to the default metadata template that will need to be addressed.

- It may be a challenge to develop a general use case that is suitable for multiple disciplines.
- In attempting to align “required” fields in the template with those in the DataCite schema, we are not sure what to do with “Resourcetype” because there is no equivalent field in Dataverse. In Dataverse, resource type is identified at the file level not the Citation level. The field “KindofData” in Dataverse is similar but not the same.
- Information associated with Rights or Licenses is not captured in the Citation metadata. This is part of a larger topic (e.g. local considerations) and therefore requires a broader conversation outside of metadata.
- Different institutions use different versions of Dataverse (v3.6 versus v4.x). Decisions made for the recommended metadata template were made based on v4.x and may not be applicable to v3.6.

Recommended metadata template

The following table describes required, recommended, and optional fields for a CARL Dataverse metadata template. The entire template, including examples, is available in *Appendix: Metadata Template*.

Required fields

- mandatory in the system

Name	Title	Description
title	Title	Full title by which the Dataset is known.
authorName	Name	The author's Family Name, Given Name or the name of the organization responsible for this Dataset.
datasetContactName	Name	The contact's Family Name, Given Name or the name of the organization.
datasetContactAffiliation	Affiliation	The organization with which the contact is affiliated.
datasetContactEmail	E-mail	The e-mail address(es) of the contact(s) for the Dataset. This will not be displayed.

dsDescriptionValue	Text	A summary describing the purpose, nature, and scope of the Dataset.
subject	Subject	Domain-specific Subject Categories that are topically relevant to the Dataset.
producerName	Name	Producer name

Recommended fields

- considered best practice but not required; depending on type of data, this field may not be applicable

Name	Title	Description
authorAffiliation	Affiliation	The organization with which the author is affiliated.
authorIdentifierScheme	Identifier Scheme	Name of the identifier scheme (ORCID, ISNI).
authorIdentifier	Identifier	Uniquely identifies an individual author or organization, according to various schemes.
keywordValue	Term	Key terms that describe important aspects of the Dataset. Can be used for building keyword indexes and for classification and retrieval purposes. A controlled vocabulary can be employed. The vocab attribute is provided for specification of the controlled vocabulary in use, such as LCSH, MeSH, or others. The vocabURI attribute specifies the location for the full controlled vocabulary.
producerAffiliation	Affiliation	The organization with which the producer is affiliated.
productionDate	Production Date	Date when the data collection or other materials were produced (not distributed, published or archived).
productionPlace	Production Place	The location where the data collection and any other related materials were produced.
contributorType	Type	The type of contributor of the resource.
contributorName	Name	The Family Name, Given Name or organization name of the contributor.
distributorName	Name	Distributor name
distributorAffiliation	Affiliation	The organization with which the distributor contact is affiliated.
distributionDate	Distribution Date	Date that the work was made available for distribution/presentation.
depositor	Depositor	The person (Family Name, Given Name) or the name of the organization that deposited this Dataset to the repository.
dateOfDeposit	Deposit Date	Date that the Dataset was deposited into the repository.

timePeriodCoveredStart	Start	Start date which reflects the time period covered by the data, not the dates of coding or making documents machine-readable or the dates the data were collected.
timePeriodCoveredEnd	End	End date which reflects the time period covered by the data, not the dates of coding or making documents machine-readable or the dates the data were collected.
dateOfCollectionStart	Start	Date when the data collection started.
dateOfCollectionEnd	End	Date when the data collection ended.
kindOfData	Kind of Data	Type of data included in the file: survey data, census/enumeration data, aggregate data, clinical data, event/transaction data, program source code, machine-readable text, administrative records data, experimental data, psychological test, textual data, coded textual, coded documents, time budget diaries, observation data/ratings, process-produced data, or other.
subtitle	Subtitle	A secondary title used to amplify or state certain limitations on the main title.
publicationCitation	Citation	Other identifier that corresponds to this Dataset.
grantNumberAgency	Grant Agency	Grant Number Agency
grantNumberValue	Grant Number	The grant or contract number of the project that sponsored the effort.
seriesName	Name	Name of the dataset series to which the Dataset belongs.
seriesInformation	Information	History of the series and summary of those features that apply to the series as a whole.

Optional fields

- good to use, but not essential for best practices

Name	Title	Description
alternativeTitle	Alternative Title	A title by which the work is commonly referred, or an abbreviation of the title.
alternativeURL	Alternative URL	A URL where the dataset can be viewed, such as a personal or project website.
otherIdAgency	Agency	Name of agency which generated this identifier.
otherIdValue	Identifier	Other identifier that corresponds to this Dataset.
dsDescriptionDate	Date	In cases where a Dataset contains more than one description (for example, one might be supplied by the data producer and another prepared by the data repository where the data are deposited), the date attribute is used to distinguish between the two descriptions. The date attribute follows the ISO convention of YYYY-MM-DD.

