

Table 2. Comparison of open source repository software features

Category	Feature	DSpace 5.x	Fedora 4.x with Islandora CLAW	Fedora 4.x with Hydra	Dataverse 4.x
	VISION	The DSpace Project will produce the world's choice for repository software providing the means for making information openly available and easy to manage.	The Fedora Project's mission is to lead the advancement of Free and open source software and content as a collaborative community.	One body, many heads - Tailored applications and workflows for different content-types, contexts and user interactions. Building a framework, not an application.	The Dataverse is an open source web application to share, preserve, cite, explore and analyze research data.
INFRASTRUCTURE	Server resources	Linux, Windows, postgres, mvn, ant, Java Serverlet (TomCat etc)	Linux	Linux 5 (Red Hat), Windows, TomCat, solr etc...Deployment hardware information and server structure is complex and variable see https://goo.gl/jbHxgf .	linux, glassfish, postgresSQL, solr
	Local / Hosted solutions	Local and hosted @ ~ \$5000 p.a.	Local and hosted. Hosting through Discovery Garden @ \$12000 p.a.	Local	Local and hosted
	Hardware required	8G RAM, 1TB storage (dependant on content)	VMs to run docker containers	vms for: Fedora & solr and hydra & rails and oracle database. Deployment hardware information and server structure is complex and variable see https://goo.gl/jbHxgf	min: 1 VM, 2 2.8 GHz, 8 GB mem, 50 GB disk; max: 6 servers, 64 GB ram each, TBs storage from NAS using NFS, hardware load balancer
	Build scripts	mvn	Docker, Ansible	ansible, vagrant	vagrant, puppet, ansible
	Open source	Yes	Yes	Yes	Yes

Table 2. Comparison of open source repository software features

	Platform customization	Yes, modularity through Manakin, xmlui much easier than jsui which required java development to customize	Through community	Through community. Ruby on Rails development	Possible with development in Java
	Support community	Yes - thousands of institutions	130 Institutions or DG company	30 official partner institutions (and growing annually) and many working collaborators	17+ institutions
	Administrator configurations	Yes, with restrictions	Highly configurable	Highly configurable	Yes, per dataverse
	User roles	Yes, with restrictions	Yes	Yes	Yes, per dataverse and per user and per file. Pre-defined and custom roles
USER INTERFACE/FRONT-END	Integrated front-end	Yes, JSUI(java) or Manakin (xml)	Islandora	Hydra (see also Sufia http://sufia.io/ hydra powered repository front end)	Yes
	Customisable design	Somewhat through Manakin, including branding	Yes	Yes, easily skinned UI	Yes. Group dataverses can be skinned too
	Mobile optimised design	Yes - Manakin	Yes	Yes	Yes
	Browse options	Yes - faceted (any field provided in item description) provided by Solr	Yes, faceted (via Solr)	Customisable, gated discovery	Yes, faceted browsing for citation metadata only. Faceted search is possible (see below)

Table 2. Comparison of open source repository software features

METADATA	Customisable	Yes - Requires developer input: extensive and collection specific.	Yes, requires developer input: default set to Dublin Core. Customizable for MODs and MODS to Dublin transforms. Any xml-based schema possible. Solution packs provide domain support	Yes. MODS default and ability to attach descriptive metadata. Can convert metadata to another schema 'on the fly' and DC, EAD & PBCore are all being used by partner institutions	Potential to develop in conjunction with authors
	Required fields (DC)	Yes	Yes - can set	Yes, basic and extendible (customisable)	Yes. DDI, DCMI for citation metadata
	Exporting schemas	OAI-PMH, xml, METs	OAI-PMH, DC, MODS, xml, and others depending on solution pack	OAI-PMH and most others due to conversion ability	JSON and XML
	Schema flexibility	Flexible	Flexible - any metadata schema.format	Flexible, but some customisation requires developer input	3 levels: Citation, domain (astronomy, life sciences, social & humanities), file-level metadata
	Validation	Yes	Yes	Yes	Yes
	Versioning	Yes	Yes	Yes	Yes
	Attachable discipline-specific metadata document	Yes	Yes	Yes, and ability to convert schema	No but automatic ingest from select file types
	Compound object handling	Somewhat, but in development? Manakin can link metadata	Yes - with solution package	Yes, compound and complex	No, but dataverses can be nested
PRIVACY/SECURITY	Embargo	Yes	With scholar module	Yes	Yes, up to the curator of the community to publish.

Table 2. Comparison of open source repository software features

	Auto-lift embargo support	Not sure	Yes- admin controlled, with scholar module and possibly Islandora IP embargo https://wiki.duraspace.org/display/ISLANDORA/Islandora+Scholar	Not sure	Not sure
	Private/public data	Yes	Yes	Yes, groups too	Yes, groups too
	Authentication mechanisms	Yes, integration mechanisms: Password, Shibboleth, LDAP, IP, X.509	Yes - integrates with LDAP, CAS and Shibboleth	Yes - integrates with LDAP, CAS and Shibboleth	Shibboleth (LDAP apparently)
	Access controls	Yes, 'groups'	Yes, highly configurable	Yes, multi level for differential access.	Per dataverse and per file. Role-based as well & custom roles. Notification of access requests
SUBMISSION/INGESTION	Data types accepted	All types	All types	All types	All types
	Size (storage allocated, upload limits etc.)	1GB file limit, but can be increased to 2GB. Without Cacoon, can be increased to 4GB.	1 TB via REST API	1 TB via REST API	2 GB limit via API vs larger for API. Compressed archives limited to 1000 files for automatic unpack
	Metadata data submission	Basic DC fields mandatory upon submission process, other fields can be added	Mandatory fields can be customised	Mandatory fields can be customised	Mandatory fields can be set.
	Automated metadata import from files	Yes, from csv	Yes, from csv	Yes, from csv	Yes for FITS data
	Data deposit API	Yes, SWORD	Yes, SWORD	Yes, SWORD	Yes, SWORD

Table 2. Comparison of open source repository software features

	Error reporting	Unsure - to be determined	Unsure - to be determined	Unsure - to be determined	Unsure - to be determined
	Declaration form	Yes - upon ingest, can be customised	Yes - upon ingest, can be customised	Yes, each digital object requires a rights declaration on ingest	Templates are available but unsure if presented each time upon submission.
	Curation layer	Yes - collection level	Yes, 'mediated deposit system' and ability to set workflow systems	Yes, 'mediated deposit system' and ability to set workflow systems	Yes
ACCESS/SHARING	Licensing	Yes, customisable - as part of manual submission process. (creative commons, community/collections, repository submission agreement)	Yes, customisable	Yes, customisable	Yes, CC0 by default but opt-out possible
	Integration with other institutional repositories/collections	Linkable via metadata harvesting (OAI-PMH)	Linkable via metadata harvesting (OAI-PMH)	Yes, using Blacklight it can aggregate content from multiple repositories with links back to the source systems, allowing unified search and browse of separate repository systems/software	Very efficient at linked Dataverse repositories and has OAI-PMH functionality so should be linkable via metadata harvesting
	Linked objects	Yes	Yes	Yes, complex arrangements are possible	Unsure

Table 2. Comparison of open source repository software features

	Private project/group spaces	Yes, but not collaborative	Unsure, but likely possible with development	Yes - development required	Yes - collaborative working spaces and options to share final datasets in project spaces as private or public
ANALYSIS	Visualisation and playback of multimedia	No, but possible with filter plug-ins (requiring development). Claimed access to text, images, moving images, mpegs and data sets	Depends on modular media solution packs: images, video, documents, binaries.	Yes, visualisation capabilities are extensive, but requires development (uses Blacklight for tailored views of different types of media).	Unsure - likely possible, but limited
	Graphing of tabular data	Unsure	Yes - with implementation of solution packs	Yes, with development	Yes, built-in capability - Stata, SPSS,R,xls, CSV can be ingested and viewed with TwoRavens
	Geographic data analysis	Yes, but requires development	Yes - solution packs	Yes	Yes, built-in capability - Shapefiles viewed with WorldMap
	Images	Yes, but requires development	Yes	Yes	Yes
	Other	-	-	-	Astronomy FITS
	Access to data via API	No - download only	Yes REST API	Yes REST API	Yes
POLICY	Guidelines available through interface: (DMP, deposit, preservation, sharing)	Unknown	Unknown	Can be added to terms of use	Can be added to terms of use
	Automated acknowledgement of data policies: (preservation, RDM, open access, ownership, etc)	Unknown	Unknown	Unknown, but likely possible through development	Yes (data usage agreement), templates exist

Table 2. Comparison of open source repository software features

	Policies section on landing page	Yes, can be added with simple development	Yes, can be added with simple development	Yes, can be added with simple development	Yes, templates exist, to be determined in action
	Succession plan	Well supported by extensive community. Does not require large funding to be sustainable	Crowd/community-sourced funding model. Not currently secured and struggling to raise necessary funds.	MoU between partner institutions ensures long-term cooperation and sustainability	Well supported by Harvard and partner institutions.
	Origin of funding	Originally Hewlett Packard and MIT, but now funded by The Duraspace Community Sponsorship Programme.	FEDORA: Originally well funded by the Gordon and Betty Moore Foundation (Cornell U. and U. Virginia). Now supported by The Duraspace Community Sponsorship Programme. ISLANDORA: Islandora Claw development funded through incentivised, optional membership fees.	FEDORA: Originally well funded by the Gordon and Betty Moore Foundation (Cornell U. and U. Virginia). Now supported by The Duraspace Community Sponsorship Programme HYDRA: Not funded, institutional partner model is stable	Harvard, MIT, IQSS partnerships with extended community - long history of university funding and no threat to sustainability.
ARCHIVING	Author ID (orcid, scopus,DAI)	Yes, ORCID	Yes - ORCID	Yes, ORCID solution pack	No - ORCID integration is planned for future development
	Version control	Yes - new function and still being improved	Yes	Yes	Yes, data and metadata. No deletion

Table 2. Comparison of open source repository software features

	Citation and references	No, but can be integrated	Scholar module	No, but can be integrated. (Integration in Sufia for Endnote, Zotera, Mendeley)	EndNote XML, RIS Format, or BibTeX Format
PRESERVATION	Redundancy	No, but can be integrated	No, but can be integrated	No, but can be integrated	LOCKSS network
	Persistent IDs	Yes - CNRI handle system	Yes, with plugins	Yes, with plugins	Yes
	DOI capability	Yes - interoperability with EZID and Datacite API and conversion of DSpace metadata into Datacite format for DOI retrieval BUT no support for community/collection level DOIs	Yes- DataCite Community/collection level DOIs?	Yes- DataCite Community/collection level DOIs?	DataCite & EZID
	Persistent data deposit	Checksums with audit trails	Checksums with audit trails.	Checksums with audit trails	Universal Numeric Fingerprint
	Integration with Archivematica	Yes - ASpace	Yes, Archidora	Yes. Integration plugins being developed (Hull/York collaboration through Jisc funding)	In development, proof of concept stage
	LOCKSS compliance	Yes, integration option	Yes, integration option	Yes, integration option	Yes: + EZ-Proxy, DDI, Z39.50
	Content back-up	Yes - but issues with versioning depending on choice of back-up procedure	Yes	Yes	PostgresQSL database backup and study files. Glassfish configand customizations too
ADMIN/REPORTING	Tracking	Yes, SOLR	Yes, SOLR	Yes, SOLR	Yes, via guestbooks. Access requests too

Table 2. Comparison of open source repository software features

	Usage/download reports	Yes, SOLR & elastic search (Manakin)	Yes, SOLR	Yes, SOLR	Yes, IDs and custom questions. Sharing and reporting too.
	Stakeholder reports	Customised monthly reporting	not known	not known	Not automatically
	Author reports	Customised monthly reporting	not known	not known	Not automatically
	Altmetrics	Altmetrics badges can be embedded. collection stats are visible	not known	not known	In development
DISSEMINATION	search API	Yes - Lucene (Java)	Yes - GSearch?	Yes - GSearch?	Yes
	OAI-PMH compliance	Yes	Yes	Yes	Yes
	Publishing workflows	JSPUI & XMLUI have submission workflow including Author and Publisher roles	Yes (details not known)	Yes (details not known)	Yes: draft, review, published
	Faceted search	Yes	Yes	Yes	Yes
	Search engine optimisation	Yes, SOLR, Google, optimised for Google Scholar	Yes, SOLR	Yes, SOLR	Yes, SOLR
	Social media share options	Yes, may require development	not known	Yes (many options with Sufia: Facebook, Twitter etc)	Yes (Facebook, Twitter, Google Plus)
	Features page	Not sure	Yes, requires development	Yes, requires development (through customisation or with Sufia and also upcoming Hydra-in-aBox)	Display featured dataverses on main page

Table 2. Comparison of open source repository software features

	RSS	Yes	Yes	Yes	Yes
	Integration with discovery platforms	Yes, - OAI-PMH, Open Search, RSS, SWORD	Yes	Yes	Yes
	Integration with research profiles (ORCID)	Yes	Yes	Yes	Planned for development
DATA PUBLISHING WORKFLOW	Versioning of published datasets	Yes - but only backs up latest version	Yes (details not known)	Yes	Yes
	Communities and curators of data collections	Yes	Yes, requires development	Yes (highly customisable through development)	Yes - easy and extensive
	Deletion possible	Yes	Yes, CRUD (create, read, update, delete)	Yes, CRUD (create, read, update, delete)	No, only deaccession