The International Journal of Digital Curation
Volume 8, Issue 2 | 2013

Here, KAPTUR This! Identifying and Selecting the Infrastructure Required to Support the Curation and Preservation of Visual Arts Research Data

Leigh Garrett,
Director, Visual Arts Data Service,
University for the Creative Arts

Marie-Therese Gramstadt,
Research Associate, Visual Arts Data Service,
University for the Creative Arts

Carlos Silva,
Planning and Technical Manager, Visual Arts Data Service,
University for the Creative Arts

Abstract

Research data is increasingly perceived as a valuable resource and, with appropriate curation and preservation, it has much to offer learning, teaching, research, knowledge transfer and consultancy activities in the visual arts. However, very little is known about the curation and preservation of this data: none of the specialist arts institutions have research data management policies or infrastructure and anecdotal evidence suggests that practice is ad hoc, left to individual researchers and teams with little support or guidance. In addition, the curation and preservation of such diverse and complex digital resources as found in the visual arts is, in itself, challenging. Led by the Visual Arts Data Service, a research centre of the University for the Creative Arts, in collaboration with the Glasgow School of Art; Goldsmiths College, University of London; and University of the Arts London, and funded by JISC, the KAPTUR project (2011-2013) seeks to address the lack of awareness and explore the potential of research data management systems in the arts by discovering the nature of research data in the visual arts, investigating the current state of research data management, developing a model of best practice applicable to both specialist arts institutions and arts departments in multidisciplinary institutions, and by applying, testing and piloting the model with the four institutional partners. Utilising the findings of the KAPTUR user requirement and technical review, this paper will outline the method and selection of an appropriate research data management system for the visual arts and the issues the team encountered along the way.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

Led by the Visual Arts Data Service (VADS)[1], a research centre of the University for the Creative Arts, and funded by the JISC Managing Research Data Programme (JISC MRD 2011-13)[2] KAPTUR[3] seeks to 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.

The first stage of the project focused on an environmental assessment (Garrett & Gramstadt, 2012; Garrett et al., 2012) which included eight short informal interviews, sixteen in-depth recorded and transcribed interviews, a literature review, information gathered through attendance at various meetings and events, desk research and information collected from projects reporting from the previous round of JISC MRD funding (2009-11).

Following on from the publication of the environmental assessment report in February 2012, the Technical Manager embarked on a series of interviews with the four KAPTUR Project Officers and with information technology staff at each partner institution, with the purpose of creating a user requirements document for the curation and preservation of research data in the visual arts. The draft was circulated to the project team for additional comments and review (the final analysis can be found in Appendix 1).

With reference to the user requirements, the Technical Manager identified seventeen potential systems that could be relevant to the curation and preservation of visual arts research data (details can be found in Appendix 2). Using a basic scoring mechanism, based on one point per requirement, five of these systems were identified as potential solutions and selected for further detailed analysis. The Technical Manager created an online questionnaire and the KAPTUR Project Officers were asked to enter priority scores for each of the requirements in order to calculate a more accurate score for each of the five potential solutions (see Appendix 3 for analysis). EPrints, figshare and DataStage were selected as the preferred options for the KAPTUR project.

User Requirements

As outlined, the selection criteria were agreed with appropriate representatives from the four institutional partner institutions and used to evaluate potential software solutions, bearing in mind the scope and resources of the project. Throughout this stage of the project the team identified five key requirements (full details can be found in Appendix 1).

[1] VADS: http://www.vads.ac.uk

[2] JISC MRD: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx

[3] KAPTUR: http://www.vads.ac.uk/kaptur

 doi:10.2218/ijdc.v8i2.273

Solution Types

Research data management software costs vary widely but generally can be ascribed to two main types: open source or commercial software. The partners expressed a preference for open source solutions, which aligns with recommendations made by the funders.[4]

Storage

The project team identified a requirement for the research data management solution to be able to handle a variety of different types of data, from simple and small text items to large complex multimedia items, with the flexibility or potential to include non-standard file formats.

Interface

It was agreed that the solution should comply with W3C standards, provide quality assurance features and a user-friendly and engaging upload service.

System

System requirements identified the environments – such as operating systems, virtual servers and cloud storage environments – that any potential solution might need to address. Consideration was also 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.

Institutional

Institutional requirements included specific requirements from each partner institution in terms of workflow, statistical reporting, legal compliance, preservation and disposal of data.

Technical Review

From the total of seventeen systems that were identified (Appendix 2), five were selected as the most suitable for use with visual arts research data: DataFlow, DSpace, EPrints, Fedora, and figshare (Appendix 3). Each of these were then considered by the team during the selection process with reference made to issues facing the visual arts.

DataFlow

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

[4] See: http://www.jiscinfonet.ac.uk/infokits/open-source

The project is based on the prototype developed by the JISC-funded ADMIRAL project (2009-11)[5] which developed a two-tier federated data management infrastructure for use by life science researchers. DataFlow provides services to meet researchers' local data management needs for the collection, digital organisation, metadata annotation and controlled sharing of research datasets; and provides 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;
- A structured metadata collection interface;
- A popular storage approach, similar to Dropbox.

Weaknesses

- Although DataFlow has been releasing development versions of the software for both its DataBank and DataStage solutions, its current version is not yet ready for public release;
- There are issues with the installation and setup of the current version, which the developers of DataFlow are investigating;
- Additional tools, such as WebDAV[6] and compatibility with the SWORD v2[7] resource deposit protocol, have recently been released in beta version. However, further tests and trials must be undertaken before considering the application stable and ready for use in a production environment.

DSpace

Dspace[8] is a web based application designed to capture, store, index, preserve and provide access to institutional digital research outputs. It was created by the Massachusetts Institute of Technology (MIT) and Hewlett-Packard, and has a large community of developers and users.

DSpace is written in Java and will run on any Linux or UNIX system and Windows XP. It is available under an open source license, which permits proprietary commercial use.

[5] ADMIRAL: http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmi/admiral.aspx
[6] Web Distributed Authoring and Versioning
[7] SWORD v2: http://swordapp.org/sword-v2
[8] Dspace: http://www.dspace.org

 doi:10.2218/ijdc.v8i2.273

Strengths

- DSpace provides a comprehensive workflow system where users can upload items and associated metadata, and each repository installation can tailor the workflow process to accommodate the needs of its host institution and users;
- The metadata is based upon the Dublin Core Metadata Schema, 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, 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 for each file format type in the system;
- Concepts from the OAIS Reference Model[9] 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 that provided by the EPrints Kultur plugin.

EPrints

EPrints[10] was developed at the University of Southampton and is freely available under an open source licence. Originally designed for creating and managing open access institutional 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 project (2007-09)[11] 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 and University for the Creative Arts.

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 manager during configuration, and multiple representations of the same content are permitted;

[9] Open Archival Information System
[10] Eprints: http://www.eprints.org
[11] KULTUR: http://kultur.eprints.org

- With the release of EPrints version 3.3 (September 2011) repository managers can install applications, plugins and updates with the EPrints Bazaar. These can be downloaded and installed without affecting the core configuration and settings of the repository, and applications can also be easily disabled or deleted.

Weaknesses

- EPrints, as with other open source software, often relies on project funding. This means that once a project completes the plugins may not be supported or upgraded to fit with subsequent versions;
- To be useful to the visual arts, a repository manager must install and test a series of plugins;
- With the exception of the applications made available via the Bazaar, most of the configuration must be performed manually.

Fedora

Fedora[12] is a general-purpose, open source digital object repository management system for managing and delivering digital content. Developed by Cornell University and 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 databases.

Strengths

- The system is highly scalable and can provide support for upwards of 10 million objects;
- 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 Simple Object Access Protocol (SOAP) and the Metadata Encoding and Transmission Standard (METS);
- Concepts from the OAIS Reference Model will map to Fedora.

Weaknesses

- Fedora 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;

[12]Fedora: http://fedoraproject.org

- 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.

figshare

figshare[13] is a web-based platform aimed at addressing the needs of individual researchers. Originally developed as an 'open science project' by Mark Hahnel whilst he was completing his PhD at Imperial College, University of London, figshare is now supported by Digital Science (from September 2011) and was re-launched with improved functionality in January 2012.

Researchers are encouraged to publish all their research data online, including negative data and unpublished data. Persistent identifiers are provided by the Handle System, 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;
- The interface is interactive, presenting published data according to its file type;
- The upload tool allows multiple uploads using WebDAV and Javascript;
- The development team is currently working on a desktop uploader to create a more streamlined submission process, collaborative spaces and the release of an API;
- 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;
- The software is not available for download, which means that the research data is hosted by a commercial hosting service, Amazon Web Services, (figshare's hosting solution);
- It is not SWORD compliant, although integration with EPrints or other repository software may be possible in the future.

[13] figshare: http://figshare.com

Selection: Round One

Following presentation of the initial draft of the technical review, and in discussion with the project partners the following recommendations were made:

- To update the user requirement to include a matrix of priorities, including those which could be reasonably expected to be essential for future use (Appendix 1) and added additional features (Appendix 2);
- To select an open source option as the preferred solution, although it was recognised by the project team that such solutions are also associated with risks, particularly in terms of ongoing development and support;
- To select five potential solutions, based on the user requirement from the original seventeen systems: DataFlow, DSpace, EPrints, Fedora, and figshare (Appendix 3). All five scored highly for the visual arts;
- To select EPrints as the preferred option to curate and preserve research data for the visual arts. This was reinforced by the fact that the four institutional partners currently use EPrints to support the publication of research outputs. This is of particular relevance due to the relationship between, and characteristics of, research data and research outputs in the visual arts.

Selection: Round Two

Following the initial selection of five potential solutions, a further review was undertaken using a matrix of priorities defined by the Project Officers. This returned the following scores, in order of usefulness to the visual arts:

1. Eprints (184.00),
2. DSpace (180.00),
3. DataFlow (177.00),
4. figshare (171.75),
5. Fedora (159.00).

EPrints was graded and verified by the Project Officers as the most viable option because it fulfilled most of the requirements of visual researchers and their host institutions. However, it was also acknowledged that the scoring of all the solutions was extremely close and there were elements in two (figshare and DataFlow) which fulfilled some of the requirements that the EPrints software was not able to perform without further development work. These included a local file management environment, improved visualisation of documents and multimedia, an enhanced user friendly upload feature, and increased WebDAV functionality.

Recommendations

To fully appreciate and understand how best to meet the research data management requirements of researchers and their institutions, it was recommended that two pilots

 doi:10.2218/ijdc.v8i2.273

were undertaken in parallel: an integration of figshare with EPrints; and DataFlow's DataStage with EPrints.

figshare with EPrints

By integrating figshare with EPrints, the advantage is a system which has been built with, and for, researchers to handle research data specifically, with a visually engaging interface, which will be of particular appeal to visual arts researchers. In addition, figshare anticipates future developments, including integration with DataCite for persistent identifiers and a desktop uploader to make uploading research data even more straightforward for researchers.

However, the project team recognise that there are some risks associated with using figshare:

- Currently the service is free to use as long as the research data is published. If data needs to be private there is an allowance of 1Gb, after which a charge is made;
- 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 to meet the partner requirements.

DataFlow's DataStage with EPrints

By integrating DataStage with EPrints the research data storage and solution will be hosted within each institution, which may provide greater control and standardisation for the institution. The integration will also enable content uploaded in DataStage to be securely backed up by the institution and accessible through a Web browser interface. A 'Dropbox'-like tool is featured in the latest beta version, providing a user-friendly interface which will benefit visual arts researchers. EPrints would effectively provide the role of DataFlow's DataBank.

The risks associated with using DataStage are:

- It is currently in development and the current version is a beta release;
- Support is not guaranteed after the project completes (July 2013). This could mean 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 being made available. Currently, virtual machines are available for download but further configuration and fixes are required.

Conclusion

There is no single solution which can completely fulfil all the requirements of researchers, research teams and their host institutions in the visual arts. The piloting of EPrints, as the preferred choice, with the addition of features from two of the other systems will allow the project team to investigate, test, document and identify a more comprehensive and viable research data management system for the visual arts.

Acknowledgements

We would like to thank the KAPTUR Project Officers: Robin Burgess, John Murtagh, Tahani Nadim and Anne Spalding; and their information technology colleagues for their contributions to this report. We would also like to thank our funders, JISC, and in particular Simon Hodson, JISCMRD Programme Manager for his ongoing support to the project team.

References

Garrett, L. & Gramstadt, M. (2012). *KAPTUR: Exploring the nature of visual arts research and its effective management.* Paper presented at the Electronic Visualisation and the Arts conference. London, UK. Retrieved from http://ewic.bcs.org/upload/pdf/ewic_ev12_s5paper1.pdf

Garrett, L., Gramstadt, M., Burgess, R., Murtagh, J., Spalding, A. & Nadim, T. (2012). JISC-funded KAPTUR project environmental assessment report. Visual Arts Data Service (VADS), University for the Creative Arts. Retrieved from http://www.research.ucreative.ac.uk/1054

 doi:10.2218/ijdc.v8i2.273

Appendix 1: User Requirements

Table 1 was created following interviews with four KAPTUR Project Officers and IT staff at each partner institution.

Table 1. User requirements for the curation and preservation of research data in the visual arts.

Requirement	Category and Explanation	Detail
Storage		
Metadata	The system should be able to integrate with, and/or make available content into existing institutional systems. For example, the project partners use the EPrints repository software to publish their research outputs.	Additional Information (large text field) Creators (text field)* Date Created (date field)* Date Embargo (date field) Date Last Accessed (date field) Description (large text field) DOI 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 or JACS) Title (text field)* Unique ID (integer field)* * mandatory

Requirement	Category and Explanation	Detail
Item type	Multimedia	Audio (AC3)
		Audio (FLAC)
		Audio (MP3/MPEG)
		Audio (OGG)
		Audio (WAV)
		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)
	Text	Microsoft Word
		N3
		PDF
		Plain Text
		RDF/XML
		Rich Text (RTF)
		XML
	Other	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

 doi:10.2218/ijdc.v8i2.273

Requirement	Category and Explanation	Detail
Interface		
	Integration	LDAP Authentication
		Upload tool for files and metadata
		QA/approval
		Publication of data
		Preservation of data
		Data disposal
	Capture	Evidence indicates that the best capture method for the visual arts is a "Dropbox like" folder whereby users are able to create as many folders as needed per project and upload content into the system without the need for authenticating more than once
	Search	At a minimum, a single Boolean search tool is required in order to find items stored within the system
	Interface	Accessibility and semantic guidelines
	The user interface will need to comply with the W3C standards and recommendations.	Browser compatibility
		Character encoding
		Compliance with W3C Markup Validation Service
		Standards for harmonization and the web accessibility initiative
		Valid CSS
		Valid HTML pages
		Valid JavaScript pages
		Valid metadata
		Valid XML (when needed)
System		
	Operating environment	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
	Virtual and physical	The preferred option is virtual servers with flexible and resizable disk space
	Storage	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

Requirement	Category and Explanation	Detail
	Cloud	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 service. Sustainability is a major factor to be considered; once the project moves into the production environment rolled there are ongoing costs of hosting, maintenance and other overheads to consider
	File size	For the purposes of this project it is proposed that file sizes are restricted to 1GB per upload; unless allowed otherwise by the partners information technology staff and/or hosting service
	Integration	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
	Disaster recovery	Daily incremental backups Weekly full backups Monthly full backups Daily replication data Tapes Scheduling and backup media rotation Tape labelling Retention cycle Backup tape testing
	Security	Firewall enabled for internet facing software Password required for private area/content SSL for encryption when users need to authenticate and submit credentials Ensure W3C standards; minimise cross-site scripting and injection attacks Penetration testing Source code reviews Informal reviews by developers Formal reviews by a review group

 doi:10.2218/ijdc.v8i2.273

Requirement	Category and Explanation	Detail
	Access and permissions	Defined users in the LDAP from each institution Users who will need to upload and publish research data Research data managers will need editorial rights, to upload, publish, review, restrict, return and take down research data System administrators, who will require administrator rights – including those previously detailed and administration of the research data repository
Institutional		
	Workflows	Three workflows are required: uploading content and metadata; publishing content; and take down content. In addition, at least one administrator 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.
	Statistical reporting	Google analytics to be setup for website traffic analysis and monitoring _addItem() function to track individual items from the repository
	Legal compliance	The software selected will need to comply with all legal and institutional 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 The data will need to be held within the European Union to comply with data protection law and comply with IPR, FOI and Data Protection Act

Requirement	Category and Explanation	Detail
	Preservation and disposal	In order to comply with funder requirements, and because research data is a valuable institutional asset, selected research data will need to be preserved. This means that the solution will need to provide scalability to store large amounts of data stored over long periods of time. The administrator will be responsible for the disposal of data according to the institution's policies and procedures

Appendix 2: Solutions Comparison

Five of the initial seventeen systems were not short-listed for the following reasons:

- arXiv was not considered as it is an e-print service in the fields of physics, mathematics, non-linear science, computer science, quantitative biology, quantitative finance and statistics.
- Dropbox was only considered as part of the data ingest stage. However it doesn't fulfil the complete set of requirements and at the moment can't be modified as required.
- Google Drive was only considered as part of the data ingest stage. However it doesn't fulfil the complete set of requirements at the moment, as required.
- Mendeley was not considered as its primary focus is on making PDF files available.
- Sybille is a SAP company with an enterprise software and services company offering software to manage, analyse, and mobilise information, using relational databases, analytics and data warehousing solutions and mobile applications development platforms. However, the system is focused on mobile solutions rather than research data management.

The following were analysed against the user requirements:

- CUBRID
- DataFlow
- Drizzle
- DSpace
- EPrints
- Fedora

 doi:10.2218/ijdc.v8i2.273

- figshare
- Firebird
- InfoSphere
- Ingres
- Invenio
- MS Zentity

Table 2. Initial systems comparison with user requirements (X: meets requirement; L: limited requirement; R: requires development).

Requirement/ Solution	CUBRID	DataFlow	Drizzle	DSpace	EPrints	Fedora	figshare	Firebird	InfoSphere	Ingres	Invenio	MS Zentity
Software Type												
Open source	X	X	X	X	X	X	X	X		X	X	
Storage												
Metadata	X	X	X	X	X	X	X	X	X	X	X	X
Multimedia	X	L	X	L	X	L	X	X	X	X	L	
Text items		X	X	X	X	X	X	X	X	X	X	X
Other types of items		X	X	X	X	X	X	X	X	X	X	
Interface												
Upload tool for files and metadata	X	X		X	X	X	X		X		X	X
QA/approval		L		L	X		L				X	
Publication of data	X	X		X	X	X	X				X	X
Preservation of data	X	X		X	X	X	X		X		X	X
Data disposal		X		X	X	X	X		X		X	
User friendly upload feature		X					X				X	
Search tool		X	X	X	X	X	X	X	X	X	X	L
Compliant with W3C standards	X	X	X	X	X		X	X	X	X	X	
System												
Windows OS	X			X			X	X	X	X		X
Virtual servers	X	X	X	X	X	X		X	X	X	X	
Unlimited storage		X		X	X	X	L		X			
Cloud storage	X	X	X	X	X	X	X	X	X	X	X	

Requirement/ Solution	CUBRID	DataFlow	Drizzle	DSpace	EPrints	Fedora	figshare	Firebird	InfoSphere	Ingres	Invenio	MS Zentity
Upload large files up to a maximum of 1GB per upload		X		R	R	R	R		X	X	X	
Integration with LDAP	X	X	X	X	X	X	X	X	R	X	X	L
Integration with existing institutional repositories		X		X	X		X		X			
Backup and disaster recovery procedures		X	X	X	X	X	X	X		X	X	
Software security assurance												
Institutional Requirements												
Workflows - uploading content and metadata, publishing content and take down content		X		X	X	L	X		X		X	X
Statistical reporting			X	X	X	X	X	X	R	X	X	
Legal requirements		X		X	X	X	X					
Preservation and disposal of data	X	X		X	X	X	X	X	X	X		X
Additional												
Mobile access												
API/web service/XML outputs		X	X	X	X	X	X	X	X	X		X
Internal links with other resources such as Eprints systems					L							
SWORD 2 compliant		X		X	X	X			X			
WebDAV interface		X		L	L	L	X					
Able to handle large amounts of data	X	X	X	X	X	X	X	X	X	X		
TOTAL	**13**	**27**	**14**	**28**	**28**	**24**	**27**	**16**	**22**	**17**	**20**	**10**

 doi:10.2218/ijdc.v8i2.273

Appendix 3:
Final Scoring

Table 3. Matrix of priorities. (E: Essential requirement; D: Desirable requirement).

Requirement/ Category	Partner 1		Partner 2		Partner 3		Partner 4	
	E	D	E	D	E	D	E	D
Storage								
Metadata	X		X		X		X	
Multimedia	X		X		X			X
Text items	X		X		X		X	
Other types of items	X		X		X			X
Interface								
Upload tool for files and metadata	X		X		X		X	
QA/approval	X		X		X		X	
Publication of data	X		X		X			X
Preservation of data	X		X		X		X	
Data disposal	X		X		X			X
User friendly upload feature	X		X		X		X	
Search tool	X		X		X		X	
Compliant with W3C standards	X		X		X		X	
System								
Windows OS	X		X		X		X	
Virtual servers	X		X		X			X
Unlimited storage	X		X		X		X	
Cloud storage		X		X		X	X	
Upload large files up to a maximum of 1GB	X		X		X		X	
Integration with LDAP	X		X		X		X	
Integration with EPrints	X		X		X		X	
Backup and disaster recovery procedures	X		X		X		X	
Software security assurance	X		X		X			X

Requirement/ Category	Partner 1		Partner 2		Partner 3		Partner 4	
	E	D	E	D	E	D	E	D
Institutional								
Workflows - uploading content and metadata, publishing content and take down content	X		X		X		X	
Statistical reporting	X		X		X		X	
Legal requirements	X		X		X		X	
Preservation and disposal of data	X		X		X		X	
Additional								
Mobile access				X				
API/web service/XML outputs			X					
Internal links with other resources, such as Eprints systems		X						
SWORD 2 compliant		X						
WebDAV interface		X						
Able to handle large amounts of data		X						

Table 4. Overall solution scores.

Requirement/Category	DataFlow	DSpace	EPrints	Fedora	figshare
Software Type					
Open source	7.25	7.25	7.25	7.25	
Storage					
Metadata	7.75	7.75	7.75	7.75	7.75
Multimedia (display)	4.13	4.13	8.25	4.13	8.25
Text items	8.50	8.50	8.50	8.50	8.50
Other types of items	8.50	8.50	8.50	8.50	8.50
Interface Requirements					
Upload tool for files and metadata	8.50	8.50	8.50	8.50	8.50
QA/approval	3.88	3.88	7.75		3.88
Publication of data	7.50	7.50	7.50	7.50	7.50
Preservation of data	6.50	6.50	6.50	6.50	6.50

doi:10.2218/ijdc.v8i2.273

Requirement/Category	DataFlow	DSpace	EPrints	Fedora	figshare
Data disposal	6.00	6.00	6.00	6.00	6.00
User friendly upload feature	7.50				7.50
Search tool	7.00	7.00	7.00	7.00	7.00
Compliant with W3C standards	6.50	6.50	6.50	6.50	6.50
System Requirements					
Windows server		6.50			6.50
Virtual servers	6.00	6.00	6.00	6.00	
Unlimited storage	6.00	6.00	6.00	6.00	3.00
Cloud storage	6.00	6.00	6.00	6.00	6.00
Upload large files up to a maximum of 1GB per upload	6.50	6.50	6.50	6.50	6.50
Integration with LDAP	6.50	6.50	6.50	6.50	6.50
Integration with existing institutional repositories	6.75	6.75	6.75		6.75
Backup and disaster recovery procedures	6.50	6.50	6.50	6.50	6.50
Software security assurance					
Institutional Requirements					
Workflows - uploading content and metadata, publishing content and take down content	6.50	6.50	6.50	3.25	6.50
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	6.50	6.50	6.50	6.50	6.50
Internal links with other resources such as Eprints systems			3.38		
SWORD 2 compliant	6.00	6.00	6.00	6.00	
WebDAV	5.50	2.75	2.75	2.75	5.50
Able to handle large amounts of data	7.25	7.25	7.25	7.25	7.25
TOTAL	**177.00**	**180.00**	**184.00**	**159.00**	**171.75**