The role of libraries in supporting data exchange

Susan Reilly
Project Officer
LIBER : Association of European Research Libraries
The Hague, Netherlands
E-mail: Susan.reilly@kb.nl

Session: 116 — *The role of libraries in data curation, access and preservation: an international perspective* — Science and Technology Libraries

Abstract:

In November 2011 the EU funded Opportunities for Data Exchange project launched a Report on Integration of Data & Publications. This report examined the impact that data sharing is having on scholarly communications.

This report sought to reveal opportunities for supporting a more connected and integrated scholarly record. Four perspectives were considered, those of the Researcher, who generates or reuses primary data, Publishers, who provide the mechanisms to communicate research activities, and Libraries & Data Centers, who maintain and preserve the evidence that underpins scholarly communication and the published record.

There are clear opportunities for libraries to enable more joined up and vital scholarly record of modern research.

The library is well placed within the scholarly ecosystem to improve the sharing of scientific data through the provision of support in the following areas:

1. Availability 2. Findability 3. Interpretability 4. Reusability 5. Citability

6. Curation 7. Preservation

To further clarify the role of the library in supporting data exchange a survey was sent out to the LIBER network of libraries. Based on the report, the survey sought to establish what sort of support is being demanded by researchers of research libraries and what support is currently being offered. Responses from European research libraries were compared to leading libraries in the area of data management in the US and Australia. The survey showed, that although libraries are aware that there is a need to provide data

management support services to researchers, they feel they are currently lacking the skills to do so.

Keywords: *Opportunities for Data Exchange, Data Curation, Libraries, Training, Research Infrastructure*

The Big Opportunity

The Riding the Wave Report[1] points to the rising tide of scientific data as changing the nature of research, allowing researchers from different backgrounds to work on the same data sets, potentially solving some of society's grand challenges.

The Data Publication Pyramid (fig.1) illustrates a growing problem. The pyramid visualises the ways in which research data can be made available for sharing and reuse. The base of the pyramid represents, what is still a large proportion of research data[2], data stored locally in its raw form on hard drives and disks. There are several reasons[3] why these research data are not shared, varying from ethical, technical, security, or cultural reasons. The second layer of the pyramid visualises data which is already stored in repositories. This data is available for use and reuse. The third layer of data is linked to publications as supplemental files to articles. The top layer is the traditional view of an article or publication with the data embedded within.

The PARSE.Insight [4] project found that around 60 % of researchers would like to use the research data of others. Interestingly this study also found that over 40 % of researchers have real problems in sharing their own data. To reduce the size of the bottom layer of the pyramid, and increase the amount of sharing and reuse, the issues preventing researchers from sharing their data need to be addressed.

---

[1] The High Level Expert group on Scientific Data (2010), Riding the Wave, http://www.cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

[2] ParseInsight Survey (2009) http://www.parse-insight.eu/downloads/PARSE-Insight_D4-3_GapAnalysisFinalReport.pdf

[3] Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, et al. (2011) Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6(6): e21101.doi:10.1371/journal.pone.0021101

[4] Survey Report PARSE.Insight : http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf

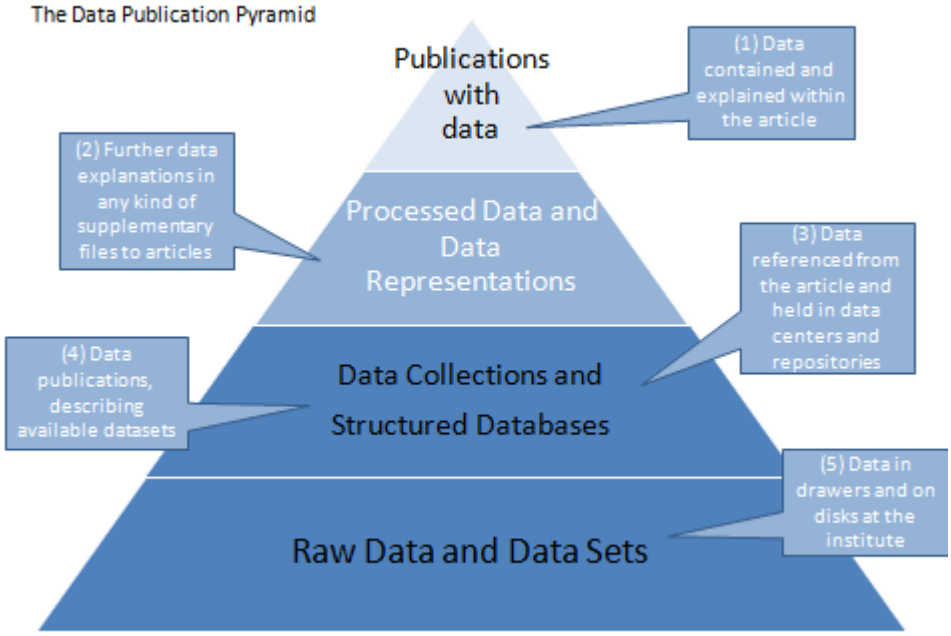

Fig.1. The ODE Data Publication Pyramid[5]

Some Perspectives on data sharing

The ODE report on the integration of data and publications[6] looked at the key concerns surrounding data sharing for researchers, publishers and archives. These issues revealed opportunities for each of the stakeholder groups, but for libraries in particular.
Of the key issues or areas of focus for researchers identified through the report the following present opportunities for libraries:

1. somewhere to put data and make it safe for reuse
2. the need to control its sharing and access
3. the ability to integrate data and publication
4. credit for data as a first class research object

[5] Reilly et al. (2011) ODE report on the integration of data and publications: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/ODE-ReportOnIntegrationOfDataAndPublications.pdf
[6] Reilly et al. (2011) ODE report on the integration of data and publications: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/ODE-ReportOnIntegrationOfDataAndPublications.pdf

A key area of focus for publisher, which libraries could also help address, is presenting data in more sophisticated formats to increase reuse, as good metadata is essential for this.

Based on the key areas of focus for each of the stakeholder groups, 7 areas of opportunity[7] for libraries can be identified:

1. Availability: in terms of helping researchers make their data available and also providing search services for data.
2. Findability: through the provisions of support for best practice in managing data.
3. Interpretability: through the provision of and training in metadescriptions
4. Reusability: through the provision of advice on the availability of subject archives and licensing for reuse
5. Citability: encouraging best practice in citations, the use of persistent identifiers
6. Curation: curation of data and provision of training on data curation
7. Preservation: advocating for good data management practices and archiving of data

Scope for current and future roles

It is clear from this study that there are very real opportunities for libraries to play a role in data exchange. The next step is to establish the impetus to harness these opportunities or, in other words, the demand from researchers that is being placed on libraries to provide support in these areas. With this in mind, a survey, based on the 7 areas of opportunity identified, was designed and sent out to the 430 libraries in the LIBER network. This survey asked respondents questions regarding the perceived demand for support in the areas of opportunity. It also asked respondents about how prepared they were to meet this demand and what they believed are the skills necessary to fully realise these roles. Finally, respondents were asked about the best way to deliver training to

[7] Reilly et al. (2011) ODE report on the integration of data and publications: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/ODE-ReportOnIntegrationOfDataAndPublications.pdf

develop the skills necessary. The survey received a very high response rate of nearly 20% and so can be judged as very representative of the state of play across research libraries in Europe. The survey was also send to a small selection of libraries in the US and Australia that were known to be leaders in the field of data management support provision. This provided an interesting contrast, especially in terms of the perception of the new skills necessary in this field.

The survey shows that research libraries perceive a very high demand from researchers for support in the area of data management. Over 80% of respondents perceive a demand for data management support. When we drill down and compare the areas of demand with corresponding support provision, the provision of support is disappointingly low. This is not surprising given that data sharing is still in a relatively early stage in most disciplines. Demand was the highest for archiving data (fig.3), with roughly only half of the demand being met. The biggest gap between demand and supply was for support in the creation of data management plans, with only 19% of libraries providing support in this area. The area of demand where most support is already being provided is for finding data as 60% of respondents provide support for finding data.

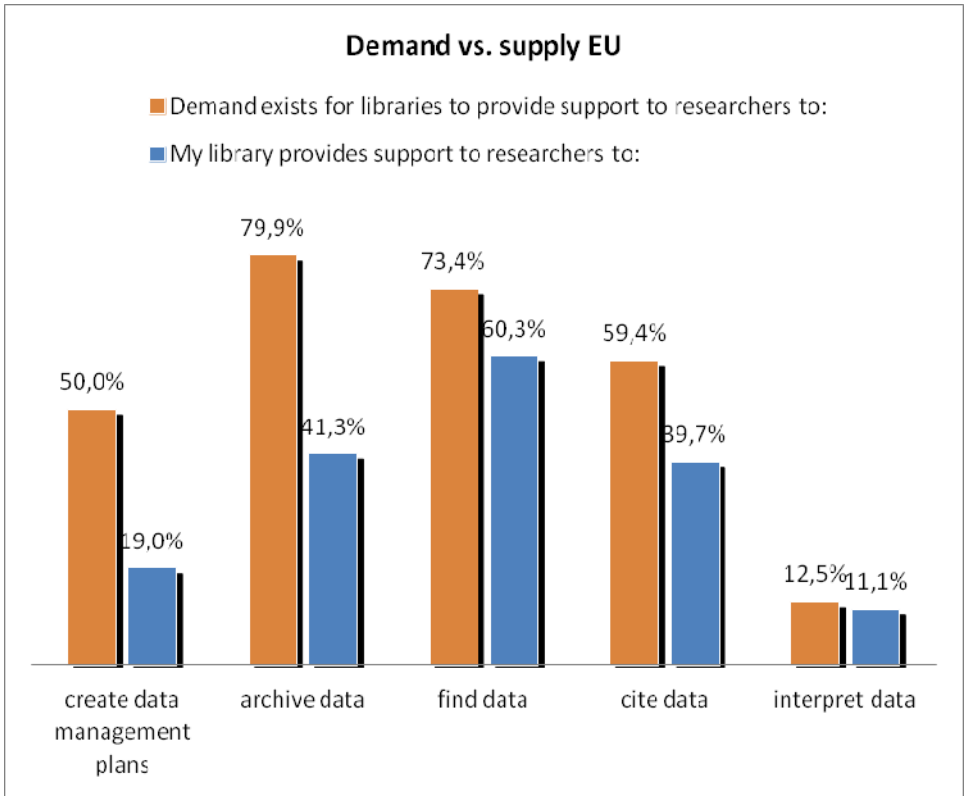

Fig.3.

Only 12% of libraries believe that they have the skills in place to meet the demand for support. Reassuringly 56% of libraries are now working on developing the skills that they perceive as necessary to provide data management support.

Developing skills

So, what are the additional skills that need to be developed if libraries are to meet the need for support in data management?
Respondents prioritised skills in the following order:

1. IT skills
2. Data curation and archiving
3. Subject specific research experience
4. Appraisal of research data
5. Visualisation of data holdings

What is interesting about this response is that the 'expert' library respondents overwhelmingly priorities the need for data curation and archiving skills. For them, IT skills came in 4th place in terms of priority. Could it be that experience shows that IT skills are not as important as perceived for libraries who are actively involved in data management support? Subject specific research experience is also a strong priority in terms of necessary additional skills identified by the 'expert' libraries.

The best means of developing all such skills, according to the libraries, is through the provision of continuing professional development. This is possible in all skills areas except for subject specific research expertise, which implies the need to recruit librarians with experience of discipline specific research.

The integration of data management into professional training courses was also a favoured option.

Conclusion

Regardless of how the necessary skills are prioritised, or the means by which they are developed, it is clear that significant investment needs to be made to develop these skills. Data sharing and e-science is changing the nature of research and, in doing so, is surfacing new opportunities for libraries. Libraries need to be quick to seize these opportunities but they also need to think carefully about the skills they decide to invest in developing. They need to reflect on the areas of support that are most in demand and also the areas that they can have the most impact given a limited availability of resources. In practice, certain skills may prove a better investment than others. Librarians are interested in and should receive continuing professional development, so that they can build on existing relevant skills. The library profession may also need to think carefully about the recruitment of new professionals and actively recruiting practising scientists into the profession.

References

ParseInsight Survey (2009) http://www.parse-insight.eu/downloads/PARSE-Insight_D4-GapAnalysisFinalReport.pdf

Reilly et al. (2011) ODE report on the integration of data and publications: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/ODE-ReportOnIntegrationOfDataAndPublications.pdf

Survey Report PARSE.Insight : http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, et al. (2011) Data Sharing by Scientists: Practices and Perceptions. PLoSONE 6(6): e21101. doi:10.1371/journal.pone.0021101

The High Level Expert group on Scientific Data (2010), Riding the Wave, http://www.cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf