

**2011 National Leadership Grant for Libraries
Advancing Digital Resources Category
Cornell University and Washington University in St. Louis**

**Enabling Data Sharing and Discovery
with the
DataStaR Semantic Platform**

Narrative

Assessment of need

The promise and challenge of sharing, reusing, and preserving research data have recently attracted a great deal of attention (Westra, Ramirez, Parham, & Scaramozzino, 2010). Widespread sharing of data has the potential to facilitate new ways of conducting research and to enable new discoveries. In addition, research institutions need to consider how they will support their researchers in complying with the new or strengthened requirements for data sharing from funders such as the National Science Foundation (NSF, 2010) and the National Institutes of Health (NIH, 2007). Research libraries, in their roles as custodians of the intellectual output at their institutions as well as potential “last mile” service providers (Gabridge, 2009), are well-positioned to provide such support, particularly for small-scale datasets.

Significant challenges impede this support, primarily because the infrastructure for services and best practices for research data curation and discovery are still in development. Some active areas of research include efforts such as the NSF’s DataNet projects¹ and the UK’s Digital Curation Centre². Libraries are faced with a difficult task: provide robust support for the breadth of data needs of individual researchers and their institutions as a whole, while the means for actually doing so are still under development.

The full scope of tasks involved in data curation is articulated concisely on the home page of the Digital Curation Education Program (DCEP) of University of Illinois Urbana/Champaign³:

“Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarly and educational activities across the sciences, social sciences, and the humanities. Data curation activities enable data discovery and retrieval, maintain data quality, add value, and provide for reuse over time. This new field includes representation, archiving, authentication, management, preservation, retrieval, and use.”

The DataStaR (Data Staging Repository) software serves as one component in a suite of curation services under development by the Cornell University Library to address the data management needs of faculty and other researchers. DataStaR was originally developed with support from the National Science Foundation (NSF award III-0712989), and was conceived of as a local data staging repository to facilitate the documentation and transmission of research datasets from a variety of disciplines to domain-specific repositories and/or institutional repositories (Dietrich, 2010; Khan, et.al., 2010; Lowe, 2009; Steinhart, 2010; Steinhart, Dietrich, & Green, 2009). During the NSF grant period, librarians have worked with researchers to provide data curation services early in the research cycle, and then promoted the transmission of data to repositories better suited for long-term curation and preservation. Development of DataStaR’s metadata capabilities proceeded in tandem and was informed by the needs of researchers working with librarians.

We now propose to take DataStaR in important new directions, enabling research data discovery and data sharing by collecting and maintaining metadata about research data created at an institution and providing the option to link these detailed dataset descriptions to research outputs, research activities, and personal profiles of researchers within or outside that institution. DataStaR will be documented and packaged for public distribution via an online open-source community environment. The online community will be used to foster support for the ongoing involvement of libraries and other research institutions in creating semantically

¹ http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141

² <http://www.dcc.ac.uk/>

³ <http://cirss.lis.illinois.edu/CollMeta/dcep.html>

descriptive metadata for discovering and sharing datasets across institutions and over time. By focusing on metadata rather than permanent data repository functions, an expanding network of DataStaR institutions can create a rich resource for data discovery and sharing while avoiding the costs and complexities of a large, centralized data store.

DataStaR encourages metadata creation about datasets and makes them discoverable at the individual dataset level, not just at the level of entire collections. While individual faculty members may refer to published datasets as research outcomes, departmental or university profile pages rarely provide any way to list even cursory information about datasets. And while institutional repositories that accept data in addition to text publications may be able to store metadata files along with research datasets, it would be rare for the metadata files to be indexed for discovery or comparison purposes (e.g., by geographic, temporal, or thematic coverage).

This proposed project has been inspired in part by the Australian National Data Service⁴ (ANDS) model for institutional participation in a broad national data registry infrastructure, with a few important differences. Since to our knowledge no national, cross-disciplinary registry model yet exists in the United States, the DataStaR approach relies primarily on Linked Data as the discovery and aggregation method from distributed local staging repositories rather than harvesting metadata to a central metadata repository. The Linked Data approach is a rapidly expanding mode of information sharing and exchange promoted by the World Wide Web Consortium (W3C) and other organizations that use the standard Web protocol (HTTP) to allow Web pages to be exposed as structured data and not just human-readable text. By exposing the full metadata from publicly sharable datasets as Resource Description Framework (RDF) statements, those metadata are available for harvesting and indexing as part of any research data registry or discovery service, yet they remain fully in the hands of the local institution and its participating researchers for updates and additions over time, including usage information, links to publications, additional data, and follow-on projects.

Another critical aspect of need is providing motivation for researchers to document their datasets. A great deal of effort is being directed at understanding the needs of researchers with respect to data curation (e.g. Data Curation Profiles project⁵), as well as information management in general (Kroll & Forsman, 2010). A library service model designed to support researchers in making their data more visible and reusable must be developed in the context of researchers' own practices and priorities, and be applicable throughout the research process. This project will address conceptual clarity, ease of use, and the value proposition for researchers of documenting and sharing their data via DataStaR through case studies and extensive user testing.

National Impact and Intended Results

Research libraries are faced with the daunting task of supporting data services without well-established tools or guidelines. This project will advance the capabilities of research libraries by providing a tool for data documentation, data sharing and ongoing discovery in the context of broader case studies assessing ongoing needs and informed by iterative user testing to validate and improve our approach. While not addressing all data-related service needs, a locally implemented DataStaR will help to make research data services and the research data itself more visible in the local institutional landscape as well as enable that information to be linked or aggregated for indexing and discovery functions across multiple institutions or by domain of interest.

DataStaR will be extended to interoperate with VIVO⁶, a semantic researcher networking platform also developed at Cornell using closely related technology. It will also support interoperation with other linked-data capable applications by offering data for harvest or direct linking that has been defined using an open and

⁴ <http://ands.org.edu/funded/eif-fast-start.html>

⁵ <http://datacurationprofiles.org>

⁶ <http://vivoweb.org>

documented ontology. As part of this work, DataStaR will be capable of augmenting its own information with a selected set of data elements from the VIVO researcher networking platform or any data source compatible with VIVO's ontology. Any other application capable of issuing a linked data request to DataStaR can in turn opt to include information on research datasets documented in DataStaR, associated via a common personal identifier or a common publication Digital Object Identifier (DOI)⁷. Making VIVO and DataStaR interoperate as complementary platforms at the institutional level provides additional flexibility for augmenting researcher profiles with research, datasets and creating coordinated data feeds to other websites, inside or outside the university.

We believe the long-term impact of dynamic linked-data connections between distributed, locally maintained repositories of information about researchers, research datasets, and the datasets themselves has not yet been fully recognized. While many challenges remain in establishing and maintaining common identifiers (or efficient techniques to resolve multiple identifiers for the same individual), the ability to query and assemble semantically identified information from any available source without specialized protocols and format conversions will substantially accelerate our ability to discover and share data as well as help raise awareness among researchers of the potential value for them in participating more fully in Web-enabled collaboration.

The project will demonstrate how research libraries can implement a staging repository to support data documentation, discovery and reuse in their institutions. By positioning DataStaR as an open-source project in a documented community development environment, we will provide other institutions a platform for developing services pursuant to the needs of their own researchers and the service support capabilities of the libraries or other units potentially involved.

As part of our ongoing effort to match development efforts to researchers' needs, we plan to develop a set of Data Curation Profiles representing a cross-section of disciplines and addressing researchers' needs and preferences with respect to the documentation, sharing, dissemination and reuse of their datasets. By adopting the standardized Data Curation Profile format for case study interviews and analysis (and by seeking IRB approval for the study), we will be able to add our profiles to the growing collection of completed DCPs, which is becoming a valuable resource for curators seeking to understand how to best meet the needs of researchers in this area.

Project Design and Evaluation Plan

Project Design

The technical architecture for a local data staging repository has been implemented as a working prototype at Cornell using a unique metadata management architecture supporting heterogeneous data and metadata in a flexible way while leveraging appropriate discipline-specific metadata schemas for the disciplines locally represented. The participating researcher can create preliminary metadata for research datasets; upload preliminary datasets to the staging repository maintained by the project; share preliminary data publicly, or only with selected colleagues; complete a more detailed metadata record using a form-based editor; reuse elements of existing metadata records in the creation of new metadata records; and optionally export "publication ready" metadata and data to the permanent institutional repository or (depending on required metadata and data formats) to an external repository. Librarians offer assistance with any of these processes through general curatorial expertise enhanced by domain-specific knowledge in one or more disciplines.

⁷ <http://www.doi.org/>

DataStaR Development

We are requesting funds to transition DataStaR from a single-library software prototype to a well-documented open-source platform ready for adoption and extension at other institutions wishing to provide research data staging repository, sharing, and discovery services.

DataStaR extends the VITRO⁸ software which underlies VIVO⁹ and which “combines a Web-based ontology and instance editor with a public display interface” (Lowe, 2009). DataStaR also allows users to optionally upload datasets for online storage and define access and modification permissions for different individuals and research groups. Files uploaded to a dataset are stored in the Flexible Extensible Digital Object Repository Architecture¹⁰ or FEDORA repository.

DataStaR’s ontology incorporates classes and properties from several established ontologies that collectively define and specify the relationships between datasets, the individuals who create or contribute to them, and affiliated organizations. Additional metadata elements such as spatial locations, topics, protocols, or publications may be added in accordance with disciplinary standards; a dataset’s metadata input forms are generated based on the associated ontologies and the choice of destination repository. Metadata are stored as Resource Description Framework (RDF) triples until ready for export or publication. Current efforts will result in DataStaR support for download of metadata and datasets as the requisite publication path to specified target repositories via the Simple Web-service Offering Repository Deposit (SWORD) protocol (Allinson, Francois, & Lewis 2008).

The National Institutes of Health-sponsored *VIVO: Enabling National Networking of Scientists* project, continuing through August 2011, has enabled significant enhancements to the VITRO software also supporting DataStaR. Changes to the common code base will be propagated to DataStaR for greater interoperability during the spring and summer of 2011. A consortium of Australian universities working under sponsorship from the Australian National Data Service¹¹ (ANDS) has extended the VITRO application and the VIVO ontology to create a “turnkey research data registry solution” supporting conversion of Open Archives Initiative (OAI) harvesting of research dataset metadata compliant with the RIF-CS metadata standard¹². Code for converting the RDF of VITRO to RIF-CS and for providing an OAI-PMH harvest point from VITRO and the ANDS VITRO¹³ is already available as sources for enhancements to DataStaR; the ANDS VITRO metadata store solution is currently oriented to collection-level descriptions of data, and not the more fine-grained descriptions of individual datasets supported by DataStaR.

Specific development tasks for the proposed project include:

- Updating the DataStaR ontology by phasing out remaining elements of the SWRC¹⁴ ontology in favor of the VIVO ontology and incorporation of elements of the ANDS VITRO ontology to support RIF-CS and other extensions.
- Developing an interface for specifying a new destination repository and indicating both required and optional metadata elements (or standards) accepted.

⁸ <http://vitro.mannlib.cornell.edu>

⁹ <http://vivoweb.org>

¹⁰ <http://http://fedora-commons.org>

¹¹ <http://ands.org.au/guides/metadata-stores-solutions.html>

¹² <http://ands.org.au/resource/rif-cs.html>

¹³ <http://eresearch.griffith.edu.au/ANDS/vitro/ANDS-VITRO.owl>

¹⁴ <http://ontoware.org/swrc/>

- Providing an interface for researchers to reference a full VIVO profile at their institution from DataStaR and link from VIVO to research data links and descriptions in DataStaR.
- Improving application interfaces including the researcher “dashboard” providing an overview of all datasets and their viewing and publication status, clearer distinctions between dataset-specific and reusable metadata, and better support for editing multiple properties via a single form.
- Improving search filtering and results faceting to enable viewing datasets and related activities by geography, topic, organization, destination repository, and other factors.

All of these development tasks will be informed and augmented with additional tasks through iterative user testing as described below.

Research Dataset Metadata

We have consciously avoided defining a single, fixed set of metadata elements for DataStaR in the belief that reusing classes and properties derived from existing domain-specific XML standards such as the Ecological Metadata Language or EML¹⁵, Darwin Core¹⁶, Federal Geographic Data Committee’s Content Standard for Digital Geospatial Metadata (FGDC)¹⁷ will facilitate alignment of metadata from associated disciplines. For this project, we will supplement the DataStaR ontology with classes and properties representing broader initiatives for research data description including RIF-CS registry interchange format¹⁸ for research datasets, the Core Scientific Metadata Model (CSMD) (Matthews & Sufi, 2004), the Common European Research Information Format¹⁹, the “isCitedBy” metadata scheme for DataCite²⁰ and additional ontologies such as those being developed by the Semantic Publishing and Referencing Ontologies project²¹. Where practical, we will maintain original class and property names and namespaces to simplify mapping metadata from DataStaR to other standards and to promote direct interoperability with other sources of dataset metadata.

Evaluation Plan

For a robust understanding and evaluation of the needs and uses of a semantic platform for data discovery, three activities should occur. As these are each unique, evaluation will be tailored for each area.

1) User Testing: Understand and document the potential uses of a semantic platform (DataStaR)

As a service model, DataStaR is predicated on providing a low-barrier interface that researchers will be motivated to use themselves, allowing library staff to act in the role of trainers and facilitators rather than providing all the labor required for specialized data curation services. The user experience is not just about functionality or performance; it is also about the interface providing a positive or affective experience for the user. Successful software development is based on user analysis, usability principles, information architecture, and aesthetic visual design and branding. By performing iterative user testing in the design process, the end product moves closer to representing user needs.

Direct user testing will be performed on the DataStaR software to lower barriers and improve the effectiveness of the software for researchers in terms of overall conceptual organization and visual design, the flow of data entry and review in the application, and items as detailed as the wording of instructions and definitions of data

¹⁵ <http://knb.ecoinformatics.org/software/eml/>

¹⁶ <http://rs.tdwg.org/dwc/index.htm>

¹⁷ <http://www.fgdc.gov>

¹⁸ <http://ands.org.au/guides/rif-cs-awareness.html>

¹⁹ <http://www.eurocris.org/Index.php?page=CERIF2008&t=1>

²⁰ <http://www.dlib.org/dlib/january11/starr/01starr.html>

²¹ <http://bit.ly/9d8qAi/>

elements. Testing with human subjects will be conducted throughout the project at both partner locations to provide iterative feedback on DataStaR’s interface and supported metadata elements, using multiple rounds of testing with a relatively small number of users to accelerate necessary improvements.

User testing will also test the extent that the researchers can use DataStaR independently without librarians mediating or providing metadata description and entry. The level of support required by users will have a significant multiplier effect for libraries in providing DataStaR, and will in turn affect sustainability.

Evaluation Approach: A descriptive, demonstration evaluation will be used to describe the value of DataStaR for the identified stakeholders using a limited number of participants. User tests will be conducted using software (e.g. Morae usability software²²) to capture the screen movements as well as audio and visual actions of the participants. Quantitative summaries will be created to elucidate which aspects of the software are being used, how they are used, and the time spent on each task. As this is an exploratory study, we are not seeking statistical power. However, we will attempt to recruit 4-6 participants at each institution.

2) Case Studies: Identify gaps in data discovery from the perspective of librarians and researchers

We propose to conduct interviews with researchers at both Cornell University and Washington University in St. Louis and apply the Data Curation Profiles (DCP) toolkit²³ to capture information about a particular dataset itself (life cycle, purpose, perceived value) as well as information about the researcher’s needs and preferences with respect to that dataset (whether, how, and when it will be shared, needed documentation and description, preservation requirements).

To meet the specific needs of our proposed project, we plan to extend Section 5 of the DCP interview template (Organization and Description of Data) to include additional questions aimed at soliciting information from researchers regarding what aspects of metadata they deem essential to supporting discovery, access and reuse of their data. We will develop these supplementary questions in consultation with the Metadata Services unit at CUL. These findings will inform additional requirements for the DataStaR platform and its metadata models, as well as providing a body of summary information documenting the scope of need and assessing how DataStaR fits into the larger data curation services landscape for potential adopting institutions.

Evaluation Approach: A subjectivist evaluation is appropriate for this area, as we will determine and document the potential need for a semantic tool for data discovery and the potential niche in which it will most likely succeed. Case studies will be used to elucidate this information. Through working with the participants in their natural environments, we will be able to preserve ecological validity, yet qualitative objectivity can be maintained through using informed, experienced, unbiased observers. There will be a combined total of 8-10 participants recruited between the two institutions.

Anticipated Evaluation Workflow

The workflow for items 1 and 2 above are guided by the natural history of a subjectivist study (Friedman & Wyatt, 2006) and described in Table 1, Anticipated Evaluation Workflow in the supporting documents.

3) Assess the project objectives

Descriptive statistics and reporting will be used for reporting the goals of the grant. The project will be considered a success if we achieve the following:

²² <http://www.techsmith.com/>

²³ <http://www4.lib.purdue.edu/dcp/>

- 8-10 completed interviews/case studies
- Produce a set of requirements of a data staging system based on both input from active researchers and curatorial functions necessary to implement a staged approach to data curation
- Complete assessment of the DataStaR system for use in a data staging capacity
- Disseminate our findings as described in the “Dissemination” section, see below

Project Resources: Budget, Personnel, and Management

We propose a single year project to build on the momentum and leverage the skills, technologies and the lessons learned from the first DataStaR project. While we recognize the data staging model inherent in DataStaR does not meet all research data service needs, we believe it to be a constructive and flexible option that can provide the immediate benefit of gathering and sharing of metadata in a sustainable, discoverable way.

The total project budget of \$354,581 reflects the three major project activities that will be coordinated throughout the project to facilitate delivery of mature, documented software, reflecting both our increased understanding of researchers’ needs as well as direct user feedback from testing the software. Librarians and IT staff serve jointly on all project teams in Mann Library and the resulting feedback loops improve communication and lead to better outcomes in the services we provide our patrons.

The project teams at Cornell and Washington University will interact closely through coordination of the case studies and user testing at both universities, continuing a two-year working relationship nurtured on the VIVO project. Project communication and coordination will be supported using the collaboration tools in the SourceForge²⁴ site, which provides the secondary benefit of planting the seed for future and ongoing community development efforts. The primary development work will be conducted at Cornell, where both DataStaR and VIVO originated, building on the considerable application development, semantic data modeling, education, documentation, data curation, and service delivery expertise assembled through those two high-profile projects.

Personnel

Mary Ochs, the principal investigator in the project, has extensive experience leading significant long-term information access and service development projects in international agriculture, including The Essential Electronic Agriculture Library (TEEAL) and the AGORA initiative with the U.N. Food and Agriculture Organization. Mary has also served as principal investigator on a previous successful IMLS-funded project, the Home Economics Archive: Research, Tradition, and History (HEARTH) project²⁵ (IMLS ND-00002). The remaining project team at Cornell includes Gail Steinhart, the principal Investigator of the NSF-sponsored DataStaR project, as lead for the case study interviews, with Dr. Ellen Cramer serving as an investigator for the user testing and overall project operational coordination. Dr. Cramer has a Ph.D. in computing technology in education and a graduate certificate in public health informatics. She has served as senior personnel on the NIH VIVO grant, as well as lead of the Cornell VIVO implementation. The technical team includes Jon Corson-Rikert, the original VIVO developer and overall development lead for the NIH VIVO project; Brian Lowe, lead of the semantic application development team for VIVO and lead semantic developer for the first two years of the first phase DataStaR project; Huda Khan, DataStaR developer since January of 2010; and Brian Caruso, team lead for application development for VIVO and developer for the FEDORA integration and access control aspects of DataStaR.

²⁴ <http://sourceforge.org>

²⁵ <http://hearth.library.cornell.edu>

Dr. Leslie McIntosh will serve as co-principal investigator at Washington University (WU), advising the evaluation components of the project and leading the WU team, including the technical implementation, case study, and usability initiatives running in parallel with activities at Cornell. Dr. McIntosh holds a Masters in Public Health with an emphasis in biostatistics and epidemiology in addition to a Ph.D. in epidemiology. She serves as the national evaluation lead and Washington University implementation lead on the NIH VIVO project as well as facilitating research within the Center for Biomedical Informatics at WU. Sunita Koul, a programmer within the Center for Biomedical Informatics and VIVO implementation expert, will lead the DataStaR implementation and ontology effort at WU. BJ Johnston will identify researchers and facilitate the case studies at WU.

Dissemination

We propose to share our results with the academic library profession via conference presentations and publications and plan to target the repository and semantic communities as well as the more general academic library community. Potential conferences and journals include: American Library Association, Association of College and Research Libraries, International Digital Curation Conference, Open Repositories Conference, Joint Conference on Digital Libraries, American Society of Information Science and Technology, International Journal of Digital Curation, or the Data Science Journal.

The dissemination model for the DataStaR software will combine a public website aimed at adopting libraries and a technical support community on SourceForge²⁶ or similar open-source software site. In addition to software downloads, SourceForge provides collaborative tools that include mailing lists, a wiki space for documentation and user-provided content and examples, and issue-tracking capabilities. The documentation will address the support needs of the researcher as the end user of DataStaR, the librarian as the service provider who engages with the metadata ontologies in DataStaR, and the technical system administrator. System documentation will cover installation and configuration of DataStaR, augmenting existing and extensive documentation of the primary components (MySQL, Apache web server, Tomcat servlet engine, Java and Jena code libraries, Lucene and Solr search tools, and FEDORA). DataStaR will be offered with a Berkeley Software Distribution (BSD) license²⁷, as is the VIVO software. The BSD license allows for the redistribution and use of the software in source or binary form.

Sustainability

The development effort on this grant will focus on preparing the DataStaR software for broad distribution and adoption by libraries and other organizations wishing to provide data curation support services beyond basic information assistance. In addition to the technical support community described above, the DataStaR project website will also provide summary information on the case studies and user testing components of the project geared toward libraries and librarians trying to decide what to offer for data curation services and assessing the needs of their own institutions.

DataStaR's longer-term future as a software application will depend on the success of the online community participants in attracting other developers and librarians engaged in data-related services to contribute to the ongoing evolution of tools and services. However, it is important to note that all information in DataStaR is stored in a simple, documented format adhering to Semantic Web standards and interoperable with both commercial and open-source tools today, and can be migrated forward as tools improve over time.

²⁶ <http://sourceforge.net>

²⁷ <http://www.opensource.org/licenses/bsd-license.php>

References

- Allinson, J., Francois, S., & Lewis, S. (2008). SWORD: Simple Web-service Offering Repository Deposit. Retrieved from <http://www.ariadne.ac.uk/issue54/allinson-et-al/>.
- Dietrich, D. (2010). Metadata management in a Data Staging Repository. *Journal of Library Metadata* 10 (2 & 3).
- Friedgan, A. (2008). Metadata Improvements -- A Case Study. The Data Administration Newsletter. Retrieved from <http://www.tdan.com/view-articles/7353>.
- Friedman, C.P. & Wyatt, J.C. (2006). *Evaluation Methods in Bioinformatics*. New York: Springer.
- Gabridge, T. (August 2009). The Last Mile: Liaison Roles in Curating Science and Engineering Research Data. Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC, no. 265 (August 2009): 15–21. <http://www.arl.org/resources/pubs/rli/archive/rli265.shtml>.
- Khan, H., Caruso, B., Lowe, B., Corson-Rikert, J., Steinhart, G., & Dietrich, D. (2010). DataStaR: Using the Semantic Web approach for Data Curation. *Digital Curation Conference 2010*.
- Kroll, S. & Forsman, R. (2010). A Slice of Research Life: Information Support for Research in the United States. Dublin, Ohio. Report commissioned by OCLC Research in support of the RLG Partnership. Retrieved from: <http://www.oclc.org/research/publications/library/2010/2010-15.pdf>.
- Lowe, B. (2009). DataStaR: Bridging XML and OWL in Science Metadata Management. *Metadata and Semantic Research Third International Conference, MTSR 2009*. Milan, Italy: Berlin: Springer.
- Matthews, B. & Sufi, S. (2004). CCLRC Scientific Metadata Model: Version 2. *Business and Information Technology Department, CCLRC, Rutherford Appleton Laboratory*. Retrieved from: <http://epubs.stfc.ac.uk/bitstream/485/csmdm.version-2.pdf>.
- Matters, M.D., Lekichvili, A., Savel, T. & Zhi-Jie, Z. (2005). *Developing Metadata to Organize Public Health Datasets*. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560704/>.
- Metadata Requirements for Datasets. (2008). *Global Biodiversity Information Facility (GBIF) Network*. Retrieved from http://www2.gbif.org/GBIF-metadata-strategy_v.06.pdf.
- Rogers, N., Huxley, L. (December 2009). Exchanging Research Information in the UK EXRI-UK . *JISC*. University of Bristol, CLAX Ltd.
- Shadbolt, A. & Porter, S. (2010). Creating a university research data registry: enabling compliance, and raising the profile of research data at the University of Melbourne. *31st Annual IATUL Conference of the International Association of Scientific and Technological University Libraries*, Purdue University, June 20-24, 2010. Retrieved from: <http://docs.lib.purdue.edu/iatul2010/conf/day3/3/>.
- Steinhart, G. (2010). DataStaR: a data staging repository to support the sharing and publication of research data. *International Association of Scientific and Technological University Libraries, 31st Annual Conference*. West Lafayette, IN.
- Steinhart, G., Dietrich, D., & Green, A. (2009). Establishing trust in a chain of preservation: The TRAC checklist applied to a data staging repository (DataStaR). *D-Lib Magazine*, 15(9/10).
- Westra, B., Ramirez, M., Parham, S.W. & Scaramozzino, J.M. (2010). Selected Internet Resources on Digital Research Data Curation. *Issues in Science and Technology Librarianship*, 63. Retrieved from: <http://www.istl.org/10-fall/internet2.html>.