INTEGRATING DATAVERSE AND ARCHIVEMATICA FOR RESEARCH DATA PRESERVATION

Meghan Goodchild
Scholars Portal &
Queen's University
Canada
meghan@scholarsportal.info
ORCID 0000-0001-7988-8046

Grant Hurley
Scholars Portal
Canada
grant@scholarsportal.info
ORCID 0000-0003-0172-4847

Abstract – Scholars Portal sponsored Artefactual Systems Inc. to develop the ability for the preservation processing tool Archivematica to receive packages from Dataverse, a popular tool for uploading, curating, and accessing research data. The integration was released as part of Archivematica 1.8 in 2018. This paper situates the integration project in the broader context of research data preservation; describes the scope and history of the project and the features and functionalities of the current release; and concludes with a discussion of the potential for future developments to meet additional use cases, service models and preservation approaches for research data.

Keywords – research data; Archivematica; preservation pipeline; Dataverse

Conference Topics – collaboration; technical infrastructure

I. INTRODUCTION

Between 2015 and 2018, Scholars Portal contracted Artefactual Systems Inc. to develop an integration between Dataverse, a popular repository tool for uploading, curating, and accessing research data, with Archivematica, a workflow application for creating preservation-friendly packages for long-term storage and management. Scholars Portal is the information technology service provider for members of the Ontario Council of University Libraries (OCUL), a 21-member consortium of academic libraries in the province of Ontario, Canada.[1] Founded in 2002, Scholars Portal is supported by OCUL members and operated under a service agreement with the University of Toronto Libraries. Our services support both research data management via a hosted, multi-institutional instance of Dataverse[2] and digital preservation services via Permafrost,[3] a hosted Archivematica-based service that pairs with the OCUL Ontario Library Research Cloud (OLRC) for preservation storage.[4] The Dataverse-Archivematica integration project was initially undertaken as a research initiative to explore how research data preservation aims might functionally be achieved using Dataverse and Archivematica together. The results of a proof-of-concept phase were developed into a working integration released as part of Archivematica version 1.8 in November 2018. This paper situates the integration project in the broader context of research data preservation in theory and practice; describes the scope and history of the project and the features and functionalities of the current release; and concludes with a discussion of the potential for future developments to meet additional use cases, service models and preservation approaches for research data.

[1] Scholars Portal: https://scholarsportal.info/.
[2] Scholars Portal's Dataverse instance: https://dataverse.scholarsportal.info/.
[3] Permafrost: https://permafrost.scholarsportal.info/.
[4] Ontario Library Research Cloud: https://cloud.scholarsportal.info/.

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

DOI: 10.1145/nnnnnnn.nnnnnnn

iPRES 2019

II. RESEARCH DATA PRESERVATION IN CONTEXT

In this paper, the term "research data" refers to a broad set of potential outputs from research activities across sectors and disciplines. The key uniting characteristic is that these materials stand as unique evidence supporting a set of research findings, whether scholarly, technical, or artistic [1]. Furthermore, these data may constitute the research findings themselves, such as in public statistical or geospatial data gathering. The communities of stakeholders who value research findings depend on the maintenance of the originary data sources in a trustworthy manner that privileges ensuring their continued authenticity, availability and reliability into the future. These concepts have been codified within the sector as the FAIR Principles for research data: findable, accessible, interoperable, reusable [2]. While the FAIR Principles do not specifically cite long-term preservation as a requirement, preservation is crucial to the continued ability to discover and access research data into the future [3]. The FAIR principles therefore link to the stewardship responsibilities that repositories take on behalf of stakeholders: in order to fulfill the FAIR principles, organizations with access to sustained resources and infrastructure must commit to ensuring the long-term maintenance of the materials under their care. The requirements for this maintenance are outlined in standards such as the Open Archival Information System reference model (ISO 14721)[5] and audit and certification frameworks including CoreTrustSeal,[6] nestor,[7] and *Audit and Certification of Trustworthy Data Repositories* (ISO 16363).[8] Repositories with stewardship responsibilities therefore seek to translate audit and certification requirements into repeatable practices to ensure that data are kept reliably into the future. There is a series of interrelated stages that make up the lifecycle required for responsible data curation and preservation over time, including creation and receipt, appraisal and selection, preservation actions, storage, and access and discovery [4]. One tool that implements some of these stages of the lifecycle is the research data repository web application Dataverse.[9]

Dataverse is developed and maintained as an open source tool by the Institute for Quantitative Social Science (IQSS) at Harvard University. It has been developed since 2006 [5]. A large open Dataverse instance is hosted by IQSS and 38 additional individual known installations of Dataverse exist throughout the world as of the time of writing [6]. While Dataverse was developed by members of the social science community, its use is not limited to any specific disciplinary area [5]. Users can deposit and describe their data files using general and discipline-specific metadata standards, generate unique identifiers, and assign access permissions. Institutions can enable self-deposit or mediated workflows for depositors, and offer Dataverse to researchers as a method of fulfilling funder requirements to deposit data in an accessible repository. Published datasets are searchable and downloadable and tabular data files can be explored using visualization tools within the platform itself.

Dataverse includes a suite of functions that contribute to the ability for a stewarding organization to reliably preserve research data. When it comes to data receipt, it enables efficient capture of materials from a researcher's individual computing systems through user-friendly upload tools, which tackles a major initial barrier of accessing data from the risky (and often inaccessible) environments of personal storage systems [7]. Researchers can also describe and contextualize their submissions through a variety of metadata fields and by linking to related publications and datasets. All submitted files receive MD5 checksums upon receipt that can enable verification of integrity over time. File format identification is conducted using JHOVE and displayed using a set of MIME-type tags [8]. Finally, a set of tabular data formats (SPSS, Strata, R.Data, CSV, and Excel) are converted to tabular text data files upon ingest, and citation-related metadata files are created for the tabular files [9]. Dataverse converts tabular data files as accurately as possible

[5] ISO 14721:2012 (CCSDS 650.0-M-2) Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model.

[6] Core trustworthy data repositories requirements, https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf.

[7] nestor seal for trustworthy data archives, https://www.langzeitarchivierung.de/Subsites/nestor/EN/Siegel/siegel_node.html.

[8] ISO 16363:2012 (CCSDS 652.0-R-1) Space data and information transfer systems -- Audit and certification of trustworthy digital repositories.

[9] Dataverse: https://dataverse.org/.

with the caveat that some commercial applications like SPSS have not published their specifications [10]. Tabular files also receive UNF checksums that can be used to verify the semantic content of the derivatives [11].

Initiatives in research data preservation, including those using Dataverse, emphasize the necessity of storing and monitoring datasets as independent from the submission and discovery platforms that users usually interact with. This approach appears to be informed by an interpretation of the OAIS reference model, which emphasizes the flow of received data into stored and monitored units of content information as Archival Information Packages (AIPs) and Dissemination Information Packages (DIPs). However, these packages are logical rather than physical: their components may not have to be stored physically together so long as the total package can be retrieved and independently understood by members of the designated community [12]. However, OAIS and related certification frameworks do identify in a broad sense what functions preservation systems should perform, and these features may only partially exist in a software package designed primarily for receipt, access and discovery. Creating platform-independent packages means that preservation institutions can generate and manage preservation metadata, use more than one managed method for storage, run preservation-supporting functions at ingest and over time, and audit and maintain stored packages without depending on a single system to perform all of these tasks in addition to user-facing functions.

Research on research data preservation has emphasized the desirability of storing and managing independent preservation packages. A white paper authored by members of the Canadian Association of Research Libraries (CARL)'s Portage Network Preservation Expert Group theorizes the disaggregation of OAIS-type functions among a set of potential preservation service providers who take care of particular functions such as archival storage, while communicating the results of these efforts back to a centralized administrative agency [13]. In the United Kingdom, Jisc's series of three "Filling in the Preservation Gap" reports specifically investigate the use of Archivematica in service of preserving research data.[10] A series of test implementations at the University of York and University of Hull were deemed successful and Archivematica was among the preservation providers tested with the Jisc's Research Data Shared Service pilot [14]. Therefore, Dataverse's functions primarily map to the "Producer" end of the OAIS model, where materials are negotiated and accepted for ingest, and some baseline preservation-supporting functions are performed. Further research is required on how platforms like Dataverse might fulfill the requirements of the Producer-Archive Interface Methodology Abstract Standard (PAIMAS)[11] and Producer-Archive Interface Specification (PAIS)[12] standards for structuring producer-archive interactions.

Data repositories using Dataverse are taking steps to export data and metadata from Dataverse into preservation storage and/or processing. In the Netherlands, DANS' DataverseNL service exports packages using the BagIt specification[13] to their EASY preservation repository [15]. The Qualitative Data Repository (QDR) at Syracuse University is taking a similar approach with the development of a proof-of-concept implementation of exported OAI-ORE metadata and zipped Bags from Dataverse [16]. The Odum Institute at the University of North Carolina uses scripts to push data packages to iRODs, which performs preservation processing and storage replication [17]. The Dataverse software itself also includes the ability to transfer Bagged exports to DuraCloud [18].

The Dataverse-Archivematica integration takes advantage of the preservation-related actions that Dataverse performs, and makes them available to an Archivematica-based workflow. The features of this integration are discussed in the following sections.

III. History and Scope of Project

A. *Proof-of-Concept*

[10] *Filling the preservation gap* project page: https://www.york.ac.uk/borthwick/projects/archivematica.
[11] Consultative Committee for Space Data Systems, *Producer-archive interface methodology abstract standard*. CCSDS 651.0-M-1. Magenta book, 2004. https://public.ccsds.org/Pubs/651x0m1.pdf.
[12] Consultative Committee for Space Data Systems, *Producer-archive interface specification*. CCSDS 651.1-B-1. Blue book, 2014. https://public.ccsds.org/pubs/651x1b1.pdf.
[13] *The BagIt File Packaging Format (V1.0)*, https://tools.ietf.org/html/draft-kunze-bagit-17.

In response to growing community interest, Scholars Portal initiated a research project in 2015 to investigate how research datasets stored in Dataverse could be processed into AIPs using Archivematica. Initial project participants included members from Scholars Portal and the University of Toronto, Artefactual Systems, IQSS Dataverse, the University of British Columbia, the University of Alberta, Simon Fraser University, and the CARL Portage Network.

Project participants conducted an initial requirements analysis and proposed a draft workflow. Artefactual Systems developed a prototype of the Archivematica software that used Dataverse APIs to retrieve datasets for ingest and processing in Archivematica. The proof-of-concept integration was only available through a development branch of Archivematica and presumed an automated workflow in which all datasets in a target Dataverse would be transferred and processed by Archivematica.

The initial project provided an opportunity to explore best practices related to preservation of research data; investigate how Dataverse handles and stores data and metadata, processes derivatives and versions files, exports data and metadata; and the possibilities for how Archivematica could transfer and interpret Dataverse dataset packages. The project also identified the use of the Data Documentation Initiative (DDI)[14] metadata standard within Archivematica METS files for descriptive metadata. Given DDI's capacity to comprehensively describe specific characteristics related to research data for discovery and reuse, this mapping was intended to expand the scope of descriptive metadata in Archivematica METS files and make these files more hospitable to describing research data.

B. Production Release

In 2018, Scholars Portal sponsored further development work with Artefactual Systems to improve the original proof-of-concept and merge it with the public release of Archivematica in version 1.8 (developed and tested using Dataverse version 4.8.6 and above).[15] The major result of the integration is that Archivematica can be configured to use a connected Dataverse instance as a transfer source location. Datasets are queried and retrieved using Dataverse's APIs and processed using the Dataverse transfer type, which contains specific processing micro-services (described in section IV below).

The integration was designed with a series of assumptions in terms of its design. First, the design presumes a user who has an account with a Dataverse instance and who has generated an API token (a unique code for authentication). The same or a different authorized user also has access to an Archivematica instance and wishes to process certain datasets into AIPs for long-term preservation. This assumes the user has obtained the necessary rights and privileges to process and store dataset files independently from Dataverse. Secondly, the current design assumes that the user is interested in selecting specific datasets in a target Dataverse instance for preservation. This assumption conforms to specifications such as CoreTrustSeal that state that repositories must appraise and select data for preservation.[16] The current design does not include an automated function for ingest of all datasets within a Dataverse container, though we acknowledge that this functionality may meet additional use cases.

A single dataset in a Dataverse instance corresponds to a Submission Information Package (SIP). Individual files cannot be transferred from Dataverse for preservation. However, users can select individual files by using the Appraisal function of Archivematica to create a final AIP.[17] At present, only the current version of files and metadata can be selected for preservation. (Dataverse tracks versioning and provenance of metadata and file changes, with all versions retained by the system). Finally, while users may choose to create a DIP as part of the Archivematica workflow, it is assumed that the version on Dataverse will generally remain the one used for access. The scope of the integration did not include communication back with a connected Dataverse to write preservation

[14] DDI: https://www.ddialliance.org/.

[15] Archivematica 1.8 - Dataverse transfers: https://www.archivematica.org/en/docs/archivematica-1.8/user-manual/transfer/dataverse/#dataverse-transfers.

[16] See s. 8: "Appraisal," in *Core trustworthy data repositories requirements,* https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf.

[17] Archivematica 1.8 - Appraisal: https://www.archivematica.org/en/docs/archivematica-1.8/user-manual/appraisal/appraisal/#appraisal.

metadata, or the replacement of user-submitted data with the DIP generated by Archivematica.[18]

IV. Workflow and Functionality

Fig. 1 presents an overview of the workflow for the integration. Beforehand, a user must first configure the Archivematica Storage Service to connect to a Dataverse instance. Then, Transfer Source Locations can be set to filter based on query search terms or on a specific dataverse container using Dataverse's Search API.

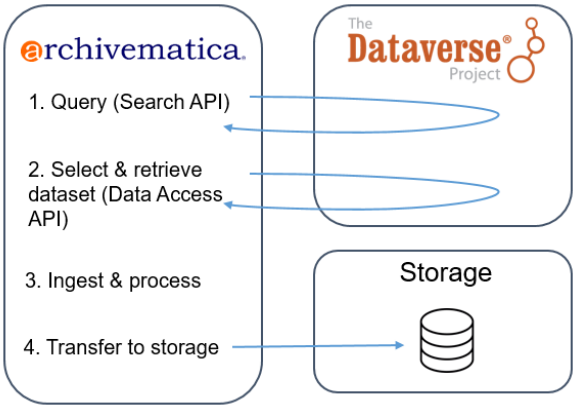

Fig. 1. Workflow for integration between Dataverse and Archivematica

First, the user would browse the datasets available in the Transfer Browser, select one dataset, set the transfer type to "Dataverse," and start the transfer. Second, Archivematica uses Dataverse's Data Access API to retrieve a copy of the most recent version of the dataset. The package contains the original user-submitted data files, and if the user submitted tabular data, a set of derivatives of the original tabular files in several formats along with associated metadata files describing the tabular files. Third, the Dataverse transfer type completes preconfigured ingest and processing steps:

- Creation of a Dataverse METS file describing the dataset as received from Dataverse, which includes a descriptive metadata section mapped to the DDI standard, a list of files grouped by type (original, metadata, or derivative), and a description of the structure of the files provided by Dataverse;
- Fixity checks of data files verified using the MD5 checksums that were generated by Dataverse;
- Other standard Archivematica microservices conducted as configured.

Archivematica produces a METS file for the AIP (see Table 1) that copies over the descriptive metadata from the Dataverse METS file, outlines the relationships between original and any derivative files that have gone through the tabular ingest process in Dataverse, and includes details of the Dataverse instance in the PREMIS agent fields, in addition to the actions undertaken through Archivematica's processing steps. Tabular derivatives are recorded with an associated PREMIS event labeled as "derivation." Though Artefactual Systems proposed this term in 2015 as part of feedback on the PREMIS controlled vocabulary for events, it has not yet been implemented in the PREMIS events controlled vocabulary.[19] Derivatives and metadata files are also identified in the METS fileGrp sections within the fileSec section.

Finally, the resulting AIP is transferred via the Archivematica Storage Service to a connected storage system.[20] Therefore, the stewarding institution or organization would manage policies and costs for long-term storage and maintenance. For example, OCUL institutions that subscribe to Permafrost would have access to Archivematica instances that could be configured to their institutional containers as part of Scholars Portal Dataverse platform. In this case, datasets processed as AIPs could be stored on the OLRC managed by the library. Other users may host Archivematica locally or take advantage of other service arrangements and still be able to connect to a target Dataverse instance of their choice. The integration also presents opportunities for centralized, collaborative services that offer Dataverse, Archivematica, and preservation storage as a service model.

