Document title: Kaptur technical report
Last updated: May 2012

JISC

Project Information			
Project Acronym	Kaptur		
Project Title	Kaptur		
Start Date	3rd October 2011	End Date	29th March 2013
Lead Institution	University for the Creative Arts		
Project Director	Leigh Garrett, VADS Director		
Project Manager & contact details	Marie-Therese Gramstadt, VADS Projects Officer mtg@vads.ac.uk		
Partner Institutions	Glasgow School of Art; Goldsmiths, University of London; University of the Arts London		
Project Web URL	http://www.vads.ac.uk/kaptur/		
Programme Name	JISC Managing Research Data 2011-13		
Programme Manager	Simon Hodson		

Document Name			
Document Title	Kaptur technical analysis report		
Author(s) & project role	• Leigh Garrett, Kaptur Project Director • Carlos Silva, Kaptur Technical Manager • Marie-Therese Gramstadt, Kaptur Project Manager		
Date	04/05/2012	Filename	Kaptur_technical_report_v1_1.doc
URL	http://www.vads.ac.uk/kaptur/outputs/		
Access	☐ Project and JISC internal	☒ General dissemination	

Document History		
Version	**Date**	**Comments**
1	3rd May 2012	Main body of report completed.
1.1	4th May 2012	Formatting, addition of appendices and references.
1.2	22nd May 2012	Revised conclusion

CONTENTS

1. INTRODUCTION

1.1. BACKGROUND

Led by the Visual Arts Data Service (VADS) and funded by the JISC Managing Research Data programme (2011-13) KAPTUR will discover, create and pilot a sectoral model of best practice in the management of research data in the visual arts in collaboration with four institutional partners: Glasgow School of Art; Goldsmiths, University of London; University for the Creative Arts; and University of the Arts London.

1.2. RESEARCH METHOD

This report is framed around the research question: which technical system is most suitable for managing visual arts research data?

The first stage involved a literature review including information gathered through attendance at meetings and events, and Internet research, as well as information on projects from the previous round of JISCMRD funding (2009-11).

During February and March, the Technical Manager carried out interviews with the four KAPTUR Project Officers and also met with IT staff at each institution. This led to the creation of a user requirement document (Appendix A), which was then circulated to the project team for additional comments and feedback.

The Technical Manager selected 17 systems to compare with the user requirement document (Appendix B). Five of the systems had similar scores so these were short-listed. The Technical Manager created an online form into which the Project Officers entered priority scores for each of the user requirements in order to calculate a more accurate score for each of the five short-listed systems (Appendix C) and this resulted in the choice of EPrints as the software for the KAPTUR project.

2. TECHNICAL REQUIREMENTS

Selection criteria were agreed across the project partners in order to evaluate software and to make sure it falls within defined requirements of the project. In this research, we evaluated the software based on the following main requirements (more detailed information can be found in Appendix A).

2.1. SOFTWARE TYPE AND COST

Software is evaluated based on its type, open source or commercial software, with a strong preference for open source software.

Research Data Management (RDM) software costs vary widely depending on the product and level and scale of the repository; the range is limited by the KAPTUR project budget.

2.2. STORAGE REQUIREMENTS

The software will need to be able to handle different types of data, from simple and small text items to complex and large multimedia items with the flexibility or potential to include unusual file formats.

2.3. INTERFACE REQUIREMENTS

The software should comply with W3C standards[1], provide quality assurance features, and have a user-friendly upload tool.

2.4. SYSTEM REQUIREMENTS

The physical system requirements describe whether it can run in certain environments such as operating systems, virtual servers and cloud storage environments. Consideration will also be given to defined limits for data upload and the ability to integrate the software with tools and other software currently in use by the partner institutions.

2.5. INSTITUTIONAL REQUIREMENTS

This includes the specific requirements from each partner institution in terms of workflow, statistical reporting, legal, preservation and disposal of data.

[1] World Wide Web Consortium Standards http://www.w3.org/standards/

3. SHORT-LISTED TECHNICAL SYSTEMS

From a total of 17 different systems that were assessed (Appendix B), in the final phase of the selection process five systems were short-listed as they were all capable of fulfilling the requirements for the KAPTUR project: DataFlow, DSpace, EPrints, Fedora, and Figshare (Appendix C).

3.1. DATAFLOW

DataFlow is an open source software project which is developing and promoting a free-to-use cloud-hosted system for management, preservation and publication of research datasets.

The project is based on the prototype developed by the JISC funded ADMIRAL[2] project (2009-11) which looked at a two-tier federated data management infrastructure for use by life science researchers. This provides services to meet researchers' local data management needs for the collection, digital organisation, metadata annotation and controlled sharing of research datasets, and an easy and secure route for archiving annotated datasets to an institutional repository, The Oxford University Data Store. The Data Store assigns Digital Object Identifiers (DOIs) and uses Creative Commons licensing, it also enables long-term preservation and access to research data.

STRENGTHS

- DataFlow offers a simple deposit interface managed by either an administrator or the researchers themselves.
- It provides a structured metadata collection interface.
- The system offers a popular storage approach similar to Dropbox.

[2] The ADMIRAL Project: A Data Management Infrastructure for Research Across the Life sciences http://imageweb.zoo.ox.ac.uk/wiki/index.php/ADMIRAL

WEAKNESSES

- DataFlow is currently under development and although it has been releasing development versions of the software for both DataBank and DataStage, its current version is not yet for public release and production.
- There are also issues with the installation and setup of the current version, which the developers of DataFlow are assessing and correcting.
- Further tools such as WebDAV (Web Distributed Authoring and Versioning)[3] and compatibility with the SWORD v2[4] resource deposit protocol will be released shortly[5], however further tests and trials must be undertaken before considering the application stable and ready for use in a production environment.

3.2. DSPACE

DSpace was designed to capture, store, index, preserve and provide access to institutional digital research materials. It is open source and was created by the Massachusetts Institute of Technology (MIT) and Hewlett-Packard; it has a large community of developers and users[6].

DSpace is written in Java and will run on any Linux or UNIX system and Windows XP. It is available under the BSD open source License, which permits proprietary commercial use of the software and incorporation of the code into proprietary products.

DSpace is a web-accessible system and any modern web browser is capable of submitting and accessing content in DSpace.

[3] WebDAV website http://www.webdav.org/

[4] SWORD (Simple Web-service Offering Repository Deposit) v2 http://swordapp.org/category/sword2/

[5] originally expected on 24th April 2012

[6] DSpace Community https://wiki.duraspace.org/display/DSPACE/Home

STRENGTHS

- DSpace provides a comprehensive workflow system where users can upload items and associated metadata via the web interface. Each individual repository installation can tailor the workflow process to accommodate the needs of its varying user-types.
- The metadata is based upon the Dublin Core Metadata Schema[7], adapted by MIT Libraries to meet DSpace requirements. DSpace calculates and retains a checksum for each item uploaded so that the integrity of the item can be verified at a later date, and the validity of the file periodically checked.
- In most cases the software is able to identify the file format of a deposit.
- DSpace supports preservation by providing a *Bitstream Format*[8] for each file format type in the system.
- Concepts from the OAIS (Open Archival Information System[9]) Information model will map to DSpace.

WEAKNESSES

- The development of separate custom modules is not as straight forward as with EPrints.
- Out-of-the-box DSpace doesn't provide a visual interface such as the EPrints Kultur plugin[10].

3.3. EPRINTS

EPrints was developed at the University of Southampton and is freely available as open source software. Originally designed for creating and managing open access institutional

[7] Dublin Core Metadata Initative (DCMI) Metadata Terms http://dublincore.org/documents/dcmi-terms/

[8] DSpace documentation http://www.dspace.org/1_6_0Documentation/ch02.html#N10463

[9] DCC OAIS Overview http://www.dcc.ac.uk/webfm_send/435

[10] Although the JISC funded EXPLORER project (2011) applied some of the Kultur features to the DSpace repository software http://explorer.our.dmu.ac.uk/

repositories of digital research papers and publications, EPrints is now used to store and manage a much broader range of content types and data.

Led by the University of Southampton, the JISC funded Kultur[11] project (2007-09) piloted a model for repositories suitable for the specialist needs of arts researchers, and founded start-up repositories for research outputs at University of the Arts London[12] and University for the Creative Arts[13].

STRENGTHS

- EPrints can accommodate different types of workflows; these can be edited to provide different options such as sending email notifications to administrators and editors.
- Content can be stored in any file format as designated by the repository administrator during configuration. Multiple representations of the same content are also permitted.
- With the release of EPrints version 3.3[14] (September 2011) repository managers can install applications with '1-click' through the EPrints Bazaar, described as an 'App Store'. These applications can be downloaded and installed in the repository without affecting the core configuration and original settings of the repository. The applications can also be easily disabled or deleted.

WEAKNESSES

- EPrints, as any other open source software relies on project funding, this means that once a project completes the plugins may not be supported or upgraded to fit with the latest version of EPrints.

[11] Kultur project website http://kultur.eprints.org/

[12] UAL Research Online http://ualresearchonline.arts.ac.uk/

[13] UCA Research Online http://www.research.ucreative.ac.uk/

[14] EPrints 3.3 Stable http://eprintsnews.blogspot.co.uk/2011/09/eprints-33-stable.html

- In order to 'kulturise'[15] a repository a series of plugins must be installed and tested before setting up a production environment.
- With the exception of the applications available in the Bazaar, most of the configuration must be performed manually.

3.4. FEDORA

Fedora (Flexible Extensible Digital Object and Repository Architecture) is a general-purpose open source digital object repository management system for managing and delivering digital content. It was developed at Cornell University together with the University of Virginia in 1999, it can manage multiple object types within a single implementation and it is used in a range of repositories around the World but mainly in the United States.

The Fedora repository is available under the Educational Community License. It runs as a service within an Apache Web Server with Tomcat; the server is backed in part by a relational database or it can be configured to work with MySQL setups.

STRENGTHS

- The system is highly scalable and can provide support for upwards of 10 million objects[16].
- Different client and end user interface applications can be installed and integrated with the core distribution to provide enhanced functionality and user services.
- Fedora incorporates a number of features that support preservation including use of XML and open standards such as SOAP[17] (Simple Object Access Protocol) and METS[18] (Metadata Encoding and Transmission Standard).

---

[15] Term arising out of the JISC funded Kultivate project (2010-11) to mean enhancing a repository for the specialist needs of arts researchers.

[16] DCC Technology Watch: Fedora http://www.dcc.ac.uk/webfm_send/463

[17] SOAP http://www.w3.org/TR/soap/

[18] METS http://www.loc.gov/standards/mets/

- Concepts from the OAIS Information model will map to Fedora.

WEAKNESSES

- Fedora's functionality is dependent on the additional functionality provided by client applications; it can be a challenge to further develop and enhance the repository from its original setup.
- Quality assurance: a researcher or user can upload a record into the repository and make it available to the community without it being checked by an editor or repository manager.
- Workflow is not integrated into the basic repository system and requires a separate application service.

3.5. FIGSHARE

Figshare is a web-based platform aimed at researchers. It was originally developed as an 'open science project' by Mark Hahnel whilst he was completing his PhD at Imperial College, University of London; it is now supported by Digital Science[19] (since September 2011) and was re-launched with improved functionality in January 2012.

Researchers are encouraged to publish all their research outputs online, including negative data and unpublished data. Persistent identifiers are provided by the Handle System[20]; Creative Commons licenses are used; and there are tools to enable searching and sharing of data.

STRENGTHS

- Figshare offers a simple deposit interface managed directly by the researchers themselves.

[19] Digital Science website http://www.digital-science.com/

[20] Handle System website http://handle.net/

- It also offers an interactive interface where any published data is presented according to its file type.
- Its upload tool allows multiple uploads using WebDAV and javascript.
- The development team is currently working on a desktop uploader[21] to allow a more streamlined process of submission.
- The application uses Web 2.0 tools to enhance the sharing experience.

WEAKNESSES

- Figshare currently lacks a quality assurance system or method where an editor or repository administrator can check a record before it is made publicly available.
- Currently the software is not available for download which means that the research data is hosted by Amazon Web Services[22] (AWS), Figshare's hosting providers.
- It is not SWORD compliant; although integration with EPrints or other repository software may be possible in the future.

4. SELECTION OF SOFTWARE

Following the analysis of the findings, there were four main recommendations:

User requirements - the four institutions selected essential for the list of user requirements (or would be essential in the future) and also added additional features (Appendix B).

Open source software - open source software is preferred by the institutional partners, as well as recommended by the project's funder, JISC. Whilst open source software has several benefits it also comes with risks in terms of ongoing development and support.

[21] Figshare features http://figshare.com/features

[22] Amazon Web Services http://aws.amazon.com/

The final five - based on the user requirements, 17 systems (Appendix C) were shortlisted to five: DataFlow, DSpace, EPrints, Fedora, and Figshare (Appendix D). It was more difficult to make a selection from these five as potentially they would all have been suitable.

EPrints for visual arts research data - the research methodology led to the choice of EPrints open source repository software for the KAPTUR pilot technical system. This decision is additionally supported by the four institutional partners' choice of EPrints for the publication of their research outputs.

5. CONCLUSION AND RECOMMENDATIONS

The first stage of the research reduced the choice of software to five options, which were all found to be suitable for managing research data in the visual arts. Of these a further selection process reduced the choice of software to three strong contenders: EPrints, Figshare, and DataFlow. EPrints is already in use at the partner institutions, and has been both graded and ratified by the Project Officers as the most viable option which fulfills most of the requirements of the project. However EPrints is not a clear-cut winner in that the grading by the partner institutions was very close between the three, and there are elements in the other two, Figshare and DataFlow, which fulfill some of the requirements that the EPrints software is not able to perform (or which would require development work): a 'local' file management environment; improved visualization of documents and multimedia; a user friendly upload feature; and a WebDAV interface.

In order to completely fulfill the project requirements, it is recommended that two pilots occur side by side: an integration of EPrints with Figshare and a separate piece of work linking DataFlow's DataStage with EPrints. By integrating EPrints with Figshare, the project can take advantage of a system which has been built with, and for, researchers to handle research data specifically, and has a user-friendly visual interface (which is constantly evolving and enhanced by Figshare directly). Future developments include: integration with DataCite for persistent identifiers (Figshare currently uses the Handle System) and a desktop uploader to make uploading research data even easier.

There are some risks associated with using Figshare:

- in principle the platform where it is based is free for use as long as the research data is published, if the data needs to remain private there is an allowance of 1Gb, after which a charge is made to the user or institution;
- certain exclusions and possibly hosting fees may be required as part of the integration with EPrints;
- additional data protection and security issues will need to be addressed such as data storage location and authentication mechanisms in order to match the partner requirements.

By integrating DataStage with EPrints the research data storage and software will be hosted within each institution, providing them with better control over the type of data that can be stored, published and managed. The integration will also enable content uploaded in DataStage to be securely backed up by the institution and accessible from anywhere in the world. A 'Dropbox'-like tool is featured in the latest beta version, providing a user-friendly interface which will benefit visual arts researchers. EPrints will effectively provide the role of DataFlow's DataBank.

The risks associated with using DataStage from the DataFlow Project are:

- it is a work in progress and currently in development; the current download is a beta release;
- support is not guaranteed after the project completes; meaning that bug fixes and other issues will rely on whether the work is undertaken by the Open Source community;
- setting up the system will depend on the appropriate documentation and technical specifications of the DataFlow project; currently virtual machines are available for download, however further configuration and fixes are required.

In conclusion, there is no single product which can completely fulfill all the requirements of the Kaptur project partners, therefore piloting EPrints, as the main choice of system, with the addition of features from two of the other systems will allow the project team to test, explore and document findings, further advantages or disadvantages and present a more comprehensive and viable pilot research data management system for the visual arts.

6. ACKNOWLEDGEMENTS

We would like to thank the KAPTUR Project Officers and their colleagues in the IT Departments for their contributions to this report. We would also like to thank our funders, JISC, and Simon Hodson, JISCMRD Programme Manager.

7. REFERENCES

arXiv. Available from: http://arxiv.org/ Cornell University Library. [Accessed 4 May 2012]

CUBRID. Available from: http://www.cubrid.org/ [Accessed 4 May 2012]

DataFlow. Available from: http://www.dataflow.ox.ac.uk/ University of Oxford. [Accessed 4 May 2012]

Drizzle. Available from: http://www.drizzle.org/ [Accessed 4 May 2012]

Dropbox. Available from: https://www.dropbox.com/ [Accessed 4 May 2012]

DSpace. Available from: http://www.dspace.org/ DuraSpace. [Accessed 4 May 2012]

EPrints. Available from: http://www.eprints.org/ University of Southampton. [Accessed 4 May 2012]

Fedora. Available from: http://fedoraproject.org/ [Accessed 4 May 2012]

Figshare. Available from: http://figshare.com/ Digital Science. [Accessed 4 May 2012]

Firebird. Available from: http://www.firebirdsql.org/ Firebird Foundation Incorporated. [Accessed 4 May 2012]

Google Drive. Available from: https://drive.google.com/ Google. [Accessed 4 May 2012]

InfoSphere. Available from: http://www-01.ibm.com/software/data/infosphere/ IBM. [Accessed 4 May 2012]

Ingres. Available from: http://www.actian.com/products/ingres Actian Corporation. [Accessed 4 May 2012]

Invenio. Available from: http://invenio-software.org/ [Accessed 4 May 2012]

Mendeley. Available from: http://www.mendeley.com/ [Accessed 4 May 2012]

MS Zentity. Available from: http://research.microsoft.com/en-us/projects/zentity/ Microsoft Corporation. [Accessed 4 May 2012]

Sybase. Available from: http://www.sybase.com/ [Accessed 4 May 2012]

8. APPENDICES

8.1. APPENDIX A: USER REQUIREMENTS

This version dated 26th March 2012

1. Storage Requirements

a. Metadata requirements

The RDM system should be able to integrate with, and/or make available content into existing local institutional systems. For example, the project partners use the EPrints[23] repository software to publish their research outputs[24].

The metadata requirements identified are listed below with an asterisk next to mandatory fields; additional metadata fields may be needed to facilitate integration with local systems.

- Additional Information (large text field)
- Creators (text field)*
- Date Created (date field)*
- Date Embargo (date field)
- Date Last Accessed[25] (date field)

---

[23] EPrints http://www.eprints.org/software/

[24] Institutional Research Repositories for the four partners are located: http://radar.gsa.ac.uk/ http://eprints.gold.ac.uk/ http://www.research.ucreative.ac.uk/ http://ualresearchonline.arts.ac.uk/

- Description (large text field)
- DOI[26]
- Funders (text field)
- Institutional or Group Creators (text field)
- Keywords (text field)
- License (text field)
- Location/Venue (text field)
- Material (text field)
- Measurements or Duration (text field)
- Number of Pieces (text field)
- Publisher (automatically generated based on the institution's name)*
- References (large text field)
- Related Exhibitions (text field)
- Related Publications (text field)
- Related URLs (text field)
- Rights (text field)*
- Subjects (based on LOCSH[27] or JACS[28])
- Title (text field)*
- Unique ID (integer field)*

b. Multimedia items

- Audio (AC3)
- Audio (FLAC)
- Audio (MP3/MPEG)
- Audio (OGG)
- Audio (WAV)

---

[25] Required by EPRSC, unless it is recorded elsewhere in the system

[26] institutions need to contact a DataCite Managing Agent in order to mint DOIs http://datacite.org/membership

[27] Library of Congress Subject Headings (LCSH) http://id.loc.gov/authorities/subjects.html

[28] Joint Academic Coding System (JACS) http://www.hesa.ac.uk/content/view/1776/649/

- Audio (WMA)
- Image (bmp)
- Image (gif)
- Image (jpeg)
- Image (pdf)
- Image (photoshop)
- Image (png)
- Image (TIFF)
- PDF
- Video (AVCHD)
- Video (AVI)
- Video (Flash)
- Video (MP4)
- Video (MPEG)
- Video (Quicktime)
- Video (Windows Media)

c. Text items

- Microsoft Word
- N3
- PDF
- Plain Text
- RDF/XML
- Rich Text (RTF)
- XML

d. Any other items

- Archive (7ZIP)
- Archive (BZ2)
- Archive (TGZ)
- Archive (ZIP)
- Blogs
- HTML
- Links to external websites and other resources

- Microsoft Excel
- Microsoft Power Point
- Postscript
- Tweeter data (transcription files)
- Wikis

2. Interface Requirements

a. Logical flow

The flow of the system should be streamlined but at the same time provide the potential for interacting with other systems. The basic requirements from the interface will be:

- LDAP Authentication
- Upload tool for files and metadata
- QA/approval
- Publication of data
- Preservation of data
- Data disposal

b. Capture method

Based on existing non-institutional systems used by interviewees in the Environmental Assessment report; it was proposed that the best capture method for active research data would be a "Dropbox[29] like" folder where users are able to create as many folders as needed per project (depending on the amount of space allocated) and upload content into the system without the need for authenticating more than once.

c. Search tool

At a minimum, a single Boolean search tool is required in order to find items stored within the system.

[29] Dropbox https://www.dropbox.com/

d. User interface compliant with web standards

The user interface will need to comply with the following World Wide Web Consortium (W3C)[30] standards and recommendations:

- Accessibility and semantic guidelines
- Browser compatibility
- Character encoding
- Compliance with W3C Markup Validation Service[31]
- Standards for harmonization and the web accessibility initiative
- Valid CSS
- Valid HTML pages
- Valid JavaScript pages
- Valid metadata
- Valid XML (when needed)

3. System Requirements

a. Operating System

The preferred Operating System across the four partner institutions is Microsoft Windows, however it is possible to install other environments with different Operating Systems such as Virtual Servers or Virtual Machines running Linux or other types of Unix based systems.

b. Virtual Server vs. Physical Server

The preferred option is Virtual Servers with flexible and resizable disk space.

c. Storage requirements

It is expected that the software can hold individual accounts with unlimited storage however, the system administrators are expected to be able to define a limit per account/user.

d. Cloud storage (allowance)

[30] W3C Standards http://www.w3.org/standards/

[31] W3C Markup Validation Service http://validator.w3.org/

Cloud storage is permitted in all the institutions; however there are policies, procedures and regulations currently in review, which might affect the choice of cloud hosting company and company location. Sustainability is a major factor to be considered; once the project is rolled out, who will pay for the hosting, maintenance and other overheads from this project[32].

e. Maximum file size to upload allowed

For the purposes of this project it is proposed that file sizes are restricted to 1GB per upload; unless allowed otherwise by the institution's IT department and/or hosting service.

f. Integration with institutional systems

Integration with LDAP is required in order to streamline the authentication workflow for users. Integration with EPrints software for the publication and display of research data is also required.

g. Backup and disaster recovery procedures

- Daily incremental backups
- Weekly full backups
- Monthly full backups
- Daily replication data
- Tapes
- Scheduling and backup media rotation
- Tape labeling
- Retention cycle
- Backup tape testing

h. Software Security Assurance

The selected software will need to provide the following security measurements:

- Firewall enabled for internet facing software
- Password required for private area/content
- SSL for encryption when users need to authenticate and submit credentials

[32] Business Costs and Sustainability Plans will be created at each institution.

- Ensure W3C standards; minimize cross-site scripting and injection attacks
- Penetration testing
- Source Code reviews
- Informal reviews by developers
- Formal reviews by a review group

i. Access and Permissions

Access to the software will need to be granted to:

- Defined users in the LDAP database(s) from each institution
- Users who will use the software (user rights – upload and publish individual content)
- Repository Managers (editorial rights – as per above plus ability to review content and restrict, return and take down items as appropriate)
- System administrators (admin rights – as per above plus general administration of the site)

4. Institutional Requirements

a. Workflows

Three workflows are required:

- Uploading content and metadata: create a folder -> upload content and metadata
- Publishing content: select content from folder -> assign a record where content will fit as research data OR create a new record based on the data
- Take down content: select file(s) from folder -> unpublish data

And in addition at least one Repository Manager with editable rights should be created to have overall control of the public facing interface and Quality Assurance of content made available online by the users/researchers.

b. Statistical reporting

- Google analytics to be setup for website traffic analysis and monitoring
- _addItem() function to track individual items from the repository

c. Legal requirements

The software selected will need to comply with general legal policies such as:

- Intellectual Property Rights (IPR)
- Freedom of Information (FOI) Act
- Data Protection Act
- Information Security Policy
- Records Management Policy
- Research Data Management (RDM) Policy

More specifically, the software and database will need to be held within the European Union to comply with data protection law and comply with IPR, FOI and the Data Protection Act.

d. Preservation and disposal of data

In order to comply with funder requirements[33], and because research data is a valuable institutional asset, selected research data will need to be preserved for the longer term. This means that the RDM system will need to provide scalability to cope with large amounts of data stored over long periods of time. The Repository Manager will be responsible for the disposal of data according to the institution's policies and procedures.

7.2. APPENDIX B: COMPARISON OF 17 SYSTEMS

This version dated 26th April 2012

Five of the 17 systems (12 are described in more detail in the spreadsheets below) were not short-listed for the following reasons:

1. **arXiv** - Not considered as arXiv is an e-print service in the fields of physics, mathematics, non-linear science, computer science, quantitative biology, quantitative finance and statistics.

[33] For example the EPSRC require research data to be available for at least "[...] 10 years from the end of any researcher 'privileged access' or, if others have accessed the data, from last date on which access to the data was requested by a third party." DCC guidance on EPSRC requirements http://www.dcc.ac.uk/resources/policy-and-legal/research-funding-policies/epsrc

2. **Dropbox** - Dropbox was only considered as part of the data ingest stage, however it doesn't fulfill the complete set of requirements and at the moment can't be modified from its original software, therefore it is not considrered.
3. **Google Drive** - Google Drive was only considered as part of the data ingest stage, however it doesn't fulfill the complete set of requirements and at the moment can't be modified from its original software.
4. **Mendeley** - Not considered as its primary focus is on making PDF files available.
5. **Sybase** - Sybase is an SAP company with an enterprise software and services company offering software to manage, analyze, and mobilize information, using relational databases, analytics and data warehousing solutions and mobile applications development platforms. The system is focused on mobile solutions rather than research data management and therefore it wasn't short-listed.

Requirement/Category	CUBRID	DataFlow	Drizzle	DSpace	EPrints	Fedora
Software Type						
Open Source	X	X	X	X	X	X
Storage Requirements – capable of handling						
Metadata	X	X	X	X	X	X
Multimedia	X	Limited multimedia tools	X	Limited multimedia tools	X	Limited multimedia tools
Text Items		X	X	X	X	X
Other types of items		X	X	X	X	X
Interface Requirements						
Upload tool for files and metadata	X	X		X	X	X
QA/approval		Limited QA		Limited QA	X	
Publication of data	X	X		X	X	X
Preservation of data	X	X		X	X	X
Data disposal		X		X	X	X
User friendly upload feature		X				
Search tool		X	X	X	X	X
Compliant with W3C standards	X	X	X	X	X	
System Requirements – capable of having/running under						
Windows OS	X			X		
Virtual Servers	X	X	X	X	X	X
Unlimited Storage		X		X	X	X
Cloud Storage	X	X	X	X	X	X
Upload large files up to a maximum of 1GB per upload		X		On request - depending on institution	On request - depending on institution	On request - depending on institution
Integration with LDAP	X	X	X	X	X	X
Integration with existing Institutional Repositories		X		X	X	
Backup and disaster recovery procedures		X	X	X	X	X
Software Security Assurance						
Institutional Requirements						
Workflows - uploading content and metadata, publishing content and take down content		X		X	X	Limited workflow modifications
Statistical reporting			X	X	X	X
Legal requirements		X		X	X	X
Preservation and disposal of data	X	X		X	X	X
Additional Requirements						
Mobile access						
API/Web Service/XML outputs		X	X	X	X	X
Internal links with other resources such as Eprints systems					Limited	
SWORD 2 Compliant		X		X	X	X
WebDAV interface		X		Limited tools to allow WebDAV	Limited tools to allow WebDAV	Limited tools to allow WebDAV
Able to handle large amounts of data	X	X	X	X	X	X
TOTAL	13	27	14	28	28	24

Requirement/Category	Figshare	Firebird	InfoSphere	Ingres	Invenio	MS Zentity
Software Type						
Open Source		X		X	X	
Storage Requirements – capable of handling						
Metadata	X	X	X	X	X	X
Multimedia	X	X	X	X	Limited	
Text Items	X	X	X	X	X	X
Other types of items	X	X	X	X	X	
Interface Requirements						
Upload tool for files and metadata	X		X		X	X
QA/approval	Limited QA				X	
Publication of data	X				X	X
Preservation of data	X		X		X	X
Data disposal	X		X		X	
User friendly upload feature	X				X	
Search tool	X	X	X	X	X	Limited Search Tool
Compliant with W3C standards	X	X	X	X	X	
System Requirements – capable of having/running under						
Windows OS	X	X	X	X		X
Virtual Servers		X	X	X	X	
Unlimited Storage	Limited		X			
Cloud Storage	X	X	X	X	X	
Upload large files up to a maximum of 1GB per upload	On request		X	X	X	
Integration with LDAP	X	X	On request	X	X	Limited to third party products
Integration with existing Institutional Repositories	X		X			
Backup and disaster recovery procedures	X	X		X	X	
Software Security Assurance						
Institutional Requirements						
Workflows - uploading content and metadata, publishing content and take down content	X		X		X	X
Statistical reporting	X	X	On request	X	X	
Legal requirements	X					
Preservation and disposal of data	X	X	X	X		X
Additional Requirements						
Mobile access						
API/Web Service/XML outputs	X	X	X	X		X
Internal links with other resources such as Eprints systems						
SWORD 2 Compliant			X			
WebDAV interface	X					
Able to handle large amounts of data	X	X	X	X		
TOTAL	26	16	22	17	20	10

7.3. APPENDIX C: COMPARISON OF 5 SHORT-LISTED SYSTEMS

This version dated 3rd May 2012

Requirement/Category	DataFlow	DSpace	EPrints	Fedora	Figshare
Software Type					
open source	7.25	7.25	7.25	7.25	
Storage Requirements – capable of handling					
Metadata	7.75	7.75	7.75	7.75	7.75
Multimedia (display)	4.125	4.125	8.25	4.125	8.25
Text Items	8.5	8.5	8.5	8.5	8.5
Other types of items	8.5	8.5	8.5	8.5	8.5
Interface Requirements					
Upload tool for files and metadata	8.5	8.5	8.5	8.5	8.5
QA/approval	3.875	3.875	7.75		3.875
Publication of data	7.5	7.5	7.5	7.5	7.5
Preservation of data	6.5	6.5	6.5	6.5	6.5
Data disposal	6	6	6	6	6
User friendly upload feature	7.5				7.5
Search tool	7	7	7	7	7
Compliant with W3C standards	6.5	6.5	6.5	6.5	6.5
System Requirements – capable of having/running under					
Windows Server		6.5			6.5
Virtual Servers	6	6	6	6	
Unlimited Storage	6	6	6	6	3
Cloud Storage	6	6	6	6	6
Upload large files up to a maximum of 1GB per upload	6.5	6.5	6.5	6.5	6.5
Integration with LDAP	6.5	6.5	6.5	6.5	6.5
Integration with existing Institutional Repositories	6.75	6.75	6.75		6.75
Backup and disaster recovery procedures	6.5	6.5	6.5	6.5	6.5
Software Security Assurance					

Institutional Requirements					
Workflows - uploading content and metadata, publishing content and take down content	6.5	6.5	6.5	3.25	6.5
Statistical reporting		6.25	6.25	6.25	6.25
Legal requirements	5.75	5.75	5.75	5.75	5.75
Preservation and disposal of data	5.75	5.75	5.75	5.75	5.75
Additional Requirements					
Mobile access					
API/Web Service/XML outputs	6.5	6.5	6.5	6.5	6.5
Internal links with other resources such as Eprints systems			3.375		
SWORD 2 Compliant	6	6	6	6	
WebDAV interface	5.5	2.75	2.75	2.75	5.5
Able to handle large amounts of data	7.25	7.25	7.25	7.25	7.25
TOTAL	**177**	**180**	**184**	**159**	**171.75**