keywordVocabulary	Vocabulary	For the specification of the keyword controlled vocabulary in use, such as LCSH, MeSH, or others.
keywordVocabularyURI	Vocabulary URL	Keyword vocabulary URL points to the web presence that describes the keyword vocabulary, if appropriate. Enter an absolute URL where the keyword vocabulary web site is found, such as http://www.my.org .
topicClassValue	Term	Topic or Subject term that is relevant to this Dataset.
topicClassVocab	Vocabulary	Provided for specification of the controlled vocabulary in use, e.g., LCSH, MeSH, etc.
topicClassVocabURI	Vocabulary URL	Specifies the URL location for the full controlled vocabulary.
publicationIDType	ID Type	The type of digital identifier used for this publication (e.g., Digital Object Identifier (DOI)).
publicationIDNumber	ID Number	The identifier for the selected ID type.
publicationURL	URL	Link to the publication web page (e.g., journal article page, archive record page, or other).
notesText	Notes	Additional important information about the Dataset.
language	Language	Language of the Dataset
producerAbbreviation	Abbreviation	The abbreviation by which the producer is commonly known. (ex. IQSS, ICPSR)
producerURL	URL	Producer URL points to the producer's web presence, if appropriate. Enter an absolute URL where the producer's web site is found, such as http://www.my.org .
producerLogoURL	Logo URL	URL for the producer's logo, which points to this producer's web-accessible logo image. Enter an absolute URL where the producer's logo image is found, such as http://www.my.org/images/logo.gif .
distributorAbbreviation	Abbreviation	The abbreviation by which this distributor is commonly known (e.g., IQSS, ICPSR).
distributorURL	URL	Distributor URL points to the distributor's web presence, if appropriate. Enter an absolute URL where the distributor's web site is found, such as http://www.my.org .
distributorLogoURL	Logo URL	URL of the distributor's logo, which points to this distributor's web-accessible logo image. Enter an absolute URL where the distributor's logo image is found, such as http://www.my.org/images/logo.gif .
softwareName	Name	Name of software used to generate the Dataset.
softwareVersion	Version	Version of the software used to generate the Dataset.
relatedMaterial	Related Material	Any material related to this Dataset.
relatedDatasets	Related Datasets	Any Datasets that are related to this Dataset, such as previous research on this subject.

otherReferences	Other References	Any references that would serve as background or supporting material to this Dataset.
dataSources	Data Sources	List of books, articles, serials, or machine-readable data files that served as the sources of the data collection.
originOfSources	Origin of Sources	For historical materials, information about the origin of the sources and the rules followed in establishing the sources should be specified.
characteristicOfSources	Characteristic of Sources Noted	Assessment of characteristics and source material.
accessToSources	Documentation and Access to Sources	Level of documentation of the original sources.

Model Dataverse dataset

The following Dataverse dataset has been created by the DVN Metadata Group to demonstrate fully-described project-level metadata according to the proposed CARL Metadata template. Required, Recommended, and Optional fields have all been used according to their descriptions above:

- Social Media Use Among Teens [Canada] - <http://dx.doi.org/10.5072/FK2/TOXB6Q>

Real world examples of Dataverse datasets

The following Dataverse datasets have been identified by the Metadata Group as real-world examples of good project-level metadata within various Canadian Dataverse instances. These datasets are actual research data from different disciplines and demonstrate Dataverse's ability to describe research data from different subject domains within its default citation metadata schema:

- Epidemiology of Neuropathic Pain in Canada, 2015 - <http://hdl.handle.net/10864/11426>
- Soil moisture over Canadian Arctic tundra: Trail Valley Creek, Northwest Territories [Canada] dataset, 2017 - <http://dx.doi.org/10.5683/SP/GSFLE3>
- Disability and Accessibility in Canada, 2015 - <http://dx.doi.org/10.5683/SP/FSJ0TO>
- Violence Risk Appraisal Guide - Revised, 2013 - <http://hdl.handle.net/10864/12053>

Recommendations for labeling, describing & organizing files (O2)

In addition to creating good metadata for a study, it's important to label, describe and organize the data files and associated documentation to make it easier for data re-use.

The DVNMG Metadata group recommends:

1. Developing best practices documentation for labeling, describing and organizing files (including file formats) using the content below and in consultation with the DVN WG Training group. Should there be one Canadian Dataverse in the future, we envision having Portage-branded documentation and consistent practices regarding file formats.
2. Developing a short list of standardized tags in Dataverse at the file level (e.g. data, documentation) to improve organization of files and refine searching.
3. Providing feedback to the Dataverse development community regarding ways to improve file organization and sorting.

Process

Rather than starting from scratch, we took advantage of existing online materials that provide guidance to users. We identified and evaluated sources from a number of institutions, including one Portage member. Key points for each area are summarized below and serve as the basis for best practices documentation. In addition, we discussed the challenge of organizing files in Dataverse and ways in which it could be improved.

Issues

In some cases the development of best practices and recommendations outlined in this section conflict with the current functionality in Dataverse. There are several issues related to file organization, file sorting, and uploading files within Dataverse, including:

- Earlier versions of Dataverse offered file categorization, which allowed for organization. Newer versions (4.x) have deprecated categories in place of tags, which facilitate search functions by faceting results. However, these tags offer no organizing function at the dataset level. The group is aware of a proposed Dataverse development that could organize files according to tags (see screenshot below);
- Following this model, in developing a short list of standardized tags for institutions to adopt as a controlled vocabulary, there isn't a framework for adding custom controlled vocabularies for tags in the Dataverse system.
- Currently it is not possible to customize file sorting at the dataset-level; options are either alphabetical or by date of upload.
- Creating a standard or best practice for all disciplines is difficult because of differences between disciplines.

Example of potential enhancements to Dataverse for file organizing and sorting:

Dataverse About Support Contact Gary King

Gary King Dataverse

Harvard Dataverse > Gary King Dataverse > Sample Data

View Dataset Versions Metrics 1,776 Downloads

Sample Data

Gary King, 2015, "Sample Data", http://hdl.handle.net/1902.1/SAMPLE_UNF:3:oSMPLOxXX0o0xXXxxXXxxx== Harvard Dataverse [Distributor] V2 [Version] Download Citation

If you use these data, please add this citation to your scholarly resources. Learn about [Data Citation Standards](#).

Description
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed blandit malesuada eleifend. Nulla consequat lacus non tellus interdum, in sagittis mi fermentum. Vestibulum sagittis turpis eu tempus aliquam. Nunc quis tellus lacus. Ut blandit dui a turpis sagittis commodo. In laoreet lectus non vulputate sollicitudin. Vestibulum mollis est tellus, sit amet dapibus leo porta quis.

Subject
Social Sciences

Files Metadata Terms Provenance Versions

Search these data files... Find Upload Files

1 to 5 of 55 Data Files Sort Edit Files Download

Tag
War (1)
Data (3)
Documentation (1)

File Type
Adobe PDF (1)
MS Excel (1)
Tabular Data (2)
Plain Text (1)

Access
Public (3)
Restricted (2)

50by1000.tab
Jul 31, 2015
Tabular Data - 56.8 KB - 22 Downloads
50 Variables, 1,000 Observations - UNF:6:x10r+Q9EK6oF/BMi+eKzGw==
Default example sample pretend description explanation outline words.
Data

Fearon_Laitin_2003.dta
Jul 31, 2015
Tabular Data - 2.1 MB - 7,510 Downloads
50 Variables, 1,000 Observations - UNF:6:x10r+Q9EK6oF/BMi+eKzGw==
Default example sample pretend description explanation outline words.
Civil War Replication Data

Labeling files

By labeling files logically and consistently, both the original creator and future users of the data can more easily identify the contents.

Best practices:

- Plan ahead with overall architecture and conventions for naming files (consider the order of the characters for logical sorting)
- Keep names concise but descriptive (less than 25 characters)
- Avoid spaces, dots and special characters (since can be interpreted as commands in some operating systems)
- Use underscores or hyphens between words (e.g. Project_Galapagos) or capitalize first letters (e.g. ProjectGalapagos)
- Format dates consistently (e.g. YYYYMMDD)
- Include versioning where appropriate
- Label files independently of folder structure or storage location to avoid ambiguity

- Maintain a readme file with explanations of any abbreviations used in file names

Common elements used in file names:

- Project name, abbreviation or number
- Type of data
- Location
- Name of creator or initials or research team
- Version number
- Creation date
- File extension

Example: ProjectGalapagos_MarineIguana_Counts_SantaCruz_ST_20180122.csv

Recommended sources:

- [UBC Library File Naming Guidelines](#) (1 page pdf)
- [Stanford Libraries File Naming Best Practices](#) (2 page pdf)
- [Oxford File Naming](#) (2 page pdf)
- [MANTRA Organizing Data](#) (course module)
- [UK Data Service File Structure and File Names](#) (web page)
- [UK Data Service Versioning](#) (web page)

Describing files (e.g. metadata, readme files)

By describing files appropriately, users will better understand why and how the data was collected and analyzed, what the files contain and how they relate to each other, and within a file, how the variables are defined.

Best practices:

- Provide documentation at study level (i.e. research question, methodology)
- Provide documentation at the file or database level (i.e. how files relate to each other, software required, explanation of changes between versions)
- Provide documentation at variable or item level (i.e. variable names, labels, descriptions, units of measurement) in an accompanying separate codebook or data dictionary.

Examples: readme files, study descriptions, protocols, questionnaires, codebooks, data dictionaries

Recommended sources:

- [MANTRA Documentation, Metadata, Citation](#) (course module)
- [UK Data Service Document your Data](#) (web page)

Organizing files (e.g. tags, hierarchical folder structures)

For studies with multiple datasets and pieces of documentation (e.g. readme file, study description, codebook, etc.), users can more easily identify which files to download when they are grouped in a logical way.

Best practices:

- Use a standardized list of tags to distinguish between data files and documentation files.
- For data with a hierarchical folder structure, it is best to upload the dataset as a tar with gzip (.tar.gz) to bypass the unpacking of zip files upon upload in Dataverse.

File formats (e.g. non-proprietary, suitable formats for data analysis)

Files in proprietary formats typically require the software used to create them in order to open and read them. This can be a challenge if the software (or the version) is no longer available. In contrast, open or standard formats (in which the format is published) can be read by more than one application and are more likely to be readable in the future. One drawback of open formats or file migration can be a loss of information and quality.

Best practices:

- Where possible, save data files in a non-proprietary format so they can be read by others in the future.
- Ideally, develop an accompanying codebook / user guide to support reuse.
- If it's not possible or desirable to save in non-proprietary formats, use formats that have widespread adoption by researchers or industry (e.g. SPSS).

Examples of non-proprietary formats: txt, asc, csv, tab, html, xml, pdf, tif, jpeg, mp4, flac

Recommended sources:

- [UBC Library File Formats](#) (1 page pdf)
- [MANTRA File Formats and Transformations](#) (course module)
- [UK Data Service File Formats](#) (web page)
- [UK Data Service Recommended Formats](#) (web page)

Dataverse North Training Group Activity Report & Recommendations

Membership

The DVNT-WG consists of nine (9) members:

- Carrie Breton - University of Guelph
- James Doiron (*Chair*) - Member at Large
- Siobhan Hanratty - University of New Brunswick
- Shahira Khair - CARL/Portage
- Kaitlin Newson - Scholars Portal
- Andrew Nicholson - University of Toronto
- Carol Perry - University of Guelph
- Kathryn Ruddock - University of Calgary
- Barbara Znamirovski - Trent University

Former members:

- John Brosz - University of Calgary
- Maggie Jean Neilson - Acadia University

Mandate

The Dataverse North Training Working Group has two broad objectives: 1) to identify Dataverse training needs of librarians, library staff, researchers and students across the Canadian landscape and; 2) to provide evidence based recommendations on optimal strategies for developing and delivering training resources to meet these needs.

As such, the mandate of the Dataverse North Training Working Group (DVNT-WG) is to:

- a. Work with the Portage Training Expert Group around Dataverse Training
- b. Identify the kinds of training that are most desirable for librarians and library staff
- c. Identify the kinds of training that are most desirable for faculty, researchers, and graduate students
- d. Gather existing training documentation from Dataverse North Working Group members & related organizations
- e. Propose models for sharing docs & offering shared training sessions
- f. Assessment & evaluation of Dataverse as a tool for different types of data

Recommendations

At this time, DVNT-WG is able to offer some initial recommendations regarding Dataverse training activities. The initial recommendations provided within this report are primarily

informed by the descriptive analyses relating to the 'Portage Dataverse Training Needs' survey. A brief summary of these are available for review at:
<https://docs.google.com/document/d/1iCDrY6zmfuk2L1weMHtMy8WHPxedNAxr5yuHfKKBLI/edit>

Work on the development and delivery of freely available Dataverse training materials should be a priority activity of Portage and the Dataverse North Group

- Over 93% of respondents indicated that they would be interested in accessing and using free web-based Dataverse training materials developed by Portage.
- In alignment with other Portage platforms and services, all training materials should be made available bilingually.

The initial audience of focus for training materials should be librarians, library staff, and Dataverse service providers

- Almost 75% of the survey respondents were librarians/library staff.
- Librarians have indicated a strong desire to receive Dataverse training and knowledge building support
- It is important that Librarians and Dataverse service providers are able to provide expertise and advisement to researchers looking to use Dataverse, whether that be at a local (institutional) or national level.
- Portage recognizes the importance of building and maintaining strong relationships between the libraries and researchers. As such, building Dataverse knowledge and expertise among librarians, library staff, and Dataverse service providers - 'training the trainers' - is important, and will help to deliver better services to researchers.

Dataverse training resources should be developed with the consideration that, moving forward, both regional and national Dataverse service models may be in place.

- It is recognized that the Dataverse North Business Models group is recommending a national Dataverse service, and that this would in turn likely lead to support in terms of funding opportunities and integration with Portage discovery and curation tools, but that this would also not prevent institutions from running their own regional and institutional services.
- Should a national Dataverse service come to fruition, it is recommended that training resources and initiatives take this into consideration, as it will potentially provide an opportunity for standardizing and streamlining their development and delivery.
- Similarly, regardless of whether a national service moves forward, it is recommended that Dataverse training resources and opportunities be developed to be supportive at the local and regional levels as, based on survey feedback and the collective experiences of the working group, this is something that is both wanted and needed.
- Building off of the previous point, it is recognized that Portage and Dataverse North are in a position to help connect the different institutions hosting Dataverse platforms in

efforts both to glean valuable feedback regarding training, as well as to potentially help guide and deliver training activities.

Dataverse training materials should focus both on general and specific topics

- Survey respondents indicated high levels of interest across a wide range of Dataverse training topics.
- The top two specific topics of interest indicated were '*How to prepare data for deposit*' and '*Dataverse functionality*'. The third highest topic of interest indicated was '*General Information*'. It is therefore recommended that these be the initial ones focused upon.
- The development of flow charts for depositing data into Dataverse, or decision making trees for repository selection, is of specific interest, and these should therefore be considered in future training resources development work. This work should occur in collaboration with RDM-TEG and other expert groups, and possibly also in consultation with FRDR personnel.
- Almost 80% of respondents indicated that they use the analytic features provided within Dataverse, though these primarily involve the summary and descriptive statistics features. As such, it is recommended that some focus be placed towards 'general analysis' features and training.

A wide range of formats and methods should be used for the delivery of Dataverse training materials

- Respondents indicated that a wide range of Dataverse training delivery formats and methods are desired.
- The top three most desired formats and methods are: '*Text based manuals*', '*Recorded videos*' and '*In-person workshops*', and so it is recommended that these methods are prioritized in planning the delivery of training materials.
- Based on open text feedback, modifiable training materials, such as PowerPoint slides, text based materials, and handouts, are of interest and should be considered when developing training materials and resources.
- Regarding the previously mentioned flowchart(s) for depositing data into Dataverse, it is recommended that 'promotional' posters/handouts be considered, as these high level visual materials can be extremely useful to all stakeholder types and in both electronic and print formats.
- Over 75% of respondents indicated that they used Dataverse either '*once a month*' or '*once every few months*'. As such, many users require both novel and 'refresher' training, and this was also highlighted within the open text comments. It is recommended that there be training materials developed which are then freely available on demand, when users need them.

While the most popular data format being deposited into Dataverse is tabular data, training materials should also focus on other data formats

- 80% of respondents indicated that they had deposited tabular data into Dataverse, but many other data formats such as text based materials (60%), microdata (26%), spatial (22%), audio/visual (18%) and code (10%) are also formats that are actively being deposited into Dataverse.
- With this in mind, it is recognized that Portage and Dataverse North have a potential role to play in promoting the use of Dataverse for a wide range of data formats and so it is recommended that this is kept in mind when developing training materials and resources.

Dataverse North should work closely with RDM-TEG in developing and delivering Dataverse training materials and resources

- Dataverse North should work closely with RDM-TEG in efforts to ensure that training materials and resources developed are standardized and in alignment with their overall recommendations and other training initiatives being led by the group. Additionally, Dataverse training activities will be able to leverage and benefit from existing training initiatives, such as the [Portage RDM Training Assistance Request](#) form and processes in place to help deliver training.
- It is recommended that Dataverse North and RDM-TEG work together to jointly oversee the creation of working groups which will focus on developing Dataverse training resources.
- It is recommended that Dataverse North and RDM-TEG work together to determine optimal approaches for the delivery of Dataverse training, including who may be recruited to help deliver in person training sessions and workshops. To this end, it is recognized that the Dataverse team at Harvard University is a rich resource of knowledge and expertise, and so collaborative opportunities with them should be considered.

Work and objectives - status update

The Dataverse North Training Working Group (DVNT-WG) was formed in June 2017, holding its first teleconference call meeting on June 26. To date, the Dataverse North Training Group has held thirteen (13) teleconference call meetings. Minutes for all meetings are available for access and review at: <https://drive.google.com/drive/folders/0B30E7e-TRWdeb0dLN2ROaUhlZ0E>

As outlined within the Working Group Mandates, the DVNT-WG is focusing on a number of issues. A brief description of each of these, along with current status updates, are provided below.

a) Work with the Portage Training Expert Group:

Information sharing and consultation with the Portage Research Data Management Training Expert Group (RDM-TEG) has been occurring as needed. Both James Doiron and

Carol Perry are members of RDM-TEG and so are able to bring forward items to their meeting agendas for discussion and feedback. A 'Dataverse Training Needs' survey was administered in December 2017, and the information collected is currently being analysed. We will soon look to RDM-TEG to provide some feedback and guidance regarding Dataverse training activities, and provide recommendations for possible collaborations and helping to develop and deliver some of these. An overview of these will be provided in the final briefing report at the end of March.

b) Identify the kinds of Dataverse training most desirable for librarians and library staff

One of the primary initiatives to help identify librarians' Dataverse training needs has been the 'Portage Dataverse Training Needs Survey'. The survey was developed by DVNT-WG, and consists of twenty-seven (27) variables spanning three overarching categories: 1) Demographics; 2) Experience with Dataverse; and 3) Dataverse Training. The survey was bilingually administered in December 2017, and garnered 115 responses in total (106 English/9 French). Approximately 75% of the respondents identified themselves as librarians/library staff and are located at institutions across Canada.

The survey was designed to capture key information across these categories in order to help to identify librarians' knowledge of, and experience with, Dataverse. Additionally, the survey aims to glean an understanding of key Dataverse topics of interest for which librarians would like to see training materials developed for, as well as what formats of training are most desirable (e.g., web-based, in-person, training videos, text based manuals, etc).

Initial descriptive analysis of the survey data has been performed, with the results of these helping to inform the preliminary recommendations made later within this report. Additional analysis will be performed, with final recommendations to be made within the briefing report due at the end of March.

c) Identify the kinds of Dataverse related training that are most desirable for faculty, researchers, and graduate students.

The 'Portage Dataverse Training Needs Survey' was developed to include key demographic variables, as well as logic rules and skip patterns, in order to be able to additionally capture Dataverse experience and training needs from faculty, researchers and graduate students. As such, all status update information, including timelines, is the same for this activity as the one previously described in section B.

d) Gather existing training documentation from Dataverse North Working Group members and related organizations.

Work has been underway towards compiling existing training materials that may be of interest and use to the Canadian Dataverse community. In August 2017 we requested that the Chair of the DV North Advisory Group (Lisa Goddard) forward a request to the Chairs of each of the Portage Expert Groups to ask for any Dataverse related training resources that might be shared. We additionally asked members of the DV North group to provide any documentation and/or links to training resources that they have created or used. As well, the 'Portage Dataverse Training Needs Survey' included a question which asked respondents if they had Dataverse related training materials that they were able to share and, if so, provided contact information to facilitate this.

A [folder](#) was created in the DVNT-WG's Google Drive space for materials to be deposited/stored, along with a [spreadsheet](#) for identifying and describing them. Thus far, eight (8) sets of training materials have been deposited.

Forward work includes:

- Members of the DVNT-WG will explore the Dataverse project website, as well as the [Dataverse Users Google group](#) to see what documentation (outside of the user guides available via Dataverse) and training resources are available
- The Chair of DVNT-WG (James Doiron) will connect with the Dataverse team at Harvard to see if there are any training documentation and resources that they have, are aware of, or are planning to develop and are able to share.

Assessment of the collected Dataverse training materials will occur through March 2018, and a summary will be provided within the final briefing report.

e) Propose models for sharing documentation and offering shared training sessions

With the delivery of the 'Portage Dataverse Training Needs Survey' now completed, and analysis of the information collected now underway. This work will soon commence, and in collaboration with RDM-TEG, and recommendations will be offered within the final briefing report and the end of March.

f) Assessment and evaluation of Dataverse as a tool for different types of data

At the time of the writing of this report the work on this activity is currently underway. Part of this work is being informed by variables within the 'Portage Dataverse Training Needs Survey' and so we are now assessing the different 'types' of data that were identified, where there are existing 'fits' and 'gaps', and compiling recommendations to be made.

We have additionally determined that it will be useful to conduct a systematic scan and review of pertinent literature regarding pilots/case studies/evaluations of training methods and models relating to Dataverse platforms and usage. The Chair of the Portage Data

Curation Expert Group (Jay Brodeur) additionally provided some insights regarding interests and discussions on behalf of that group.

A framework for the literature review has now been developed. Compiling and review of the literature will occur through March 2018, and will help to inform recommendation to be made within the final briefing report.

Since its inception in June, 2017, the Dataverse North Training Working Group (DVNT-WG) has focused on a number of key activities aimed to address the goals outlined within its mandate. Primary activities thus far have included the development and delivery of a 'Portage Dataverse Training Needs' survey, the gathering of training documentation from Dataverse North Working Group members and related organizations, and the assessment and evaluation of Dataverse as a repository solutions and tool for different types of data which is, in part, being informed by a systematic review currently underway.

Based on activities completed to date, DVNT-WG is able to provide some initial recommendations regarding the Canadian Dataverse training landscape. It is the group's consensus that Dataverse training is important, and that work on the development and delivery of freely available and bilingual training materials focusing both on general and specific topics should be a priority activity of Portage and the Dataverse North Group. In order to maximize both their accessibility and effectiveness, these training resources should be delivered through a wide range of formats. While we recognize the importance of developing these materials for all stakeholder types, it is our recommendation that the initial audience of focus be librarians, library staff and Dataverse service providers. As well, we believe that it is important that Dataverse training resources be developed in such a way as to support both regional and national Dataverse service models which may be in place.

Moving forward, work as outlined previously within this report will continue, including further analysis of the 'Portage Dataverse Training Needs' survey data and the described systematic literature review. An update on all activities, as well as final recommendations, will be provided within a forthcoming final briefing paper.

Appendix - Dataverse North Metadata Template

Field name	Field Title (dataverseAlias)	Field description	Field type	Required/ Recommended/ Optional	Rationale/ Guidance on interpretation	Examples
title	Title	Full title by which the Dataset is known.	text	required	required by DV, FRDR and DataCite	Social Media Use Among Teens [Canada]
subtitle	Subtitle	A secondary title used to amplify or state certain limitations on the main title.	text	recommended (if applicable)		Main Survey
alternativeTitle	Alternative Title	A title by which the work is commonly referred, or an abbreviation of the title.	text	optional	Acronym or short form of full title	Youth Social Media Survey
alternativeURL	Alternative URL	A URL where the dataset can be viewed, such as a personal or project website.	url	optional		http://youthsocialmedia.org
otherId	Other ID	Another unique identifier that identifies this Dataset (e.g., producer's or another repository's number).	none	n/a		
otherIdAgency	Agency	Name of agency which generated this identifier.	text	optional		Youth Communication Development Project, Education Department, University Name
otherIdValue	Identifier	Other identifier that corresponds to this Dataset.	text	optional		2202
author	Author	The person(s), corporate body(ies), or agency(ies) responsible for creating the work.	none	n/a		
authorName	Name	The author's Family Name, Given Name or the name of the organization responsible for this Dataset.	text	required	required by DV, FRDR and DataCite	Doe, Jane
authorAffiliation	Affiliation	The organization with which the author is affiliated.	text	recommended		University Name
authorIdentifierScheme	Identifier Scheme	Name of the identifier scheme (ORCID, ISNI).	text	recommended		ORCID

authorIdentifier	Identifier	Uniquely identifies an individual author or organization, according to various schemes.	text	recommended		1111111
datasetContact	Contact	The contact(s) for this Dataset.	none	n/a		
datasetContactName	Name	The contact's Family Name, Given Name or the name of the organization.	text	required	required by DV	Doe, Jane
datasetContactAffiliation	Affiliation	The organization with which the contact is affiliated.	text	required	required by DV	University Name
datasetContactEmail	E-mail	The e-mail address(es) of the contact(s) for the Dataset. This will not be displayed.	email	required	Required by DV; could be an institutional email address	jdoe@email.com
dsDescription	Description	A summary describing the purpose, nature, and scope of the Dataset.	none	n/a		
dsDescriptionValue	Text	A summary describing the purpose, nature, and scope of the Dataset.	textbox	required	required by DV and recommended by DataCite	The Social Media Use among Teens survey was conducted by the Youth Communication Development Project to understand social media communication behaviours among youth in Canada. The survey collected responses from Canadian youth using an online questionnaire that asks about social media use including, platform type, frequency of use, activity type, and location of use. This information is supplemented with the respondent's demographic and household characteristics.

dsDescriptionDate	Date	In cases where a Dataset contains more than one description (for example, one might be supplied by the data producer and another prepared by the data repository where the data are deposited), the date attribute is used to distinguish between the two descriptions. The date attribute follows the ISO convention of YYYY-MM-DD.	date	optional	flag	2018-01-18
subject	Subject	Domain-specific Subject Categories that are topically relevant to the Dataset.	text	required	required by DV and recommended by DataCite	Social Sciences
keyword	Keyword	Key terms that describe important aspects of the Dataset.	none	n/a		
keywordValue	Term	Key terms that describe important aspects of the Dataset. Can be used for building keyword indexes and for classification and retrieval purposes. A controlled vocabulary can be employed. The vocab attribute is provided for specification of the controlled vocabulary in use, such as LCSH, MeSH, or others. The vocabURI attribute specifies the location for the full controlled vocabulary.	text	recommended		Social media
keywordVocabulary	Vocabulary	For the specification of the keyword controlled vocabulary in use, such as LCSH, MeSH, or others.	text	optional		Government of Canada Core Subject Thesaurus

keywordVocabularyURI	Vocabulary URL	Keyword vocabulary URL points to the web presence that describes the keyword vocabulary, if appropriate. Enter an absolute URL where the keyword vocabulary web site is found, such as http://www.my.org .	url	optional		http://www.thesaurus.gc.ca/recherche-search/mtwdk.exe?k=these&l=60&w=4790&n=1&s=5&t=2
topicClassification	Topic Classification	The classification field indicates the broad important topic(s) and subjects that the data cover. Library of Congress subject terms may be used here.	none	n/a		
topicClassValue	Term	Topic or Subject term that is relevant to this Dataset.	text	optional		Society and Culture
topicClassVocab	Vocabulary	Provided for specification of the controlled vocabulary in use, e.g., LCSH, MeSH, etc.	text	optional		Government of Canada Core Subject Thesaurus
topicClassVocabularyURI	Vocabulary URL	Specifies the URL location for the full controlled vocabulary.	url	optional		http://www.thesaurus.gc.ca/recherche-search/mtwdk.exe?k=these&l=60&n=0&s=cid&t=&w=97&h=SO%20Society%20and%20Culture
publication	Related Publication	Publications that use the data from this Dataset.	none	n/a		
publicationCitation	Citation	Other identifier that corresponds to this Dataset.	textbox	recommended (if applicable)	RelatedIdentifier (12) recommended by DataCite. Recommend full citation of the publication be included here, including a relatedidentifier for access.	Doe, Jane. (2017). Teen use of social media: analysis of self-reported communication behaviours. Journal of Social Media Use. Vol 1. Iss. 1, 2017. doi:10.0000/SP/TEST

publicationIDType	ID Type	The type of digital identifier used for this publication (e.g., Digital Object Identifier (DOI)).	text	optional		doi
publicationIDNumber	ID Number	The identifier for the selected ID type.	text	optional		doi:10.0000/SP/TEST
publicationURL	URL	Link to the publication web page (e.g., journal article page, archive record page, or other).	url	optional		http://openjournalarticle.org/2202-1
notesText	Notes	Additional important information about the Dataset.	textbox	optional		This survey was administered online. Mode of interview has been found to impact results, therefore it is not recommended that these results are compared with other survey results where the interview mode was telephone based.
language	Language	Language of the Dataset	text	optional		English-Canada
producer	Producer	Person or organization with the financial or administrative responsibility over this Dataset	none	n/a		
producerName	Name	Producer name	text	required	required by DataCite (but field is called Publisher and also includes distributor responsibilities); flag	Youth Communication Development Project
producerAffiliation	Affiliation	The organization with which the producer is affiliated.	text	recommended		University Name
producerAbbreviation	Abbreviation	The abbreviation by which the producer is commonly known. (ex. IQSS, ICPSR)	text	optional		YCDP
producerURL	URL	Producer URL points to the producer's web presence, if appropriate. Enter an absolute URL where the producer's web	url	optional		http://youthsocialmedia.org

		site is found, such as http://www.my.org .				
producerLogoURL	Logo URL	URL for the producer's logo, which points to this producer's web-accessible logo image. Enter an absolute URL where the producer's logo image is found, such as http://www.my.org/images/logo.gif .	url	optional		http://youthsocialmedia.org/image.png
productionDate	Production Date	Date when the data collection or other materials were produced (not distributed, published or archived).	date	recommended	note: required by DataCite (ID 5 - PublicationYear) but not by DV or FRDR; flag - tricky - data can be collected over range of time, then analysed, then finally produced, then distributed, then archived. The date used here should be the date that the data was publically released by the producer.	2016-01-11
productionPlace	Production Place	The location where the data collection and any other related materials were produced.	text	recommended	flag	Toronto, Ontario, Canada
contributor	Contributor	The organization or person responsible for either collecting, managing, or otherwise contributing in some form to the development of the resource.	none	n/a		

contributorType	Type	The type of contributor of the resource.	text	recommended	recommended by DataCite	Researcher
contributorName	Name	The Family Name, Given Name or organization name of the contributor.	text	recommended	recommended by DataCite	Doe, Jane
grantNumber	Grant Information	Grant Information	none			527776
grantNumberAgency	Grant Agency	Grant Number Agency	text	recommended (if applicable)		Social Sciences and Humanities Research Council (SSHRC)
grantNumberValue	Grant Number	The grant or contract number of the project that sponsored the effort.	text	recommended (if applicable)		CCB123456
distributor	Distributor	The organization designated by the author or producer to generate copies of the particular work including any necessary editions or revisions.	none	n/a	flag	
distributorName	Name	Distributor name	text	recommended		Data Services
distributorAffiliation	Affiliation	The organization with which the distributor contact is affiliated.	text	recommended		Queen's University Library
distributorAbbreviation	Abbreviation	The abbreviation by which this distributor is commonly known (e.g., IQSS, ICPSR).	text	optional		QUL
distributorURL	URL	Distributor URL points to the distributor's web presence, if appropriate. Enter an absolute URL where the distributor's web site is found, such as http://www.my.org .	url	optional		http://library.queensu.ca/data/services
distributorLogoURL	Logo URL	URL of the distributor's logo, which points to this distributor's web-accessible logo image. Enter an absolute URL where the distributor's logo image is found, such as http://www.my.org/images/logo.gif .	url	optional		http://www.queensu.ca/encyclopedia/sites/webpublish.queensu.ca.gencwww/files/images/l/logo/QueensLogo_colour.png

distributionDate	Distribution Date	Date that the work was made available for distribution/presentation.	date	recommended	flag	2018-01-22
depositor	Depositor	The person (Family Name, Given Name) or the name of the organization that deposited this Dataset to the repository.	text	recommended	The name of the person/institution who provided the dataset(s) to the archive (i.e. not necessarily the person doing the submission into DV).	Doe, Jane
dateOfDeposit	Deposit Date	Date that the Dataset was deposited into the repository.	date	recommended	Date that dataset(s) is received/accepted/accessioned by repository service (e.g. could be an email). Could also be a researcher depositing their own data. Needs clarification!	2018-01-15
timePeriodCovered	Time Period Covered	Time period to which the data refer. This item reflects the time period covered by the data, not the dates of coding or making documents machine-readable or the dates the data were collected. Also known as span.	none	n/a		
timePeriodCoveredStart	Start	Start date which reflects the time period covered by the data, not the dates of coding or making documents machine-readable or the dates the data were collected.	date	recommended		2015-03-20

timePeriodCoveredEnd	End	End date which reflects the time period covered by the data, not the dates of coding or making documents machine-readable or the dates the data were collected.	date	recommended		2015-06-21
dateOfCollection	Date of Collection	Contains the date(s) when the data were collected.	none		flag	2015-03-20 to 2015-06-21
dateOfCollectionStart	Start	Date when the data collection started.	date	recommended		2015-03-20
dateOfCollectionEnd	End	Date when the data collection ended.	date	recommended		2015-06-21
kindOfData	Kind of Data	Type of data included in the file: survey data, census/enumeration data, aggregate data, clinical data, event/transaction data, program source code, machine-readable text, administrative records data, experimental data, psychological test, textual data, coded textual, coded documents, time budget diaries, observation data/ratings, process-produced data, or other.	text	recommended	flag	Survey data
series	Series	Information about the Dataset series.	none	n/a		
seriesName	Name	Name of the dataset series to which the Dataset belongs.	text	recommended (if applicable)		General Social Survey (Statistics Canada)
seriesInformation	Information	History of the series and summary of those features that apply to the series as a whole.	textbox	recommended (if applicable)		Established in 1985, Canada's General Social Survey (GSS) program is a series of independent, annual, cross-sectional surveys. The overall objectives of the program is to gather data on social trends in order to monitor changes in the living conditions and well being of

						Canadians, and to provide information on specific social policy issues.
software	Software	Information about the software used to generate the Dataset.	none	n/a		
softwareName	Name	Name of software used to generate the Dataset.	text	optional	useful for specialized software/instruments (vs. common products)	SPSS
softwareVersion	Version	Version of the software used to generate the Dataset.	text	optional		24
relatedMaterial	Related Material	Any material related to this Dataset.	textbox	optional	RelatedIdentifier (12) recommended by DataCite.	https://dataverse.scholarsportal.info/file.xhtml?fileId=8601&datasetVersionId=3416
relatedDatasets	Related Datasets	Any Datasets that are related to this Dataset, such as previous research on this subject.	textbox	optional	RelatedIdentifier (12) recommended by DataCite.	https://dataverse.scholarsportal.info/dataset.xhtml?persistentId=hdl:10864/10677
otherReferences	Other References	Any references that would serve as background or supporting material to this Dataset.	text	optional		http://poll.forumresearch.com/category/1/national/
dataSources	Data Sources	List of books, articles, serials, or machine-readable data files that served as the sources of the data collection.	textbox	optional		https://dataverse.scholarsportal.info/file.xhtml?fileId=8567&version=RELEASED&version=5 https://dataverse.scholarsportal.info/file.xhtml?fileId=8597&version=RELEASED&version=5
originOfSources	Origin of Sources	For historical materials, information about the origin of the sources and the rules	textbox	optional		http://poll.forumresearch.com/accuracy-and-methodology/

		followed in establishing the sources should be specified.				
characteristicOfSources	Characteristic of Sources Noted	Assessment of characteristics and source material.	textbox	optional	Describes noteworthy aspects of the data collection.	All study variables were coded exclusively from descriptive material contained in institutional files by trained and experienced research assistants.(from metadata for Violence Risk Assessment Guide, Revised, 2013)
accessToSources	Documentation and Access to Sources	Level of documentation of the original sources.	textbox	optional	lists book(s), article(s), serial(s), and/or machine readable data file(s) that serve as the source(s) of the data collection	Westlaw, Quicklaw