[18] The latter is the case for the Archidora integration between Archivematica and Islandora. See T. Hutchinson, "Archidora: Integrating Archivematica and Islandora," *Code4Lib Journal* 39, https://journal.code4lib.org/articles/13150.

[19] As Evelyn McLellan writes, "The use case is a research data publishing platform that generates tabular file format derivatives from uploaded statistical files. This is not normalization because the purpose is not preservation but rather derivation for the purpose of data manipulation and visualization." See: http://premisimplementers.pbworks.com/w/page/102413902/Preservation%20Events%20Controlled%20Vocabulary

[20] Storage service, *Archivematica Wiki*: https://wiki.archivematica.org/Storage_Service.

Table 1. METS overview

METS section	Description
METS dmdSec	Descriptive metadata section
DDI XML metadata: ddi:title ddi:IDno ddi:authEnty ddi:distrbtr ddi:version ddi:restrctn	DDI fields include: title, unique identifier (e.g., DOI), author(s), distributor (i.e., the Dataverse instance), dataset version, and licenses/restrictions
METS amdSec	Administrative metadata section (for original, derivative, and normalized files)
-techMD	Technical metadata (PREMIS), including file format information and extracted metadata
-digiprovMD	Provenance metadata, including PREMIS events for derivation (for tabular derived files), ingestion, unpacking bundled files, virus check, fixity check, normalization, and PREMIS agents of organization, software, and Archivematica user
METS fileSec	File section defining original files uploaded to Dataverse, derivative tabular files generated by Dataverse, metadata files generated by Dataverse, submission documentation, metadata files and normalized preservation copies generated during Archivematica processing
METS structMap	Structural map, showing directory structure of the contents of the AIP

V. Testing and Future Development

During the development work, test datasets were created in the Scholars Portal Demo Dataverse[21] that were representative of the types of datasets deposited in the production platform. Testing revealed a number of issues affecting transfers and processing, and allowed the research team to identify fixes and enhancements in several areas that could be incorporated in future releases.

A. Metadata

Currently, only six DDI fields (title, unique identifier, author(s), distributor, dataset version, and restriction) are included in the METS descriptive metadata section (see Table 1 above). Additional DDI fields (abstract, subject, and distDate) were proposed given that these fields are mandatory in Dataverse, but considered by Artefactual to be outside the scope of the development contract. All other descriptive metadata is stored in a JSON file stored within the AIP. Rights-related metadata could also be mapped directly to PREMIS as is supported currently in other Archivematica workflows.

B. Interface

Several improvements to the transfer browser pane were identified that would facilitate the ability to query and select appropriate datasets within Dataverse, such as showing the dataset version number and the ability to refine searches within the interface. The team also outlined the need for better logging of errors and notification to users. Issues experienced during testing included errors that had to be resolved within Dataverse, such as missing checksums for files and failed tabular data ingest.[22] A suggested improvement was an additional micro-service for verifying a Dataverse transfer before the transfer begins to make it easier to identify these errors and ensure compliance for Dataverse transfer types.

C. Conformance with Additional Archivematica Functions

AIP re-ingest functions present in Archivematica do not currently function for Dataverse AIPs. Development of this feature requires further discussion about use cases and approaches, such as whether re-ingest should take into account any updates made to the submitted dataset in Dataverse. The team also noted the potential benefit of relating datasets as part of a larger collection through defining an Archival Information Collection (AIC),[23] a function that needs further development to conform with the Archivematica workflow for creating AICs.

[21] Scholars Portal Demo Dataverse: https://demodv.scholarsportal.info/.

[22] Dataverse did not implement checksums in versions 3.6 and prior. For a list of known issues experienced during testing, see https://wiki.archivematica.org/Dataverse.

[23] AIC: https://wiki.archivematica.org/AIC.

D. *Messaging with Dataverse and DIPs*

Once a dataset has been processed and stored, it would be beneficial for Archivematica to send a notification to the Dataverse platform and surface selected preservation metadata indicating to users that the dataset is being preserved. However, this communication mechanism would require development work on both platforms. As mentioned previously in section III.B above, a larger potential development would be the automated replacement of user-submitted data with Archivematica-created DIPs, particularly when normalized access copies of files submitted by users might be desired for ease of access.

E. *Conformance with External Requirements*

As methods for standardization continue to develop in the field, an additional development opportunity is the ability for Archivematica-created AIPs and DIPs in Bags to be conformant to the RDA Research Data Repository Interoperability Working Group's *Final Recommendations* document. The *Recommendations* specify how repository outputs should be structured to promote data exchange, which could be used for redundant storage or access purposes [19].

VI. Conclusion and Discussion

Currently, Scholars Portal is hosting a public Archivematica sandbox connected to its demo Dataverse installation with several test datasets.[24] Invitations to participate in testing the sandbox and to provide feedback were shared with regional, national and international groups related to research data management, digital preservation, and archives, as well as Dataverse users and Archivematica users. Community testing is crucial to provide further information about how different users might use the integration and to identify additional needs of the community. This feedback will be used to inform future platform enhancements and contribute to the ongoing discussion surrounding best practices for preserving research data. We hope that others interested in using these tools will bring additional use cases and sponsor additional developments to improve the integration. Additionally, community members who test and implement the integration on their own infrastructure will also provide new perspectives related to its capacity and limitations in various contexts.

This research and integration work contributes to ongoing research and discussions surrounding research data preservation. Several challenges exist in this area, particularly in relation to forming research data preservation policies and strategies. A recent Jisc report *What to Keep* outlined use cases for research data retention and considerations for this emerging field, noting that the practice and procedures—the what, why, how long, and where—are still evolving [20]. Another challenge in developing policies and strategies relates to the heterogeneity of research data, resulting in a large number of data types and file formats, as well as discipline-specific practices and protocols. The *Science Europe Guidance Document: Presenting a Framework for Discipline-specific Research Data Management* provides a useful guidance framework for protocols within various research domains, informed by the FAIR principles, applicable laws, regulations, and standards [21]. The significant differences across disciplines suggest inherent difficulties in developing general policies and strategies for multi-disciplinary data repositories. Increasing our shared knowledge of various curation and preservation workflows would help to ensure that the tools and policies developed in these areas assist in properly managing different types of data for the long term.

Finally, additional research and requirements-gathering needs to be conducted in the area of service models and policy development to understand how preservation approaches can flow from individual researchers to institutions and repositories that are tasked with stewarding research data, and onto potential to shared infrastructures. Overall, the Dataverse-Archivematica integration project aims to connect pieces of the research data management ecosystem, drawing on best practices and standards in the archives and digital preservation communities, and to contribute to the development and enhancement of features within these two platforms.

Acknowledgements

The authors would like to acknowledge funding from OCUL to pursue this project and thank the Scholars Portal staff, members of Artefactual

[24] Archivematica Demo Sandbox, *Spotdocs Wiki:* https://spotdocs.scholarsportal.info/display/DAT/Archivematica+Demo+Sandbox.

Systems, the initial project participants, and the community for their support of the project. Our thanks go to: Allan Bell (UBC), Eugene Barsky (UBC), Peter Binkley (University of Alberta), Eleni Castro (Boston University, formerly IQSS), Alan Darnell (Scholars Portal), Kate Davis (Scholars Portal), Philip Durbin (IQSS), Alex Garnett (SFU), Geoff Harder (University of Alberta), Chuck Humphrey (Retired, formerly CARL Portage), Larry Laliberte (University of Alberta), Amber Leahey (Scholars Portal), Steve Marks (University of Toronto), Evelyn McLellan (Artefactual Systems), Umar Qasim (University of Alberta), Joel Simpson (Artefactual Systems), Justin Simpson (Artefactual Systems), Ross Spencer (Artefactual Systems), Amaz Taufique (Scholars Portal), Leanne Trimble (University of Toronto), and Dawas Zaidi (Scholars Portal).

REFERENCES

[1] CASRAI, *Research data*, Aug. 12, 2015. Accessed on: Mar. 14, 2019. [Online]. Available: https://dictionary.casrai.org/Research_data

[2] GO Fair, *FAIR Principles*. Accessed on Mar. 14, 2019. [Online]. Available: https://www.go-fair.org/fair-principles/

[3] M.D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data* 3, Mar. 15, 2016. doi:10.1038/sdata.2016.18

[4] Digital Curation Centre, *DCC curation lifecycle model*. Accessed on: Mar 14, 2019. [Online]. Available: http://www.dcc.ac.uk/resources/curation-lifecycle-model

[5] M. Crosas, "The Dataverse Network: An open-source application for sharing, discovering and preserving data," *D-Lib Magazine*, vol. 17, no. 1-2, January-February 2011. doi:10.1045/january2011-crosas

[6] Dataverse, *Dataverse repositories: A world view*. Accessed: Mar. 17, 2019. [Online]. Available: https://services.dataverse.harvard.edu/miniverse/map/

[7] T. H. Vines et al., "The availability of research data declines rapidly with article age," *Current Biology* vol. 24, no. 1. doi:10.1016/j.cub.2013.11.014

[8] Dataverse, *Dataverse installation guide: Installation*, Feb. 19, 2019. Accessed on Mar. 18, 2019. [Online]. Available: http://guides.dataverse.org/en/latest/installation/installation-main.html

[9] Dataverse, *Dataverse user guide: Tabular data ingest*, Feb. 19, 2019. Accessed on Mar. 17, 2019 [Online]. Available: http://guides.dataverse.org/en/latest/user/tabulardataingest/index.html

[10] Dataverse, *Dataverse user guide: Tabular data ingest - SPSS*, Feb. 19, 2019. Accessed on Mar. 17, 2019. [Online]. Available: http://guides.dataverse.org/en/latest/user/tabulardataingest/spss.html

[11] Dataverse, *Dataverse developer guide: Universal Numerical Fingerprint (UNF)*, Feb. 19, 2019. Accessed on Mar. 17, 2019. [Online]. Available: http://guides.dataverse.org/en/latest/developers/unf/index.html

[12] B. Lavoie, *The Open Archival Information System (OAIS) reference model: Introductory guide (2nd edition)*. Great Britain: Digital Preservation Coalition, 2014. doi: 10.7207/twr14-02

[13] U. Qasim, C. Davis, A. Garnett, S. Marks, and M. Moosberger, *Research data preservation in Canada: A white paper*. Portage Network, 2018. doi: 10.14288/1.0371946

[14] J. Mitcham, C. Awre, J. Allinson, R. Green, and S. Wilson, *Filling the digital preservation gap: A Jisc research data spring project. Phase Three report*. Jisc, 2016. doi: 10.6084/m9.figshare.4040787

[15] V. Tykhonov, P. Doorn, and M. Wittenberg, *The development of DataverseNL data repository as structured data hub*, 2017. Accessed on: Mar 17, 2019. [Online]. Available: https://www.slideshare.net/vty/dataversenl-as-structured-data-hub

[16] James Myers, *Data and metadata packaging for archiving*, Jan. 10, 2019. Accessed on: Mar. 17, 2019. [Online]. Available: https://github.com/QualitativeDataRepository/dataverse/wiki/Data-and-Metadata-Packaging-for-Archiving

[17] J. Crabtree, R. Moore, D. Sizemore, *Odum Institute iRODS policies to support preservation*, 2016. Accessed on: Mar. 17, 2019. [Online]. Available: https://slideplayer.com/slide/14487733/

[18] Dataverse, *Dataverse installation guide: Configuration - DuraCloud/Chronopolis integration*, Feb. 19, 2019. Accessed on Mar. 17, 2019. [Online]. Available: http://guides.dataverse.org/en/4.11/installation/config.html#id65

[19] RDA Research Data Repository Interoperability Working Group (WG), *Final Recommendations*. Accessed on: Mar. 19, 2019. [Online]. Available: https://www.rd-alliance.org/system/files/Research%20Data%20Repository%20Interoperability%20WG%20-%20Final%20Recommendations_reviewed_0.pdf

[20] Neil Beagrie, *What to keep: A Jisc research data study*, Feb. 2019. Accessed on: March 18, 2019. [Online]. Available: https://repository.jisc.ac.uk/7262/1/JR0100_WHAT_RESEARCH_DATA_TO_KEEP_FEB2019_v5_WEB.pdf

[21] Science Europe, *Science Europe guidance document: Presenting a framework for discipline-specific research data management*, Jan. 2018. Accessed on: March 19, 2019. [Online]. Available: https://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf