LIBER QUARTERLY
The Journal of the Association of European Research Libraries
Vol. 27, no. 1 (2017) 73–88 | e-ISSN: 2213-056X

Data Management: New Tools, New Organization, and New Skills in a French Research Institute

Caroline Martin

Irstea, Antony, France
caroline.martin@agreenium.fr

Colette Cadiou

Irstea, Antony, France
colette.cadiou@irstea.fr

Emmanuelle Jannès-Ober

Irstea, Antony, France
emmanuelle.jannes-ober@irstea.fr

Abstract

In the context of E-science and open access, visibility and impact of scientific results and data have become important aspects for spreading information to users and to the society in general. The objective of this general trend of the economy is to feed the innovation process and create economic value. In our institute, the French National Research Institute of Science and Technology for Environment and Agriculture, Irstea, the department in charge of scientific and technical information, with the help of other professionals (Scientists, IT professionals, ethics advisors…), has recently developed suitable services for the researchers and for their needs concerning the data management in order to answer European recommendations for open data. This situation has demanded to review the different workflows

Uopen Journals | http://liberquarterly.eu/ | DOI: 10.18352/lq.10196

between databases, to question the organizational aspects between skills, occupations, and departments in the institute. In fact, the data management involves all professionals and researchers to assess their ways of working together.

Key Words: data management; datasets; databases; data publication; skills; services; e-science; open data

1. Introduction

Irstea, the National Research Institute of Science and Technology for Environment and Agriculture, is a public research institute under the joint supervision of the Ministry of Research and the Ministry of Agriculture in France. Irstea has built a multidisciplinary and systemic approach to three domains: water, environmental technologies and territories, which today form the basis of its strength and originality. The appropriation of scientific results is a very important mission of the institute. It wants to be a link between practitioners and scientists and represents a collaborative space dedicated to the co-construction of knowledge.

With exponential growth and massive data production in all areas of science, management and recovery of data becomes, in the digital age, a crucial issue in technical, scientific and economic policy. These evolutions affect the production and use of scientific information, and consequently they impact on the practices of information professionals at several levels:

- Technical: for the increasingly rapid evolution of software tools, IT infrastructures and practices;
- Organizational and behavioural: new approaches leading to new modes of production and development of scientific and technical production.

This requires an important adaptation of skills and missions of Scientific and Technical Information (STI) professionals (ALA, 2014). In fact they are expected to serve more closely the different requirements to data processing, reporting and disseminating (Macmillan, 2014). These new missions are quite natural in the continuity of the usual accompanying activities (training

and local support) made by librarians from the scientific teams. (Schmidt & Shearer, 2016)

At Irstea STI professionals have developed support services for research projects, which include an involvement in the processes of management and enhancement of scientific data. After the presentation of the external and internal context of Irstea, we present the needs and questions of researchers around the data life cycle.

The new services proposed at the moment are:

- Guidelines on data management, archiving and diffusion;
- Managing a quality process on data management (ISO 9000 Quality certification);
- Training sessions and seminars in the institute to share this knowledge and to point out new skills and new transversal working methods. In particular, it is a management approach in order to support staff to future changes in the skills of librarians in a research institution by building a training plan, and the implementation of an STI operational organization to answer future challenges in the scientific research.
- Developing services as, e.g., DOI creation, as the development of "IrsteaData" (a catalogue of datasets), an important new part of our data publication system and based on a close collaboration between librarians, scientists, technicians, IT professionals, lawyers…;
- Working on our vocabularies: as we have opened our institutional repository (CemOA), we are working on the mapping of our referentials with major vocabularies for all our research products (mapping of our Irstea thesaurus with Agrovoc and Gemet thesauri, we plan to use Orcid and Geonames also).

These new services have brought the different workflows into question, especially for the deposit process of data publication and the management of the interoperability between the databases of the institute. Establishing the link between data, publications and other scientific or technical productions (reports etc.) related to the same research project has meant adopting another point of view about the management of the information system and involving the whole staff in a transversal work about their skills and roles in the scientific process.

We will see how STI professionals are now able to play a pivotal role in the management of data in relation with the other actors involved. STI services will be presented within their place in the data management systems. Finally, we highlight the leading role of STI professionals in a process of evolution of the interoperability between the various resources of the scientific information system of Irstea.

2. Data Management and Scientific Needs

2.1. Data Background at Irstea: Typology and Institutional Policy

The external environment including open science, open data and H2020, requires researchers to open and document their data.

The data at Irstea are characterized by their heterogeneity at all levels of description, such as:

- their typology: data from laboratory or field measurements, surveys of territorial actors (sociology, economics, agronomy, management science…), photos, videos, UAV images, GIS data, texts
- their field: biodiversity, ecology, natural hazards (avalanches, floods…), robotics, pollution and water chemistry, refrigeration engineering, economy, sociology, etc.;
- their nature: public data (e.g., www.avalanches.fr) confidential data (from companies or for patenting…), data privacy, licensing data (geographical data; weather/climatic data, social enquiries, agricultural and economic statistics (Eurostat, Agreste, INSEE…), reused or produced by researchers;
- The modes of organization of the data management: who does what, how and when and with whom? Lots of questions with many answers that depend on the type of players involved (researchers, data managers, IT experts, STI professionals…).

Irstea has a policy for data management and a quality process framework. Nevertheless, data management at Irstea still remains to be developed in order to become completely operational.

2.2. The Needs and Initiatives of Researchers

Several surveys were conducted with researchers (one about the use of the data papers, e.g.) but lack of responses did not allow for an in-depth analysis of needs. An internal seminar was organised in February 2015 to identify who works with data, and to collect their needs in terms of management, storage, archiving and dissemination of data and datasets. During this seminar, several workshops were held on the data management issues (metadata, storage and archiving conditions, terms of dissemination). The seminar was attended by about sixty people (64 members) from the Information and Communication Technology domain, and researchers covering a wide range of scientific fields (hydrology, biodiversity, remote sensing).

– Significant support needs:

Data life cycle stage	Expressed needs	Impact on organisation, and actors
Data production and processing	– Deployment of tools to describe data and datasets; – Choice and use of standards to document datasets and also what types of format to select. The granularity of the data documentation is a significant question for researchers because it determines the data quality	Upstream data management from the outset of data creation queries research methods and touches on individual and or/collective habits of use. This stage has a strong impact on the following stages
Data preservation	More IT support making available long-term and medium storage; The question of storage space is asked	Data preservation must be handled within a perspective of strategic and cultural changes in organisation to the storage and conservation of data under agreement. The choice of archiving formats is important
Access, dissemination and re-use	The thorny issue of access is inseparable from the issue of security and traceability. Law support on the conditions of access of the data is unanimously required as well as the conditions for re-use of data	How to make data available, in accordance with which strategy in a legal context which is hard to understand for the researcher. The researcher remains very dubious, even suspicious of conditions for the dissemination of his/her data, in an open science context

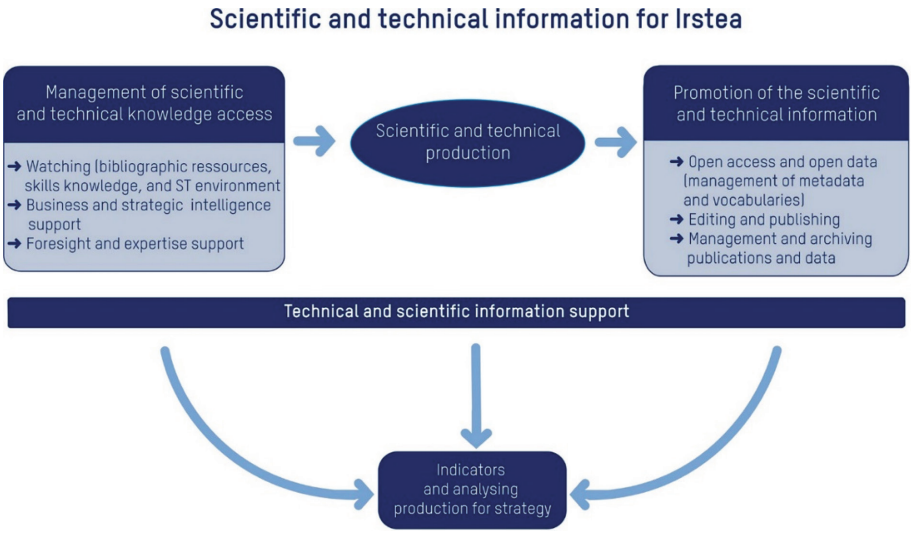
– The Bottom-up initiatives:
  Researchers organise themselves to manage and share their data (JISC, 2012). Moreover, in the last few years, research teams in the humanities have taken the initiative, such as in the case of the "Sygade" database (management and archiving system of interview's data): fruit of joint work of researchers, IT experts and librarians. Other teams have organised themselves to manage, exploit and deploy data in different fields (Hydrology, Biodiversity) and various supports (measurement data, photo collections, etc.). In this context, it should be noted that the support of STI professionals on the choice of standards, formats for archiving and tools, is often appreciated.

3. Scientific and Technical Information Department (STI): A Legitimacy to Build

3.1. STI Position on New Themes of Data Management

The scientific and technical information (STI) within the French research organizations cover aspects of access, management and processing of scientific information and mainly the management of digital resources (subscription to major publishers), but also services as reporting, cataloguing and advocacy of scientific and technical production of the researchers, bibliometrics, scientometrics, as well as publishing activities. It is an inclusive activity that occurs at different stages of scientific production. We can distinguish two types of professionals, the librarians who work within libraries in French universities and the STI professionals who work in the research institutes (CNRS, INRA…). The professional status and the activities are different. STI professionals have developed more support activities for the data management than the librarians dedicated to the management of electronic resources and the libraries.
Irstea, as a research institute, has an STI service consisting of twenty people (24 persons equal to 22 full-time employees) working on a broad panel of activities (Figure 1). Faced with technological changes and developments in practices of researchers, STI at Irstea has developed a forward study on the evolution of STI services entitled "**Which STI mission statement for research in 2030?"**. This study, conducted in 2013, identified new activities

Fig. 1: Scientific and technical information activities are present all along the research process.



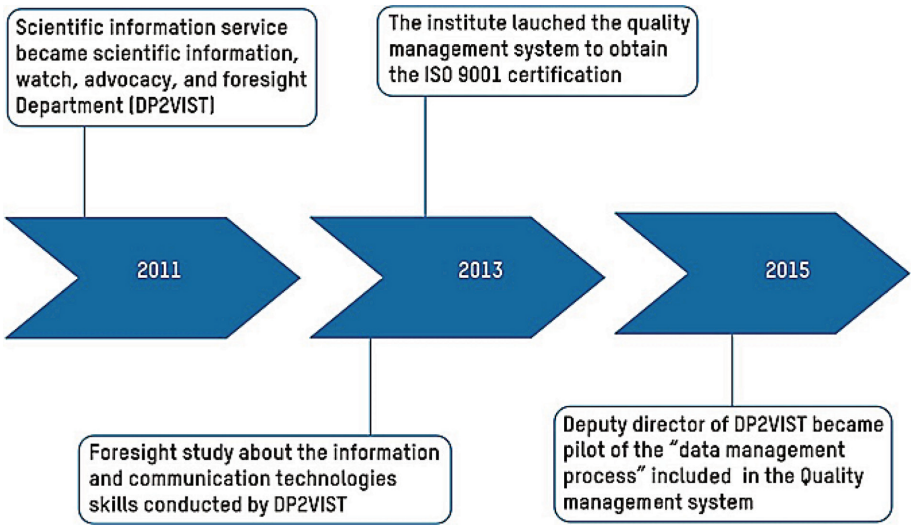
and missions around the data issue, and new services to propose for the researchers. The timeline of the events leading to and following this study are given in Figure 2.

Beyond this work, this has also allowed a better understanding of the entire scope covered by the STI staff in the institute, by researchers first and later by the top management. This initiative has also identified the cross-cutting nature of the issue of data management and put forward the necessary cooperation between services, especially with the IT services that are often considered legitimate naturally on technical aspects to process management data issues.

3.2. A New Path Worn by the Management of the Institute

The reorganization in 2011 has positioned the STI function in the research and innovation department, very close to the science strategy and also responsible for the scientific coordination. Moreover, the wish of the Presidency

Fig. 2: Timeline of the events.



of Irstea to obtain the ISO 9001 certification for the institute has developed a Quality approach of the data management. All Irstea scientific, business and support processes were identified and documented. The quality process "data management" is now steered by the head of the STI department; this legitimized the fact that STI staff is central point of the coordination of these activities, as well as computer scientists and database administrators or the IT staff (Grudzien & Hamrol, 2016). This context of quality approach has helped encourage librarians to get involved in this process, because it helps make a strategic plan for them (Saunders, 2015) A cross team "research data management" was put into place, representing all stakeholders of the data management. This team, led by the STI, organized a seminar open to all, around six objectives:

- Identifying scientific "data" advisors;
- Helping scientific teams to integrate all the operations necessary for the operational data management, and identifying support staff to contact for each step;
- Identifying cross skills in the data management issues;
- Building a model of data management plan;

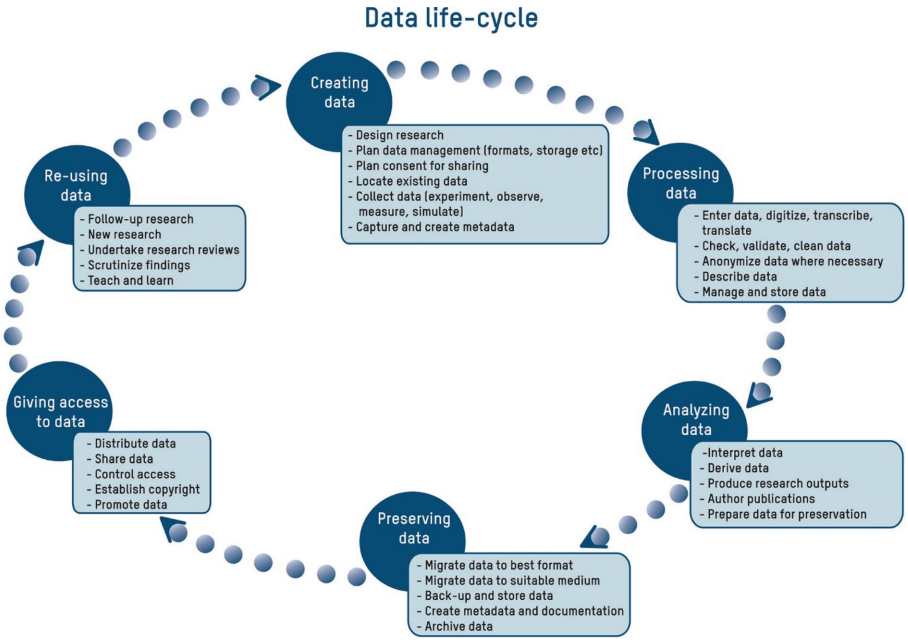
- Communication plan and training for researchers and stakeholders;
- Having communication tools for centralizing information.

Currently the team is open to IT experts, responsible for patenting and licensing, advocacy, legal and quality experts of the Institute, and it includes scientists for each project.

4. The Development of a Range of Appropriate Services

Out of the six activities concerning the data life cycle, four were identified as belonging to the natural STI services. The table below shows the type of service offered by the STI for each of these four activities, identified in the various stages of the cycle (Figure 3).

Fig. 3: From "the research data lifecycle" http://data-archive.ac.uk/create-anage/life-cycle, the scientific and technical information department works on the data lifecycle in order to position its activity in each step of the data lifecycle and to provide support for the research process concerning the data management.



Activity 1: "Creating Data"
The aim is to be involved in the framework of a research project, upstream projects or tenders.

Service offer	Description, examples
Support the researcher in the construction of a Data Management Plan (DMP)	STI and researcher: support for the creation of a DMP, questions to ask oneself, "checklist" of what needs doing. Advice on medium storage supports/formats/ standards…
Determine intellectual property rights	STI legal advice with the legal experts on intellectual property matters.

Activity 3: "Analysing the data" in a support role for the publication of articles.

Service offer	Description, examples
Link between articles and data: Creation of a DOI, advice on the choice of external data repository, publishing platforms, advocacy to register with ORCID	Watch and propositions based on institutional policy and guidelines, researcher practices and their publication strategy concerning data, international and national repositories and publishers policies.

Activity 4: "Preservation Data"

Service offer	Description, examples
Migration to the appropriate format	Advice on formats, storage medium supports, standards and on the structuring of database features
Creation of metadata	STI as data input in the IrsteaData catalogue based on Irstea policy, in relation to publications
Advice on digital archiving	Institutional archiving and procedure policy traditionally managed by STI so that this offer remains

Activity 5: "data access"

Service offer	Description, examples
Data sharing	STI: Policies covering data sharing, thematic platforms, editorials, "Open Data", "Linked Open Data", services offered
Technical support to make data accessible (publication and description)	Implementation of projects using linked data technologies and "Linked Open Data" (author-Publications-Data identification links) DOI-ORCID

5. Training and Sharing for New Skills

The IST team dedicated to the issue of data takes care of organizing the exchange of information and specific training sessions for the stakeholders. It becomes the central point of expertise and activities around data management within the STI network and the operational team for the "data management" quality process.

- **At the STI collective level**

During STI staff meetings, three areas have been prioritized:

1. the first is to share available information on the subject.
2. the second axis consists in organising training sessions on identified specific issues (see §4) including the practice of the data management plan.
3. The third axis deployed is the necessary involvement of the STI network in the sphere of Higher Education and Research at national and also at international levels. Irstea is therefore, by means of the STI head of department (and the association of STI managers of French research institutes—EPRIST) represented at a national level in the STI national coordination initiative, the "Digital Scientific Library" in segment 10 (http://www.bibliothequescientifiquenumerique.fr/bsn-10-donnees-de-la-recherche/) on data management, where a national policy on the subject is currently under discussion. Moreover, the STI network has participated for many years in inter-establishment networks of STI professionals (Renatis: http://renatis.cnrs.fr/; MEDICI: http://medici.in2p3.fr/; MATE SHS: http://mate-shs.cnrs.fr/). Lastly, several actors are members of the Research Data Alliance ("Libraries for research data interest group" and "agricultural vocabularies" interest group) and the ICSU World Data System (http://www.icsu-wds.org/community/working-groups/data-publication); actors have followed training organised within the framework of Foster projects. A general mobilisation by the scientific community and STI professionals at national scale around the French "*For a Digital Republic*" Act and its national consultation has also offered the opportunity to put the question of research data under the spotlight again, its necessary organisation and its diffusion to contribute to Open Science and Open Innovation.

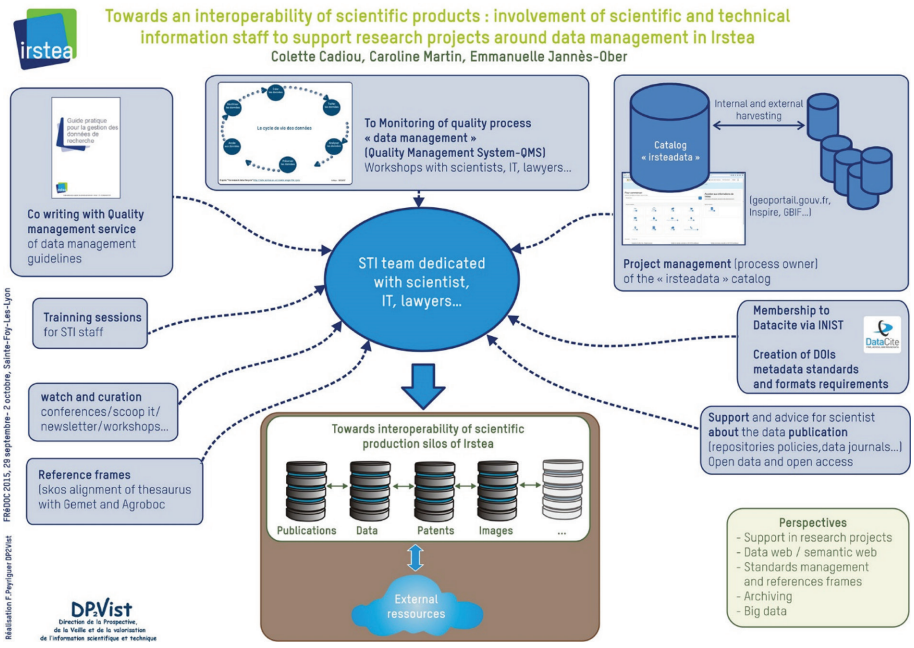
6. Services and Tools Developed or Underway

6.1. Operational Services

Operational services are developed in two ways: the first one encompasses advice and support for researchers within the framework of their research projects and the second one concerns the production of tools dedicated to data management (Figure 4). The help of services of checklists or recommendations as LIBER and RDA initiatives has been very useful (Christensen-Dalsgaard et al., 2012; Research Data Alliance, 2016).

– Advice types can be broken down into several themes: digital archiving (the scientific archiving policy is coordinated by the STI department), advice on formats to use, the metadata standards (Dublin Core…), and even on specific vocabularies (Agrovoc,

Fig. 4: Scientific and technical information department and his team offers a set of services dedicated to the data management, in the interoperable approach of tools.



Gemet…). Advice can also cover publishing strategies (type and nature of data to make it accessible and disseminate it, publication of "Data papers", data repositories policies).

- The deployment of reporting and identification tools:
  - The practical guide on data management: STI coordinated the editorial staff of Irstea's data management policy and contributed to the publication of the organisation's practical guide on data management;
  - The data management plan, developed in liaison with the quality mission, provides research teams with a tool for internal traceability of datasets produced for each research project;
  - STI has also implemented a creation service for DOIs, an identification tool for datasets in liaison with the "DataCite" initiative (https://www.datacite.org/), and encourages researchers to use ORCID as an identification tool, in collaboration with INIST-CNRS;
  - The Irstea thesaurus: a semantic project has been undertaken to update our thesaurus and its diffusion in SKOS format (mapping with other international standards such as Gemet and Agrovoc) with a view to its integration in other applications such as the future Irsteemeta catalogue of datasets and databases (IrsteaData).

6.2. Services and Tools: Development and Deployment

- The "IrsteaData" project is led by both STI (upstream) and IS management (downstream); this tool is dedicated to the identification and description of datasets produced by Irstea. The objective of the tool is to increase visibility of the scientific production while fully participating in the national movement towards open data. There is also a wish to link up "IrsteaData" and the data management plan to avoid duplicate entries and so to offer a genuine service to research teams, the only way to avoid the risk of this tool being perceived as a constraint.
- The development of communication campaigns for researchers to promote all services and tools, in particular by means of an intranet site dedicated to research data and its management.
- The deployment of "Text and Data Mining" tools (scientometrics).

7. Conclusion and Outlook

Finally, we will draw several lessons from what is going on as discussed above around research data management at Irstea:

1. The commitment of other management teams to construct a global Irstea service offer that includes STI services among others, fully affirms the legitimacy of STI to coordinate the "data management" process. Data management is actually transversal and involves all actors in scientific production, thus including complementary activities, necessitating hybrid skills that call upon, in particular, IT, legal, documentary, editorial and statistical techniques or knowledge. This requires adaptation to technological developments and changes in scientific practices. A good example would be the increase in use of text and data mining, which has already turned practices upside down, reintroducing an old problem which, in the era of Big Data, is now unavoidable: it is no longer possible, whatever the discipline and scale (team, research unit, institution, collaborative structure,…) to not organise data or to do so without integrating all its possible uses.
2. The inquiries and thinking around data management has allowed us to examine work methods, re-questioning not only STI workflows organisation, but also the organisation of links between scientific teams, IT management teams, advocacy, legal and other operational chains. Data management organisation has led us to overcome the inertia of working habits and separation of departments.
3. Data management also raises questions about the construction of information systems and interoperability of its information repositories (Agosti, Ferro, & Silvello, 2016). In the context of e-science, interoperability is more than ever a crucial issue for the dissemination of research products to decision makers and in society at large. This can only be achieved through an intelligent access and resource exploitation system both internally and externally. Moreover, the links between data, publications and all other scientific or technical productions (reports, patents…), produced by the same research project, is a middle-term objective of Irstea. Nevertheless, the institute is beginning to integrate its transversal and encompassing vision of its information system thanks to a close collaboration between STI (clearly identified as integrative spokesperson for the project

initiator) and the IT management team, whose position as project initiator is without ambiguity.

4. Data management also includes the question of the implementation of a metadata and shared standards policy at Irstea. In the medium term, and for the purpose of updating the information system, the implementation of metadata and "standards" associated with controlled vocabularies, such as the thesaurus, has become a central issue for ensuring interoperability between the different databases. Interoperability management will be achieved in particular by the creation of metadata templates, integrating the different internal policies: textual productions, data, scientific archives… This project is now ongoing in our institute.
5. Support for stakeholders in data management is crucial, and the setting up of transversal collectives bringing together scientists, librarians, IT experts and other skilled staff is necessary for the top-down and bottom-up initiatives concerning data convergence.

References

Agosti, M., Ferro, N., & Silvello, G. (2016). Digital library interoperability at high level of abstraction. *Future Generation Computer Systems, 55*, 129–146. https://doi.org/10.1016/j.future.2015.09.020.

ALA. (2014). The state of America's libraries—A report from the American Library Association. *American libraries, special issue*, *2014*, 81pp. Retrieved November 18, 2016, from http://www.ala.org/news/sites/ala.org.news/files/content/2014-State-of-Americas-Libraries-Report.pdf.

Christensen-Dalsgaard, B., van den Berg, M., Grim, R., Horstmann, W., Jansen, D., Pollard, T., & Roos, A. (2012). *Ten recommendations for libraries to get started with research data management.* Final report of the LIBER working group on E-Science, Research Data Management. Retrieved November 18, 2016, from http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf.

Grudzien, L., & Hamrol, A. (2016). Information quality in design process documentation of quality management systems. *International Journal of Information Management, 36*(4), 599–606. https://doi.org/10.1016/j.ijinfomgt.2016.03.011.

JISC. (2012). *Researchers of tomorrow: the research behaviour of generation Y doctoral students.* The British Library and HEFCE. Retrieved November 18, 2016, from http://www.webarchive.org.uk/wayback/archive/20140614205429/http://www.jisc.ac.uk/media/documents/publications/reports/2012/Researchers-of-Tomorrow.pdf.

Macmillan, D. (2014). Data sharing and discovery: What librarians need to know. *The Journal of Academic Librarianship*, *40*(5), 541–549. https://doi.org/10.1016/j.acalib.2014.06.011.

Research Data Alliance. (2016). *Going global with 23 things for research data management*. Retrieved November 18, 2016, from https://www.rd-alliance.org/going-global-23-things-research-data-management. dx.doi.org/10.15497/RDA00005.

Saunders, L. (2015). Academic libraries' strategic plans: Top trends and under-recognized areas. *The Journal of Academic Librarianship*, *41*(3), 285–291. https://doi.org/10.1016/j.acalib.2015.03.011.

Schmidt, B., & Shearer, K. (2016). *Librarians' competencies profile for research data management.* Joint task force on librarians' competencies in support of e-research and scholarly communication. Retrieved November 18, 2016, from https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf.