White Rose Libraries Repositories Renewal Project

Public report: Jan 2025

White Rose Libraries

CONTENTS

EXECUTIVE SUMMARY

White Rose Libraries (WRL) has operated shared repository services since 2004. In 2018-19, WRL began a project to review and replace the repositories platform. Much work was done to set principles (see Appendix 1) and gather requirements but work was halted in early 2020 due to COVID-19. This restarted and the updated project, which featured in the WRL Strategic Plan 2023-2027, was initially expected to run 2022-2025, and would see WRL start 2026 on the next phase of infrastructure.

WRL reviewed and updated requirements and employed Ken Chad Consulting to streamline these and make them fit to take to market. A final set of requirements was agreed by end 2024 (see Appendix 2). WRL Exec extended the project timeline to include market testing, and also set a budget ceiling for the new platform. The new timeline would now see implementation by Dec 2026.

The market testing took the form of a Prior Information Notice (PIN), supported by Leeds procurement. This closed at the end of April 2024. WRL received four PIN submissions, covering commercial and hosted open source options. The Repositories Board did consider why engagement was lower than hoped and planned to address this in any future procurement exercise.

The indicative costs for a commercial solution were beyond the budget cap suggested by WRL Exec, indicating that commercial solutions would be out of scope for any procurement exercise. Hosted and Self-hosted open source options were financially viable options. Self-hosted costs were less than for the hosted model, but that could represent good value if the platform gave increased functionality. The Repositories Board looked to identify the most desirable open source platform, therefore, and would consider the hosting model once the preferred platform was chosen.

The main options still remain DSpace and Eprints. DSpace has the largest global user base, EPrints the largest in the UK. The EPrints position is stronger than perhaps the sector realises. Engagement with the platform continues and support is ongoing, and with a new release imminent. WRL flagged concerns about the EPrints governance model and community engagement and, in direct response to this feedback, a Steering Group is being formed. The release model is less transparent than that of DSpace, masking that the same functionality is offered. With DSpace 8, DSpace plans to implement a pattern of annual releases, alongside the regular meetings of its community-led governance groups.

The DSpace community is international with significant user numbers in North and South America, and fewer users in Europe and the UK. WRL would be looking to move to DSpace 8, though it's not 100% clear what user functionality that will offer yet. From a technical delivery perspective, a reasonable developer would need 2+ years to be familiar enough with Java to properly support and develop a self-hosted DSpace instance. This might indicate an extra WRL development resource would be needed. The DSpace ‘support’ community has limited capacity at the moment and this would impact on a potential move to DSpace by a newcomer to the platform.

The Repositories Board conclusion is that DSpace does not currently offer functionality that justifies moving to it from EPrints. The recommendation is to remain with EPrints at this time. Given this, the Board also felt that there were not significant drivers to move to a hosted instance. WRL should focus on developing the self hosted instances to maximise the opportunities offered by newer versions of the software and through e.g. moving to cloud hosting.

The recommendations in full are:

Platform and model

- WRL should commit to using the open source EPrints platform as the shared WRL repository infrastructure until after the next REF (so until 2030).
- WRL should continue with the self-hosted model, and restart all the upgrade and improvement work previously paused (including identifying the best Cloud solution to move to, and implementing this).

Other recommendations

- WRL should be clear and open about this decision, reinforcing that this is not accepting the status quo, but a researched and informed conclusion to a structured process.
- WRL should implement a structured approach to upgrading the EPrints platform as new version releases are available.
- WRL should take the opportunity to engage with EPrints (e.g. as part of the steering group etc.) and contribute to making this the best solution for the specified period, in strategic as well as technical ways.
- WRL should use the period to 2030 to engage in a deeper way with other potential solutions, to fully understand current and emerging options.
- During the same period, WRL should increase strategic engagement with the repositories sector, to drive the conversation around next generation repositories and work with partners if appropriate to deliver progress on these.
- From 2027, WRL should actively review its repository provision, and position across the sector, and begin the process to secure the next solution to take the repositories forward, with the expectation to implement directly after the next REF (approx 2030).

WRL Exec accepted these recommendations, and this effectively ended the exploration phase of the project. The focus turned immediately to updating the EPrints software and our support infrastructure. That includes the planned move to the Cloud, though this may be a future phase of the project.

INTRODUCTION

White Rose Libraries (WRL) has operated shared repository services since 2004. These consist of White Rose Research Online (WRRO) and White Rose Etheses Online (WREO). Since inception, these repositories have used the EPrints platform, and the operational and developmental support is provided by 1 FTE staff resource. This resource is WRL funded but embedded within Leeds IT Services.

In 2018-19, WRL began a project to review and replace the repositories platform. This was driven by the beliefs that:

- EPrints, as an older platform, was at risk of losing central support, and would soon see a shrinking community of users;
- EPrints lacked desired (though not identified) functionality
- There would likely be newer platforms, either available or emerging, that would offer significantly more functionality than EPrints.

Much work was done to set principles (see Appendix 1) and gather requirements for the next phase of repository infrastructure and, by the end of 2019, WRL were positioned to use these to explore the options on the market. Work was halted in early 2020 due to COVID-19.

The Repositories Renewal Project was revisited in 2021 and it was agreed to restart this work, reconfirming and building on what had been done previously, but also acknowledging the passage of time and anticipating the provision on offer would hopefully have evolved. This updated project, which featured in the WRL Strategic Plan 2023-2027, was initially expected to run 2022-2025, and would see WRL start 2026 on the next phase of infrastructure.

REPOSITORIES RENEWAL PROJECT

The restarted Repositories Renewal Project was split into 3 phases of work. Initially these were:

- 2022-2023: scoping and confirming requirements;
- 2024: identifying and engaging the next platform;
- 2025: implementation and transfer to the new platform, to be fully complete by Dec 2025.

Scoping and confirming requirements completed

WRL took some time to revisit the previous list of requirements, reviewing and updating these, checking they were still fit for purpose. Towards the end of this phase, WRL employed Ken Chad Consulting, with the aim of getting external input to the project, focussed on streamlining the requirements and making them fit to take to market. WRL also hoped that this consultancy work would provide an overview of the current and emerging options, and a blueprint to adopt the new infrastructure chosen.
The consultancy outputs fed into the WRL Repository Board work, and a final set of requirements (see Appendix 2) was agreed by end 2023, as per the project timeline.

Identifying and engaging the next platform: Change to Project Timeline

The final requirements were shared with the WRL Exec Board in Jan 2024, with the request for approval to move to the “procurement” phase. On reflection, WRL Exec felt there were still many unknowns about the current market, in terms of costs, value, and the different functionality different models of system (commercial, hosted open source, self hosted open source) would bring. This made formal procurement a complicated and potentially problematic exercise that could lead to a non-optimal outcome if not a fully informed process.

To counter this, WRL Exec extended the timeline of this phase of the project to include an element of market testing. This would provide more insight into the options available and cost/value these offered. Exec also set a shared budget ceiling for the new platform, shared across the three sites.

This also led to a revised and extended timeline for the project:

- 2022-2023: scoping and confirming requirements;
- Jan 2024 - July 2025: identifying and securing the next platform;
    - Jan-July 2024: Market testing/PIN; recommendations and decision on next steps based on this;
    - Aug 2024-July 2025: identifying and securing the next platform, potentially via formal procurement;
- Aug 2025 - Dec 2026: implementation and transfer to the new platform, to be fully complete by Dec 2026.

PIN engagement and responses

The Repositories Board took the market testing forward in the form of a Prior Information Notice (PIN), supported by Leeds procurement. The PIN process was undertaken March/April 2024, and Leeds Procurement shared the PIN directly to suppliers WRL flagged and also via the usual Government channel. (It's worth noting that, however, that after the PIN submission period closed, we were contacted by suppliers on our list who had not seen this announcement). The PIN was live for about a month, during which time the Repositories Board held a briefing for interested parties giving details of the project and the WRL repository services, and offering the chance to ask questions. Suppliers could then submit answers to the PIN questionnaire, which had been drafted - again with the support of Leeds Procurement - to gather relevant information.

The PIN closed at the end of April 2024, and the Repositories Board met in May to discuss the submissions.

WRL received four PIN submissions. These covered commercial and hosted open source options. Two of the four were discounted by the Repositories Board, one due to the approach taken and one due to the supplier's lack of experience in the area. The remaining two gave strong examples of a commercial solution and of a hosted solution. The Repositories Board did consider why engagement was lower than hoped. The potential reasons considered included:

- Not seeing the PIN - if going for procurement, WRL would need to play a bigger role in dissemination;
- Not having capacity to engage with this stage, given the PIN stage has no impact on any formal procurement;
- Not having the interest to engage with this stage, given the PIN stage has no impact on any formal procurement;
- Not wanting to engage with the project at all (due to e.g. capacity, suitability, complexity of collaboration).

The information provided in the PIN questionnaires was useful, and the Repositories Board focused primarily on cost as a way to streamline discussions.

Commercial option

The costs quoted were beyond the budget allowed by WRL Exec. The Repositories Board took this to be indicative of a commercial solution costings, and so felt that **commercial solutions would be out of budget, and so should be out of scope for any procurement exercise.**

Hosted and self hosted open source options

For both these options, costs were within the budget limits outlined by WRL Exec, making them financially viable. The hosted costs were higher than remaining on a self-hosted model. However, these could represent good value if the platform gave increased functionality. The Repositories Board, therefore, explored the potential benefit in making a change to other open source infrastructure. We looked to identify the most desirable open source solution first, as

we felt that the hosting model should be a follow up consideration to identifying the preferred platform. As part of this, members of the Repositories Board attended a session on repository platforms offered by UK-CORR, engaged directly with institutions using other open source solutions, and WRL colleagues attended OR2024. This fed into the following set of conclusions.

Platform options:
It became clear that the main two open source platforms in terms of use remain DSpace and EPrints. Both remain viable, and both are being extended, and both integrate with the CRIS in use at Leeds, Sheffield and York. DSpace has the largest number of users globally of any platform but, in the UK, EPrints still has a significantly large user base. A handful of other platforms have small numbers of UK users (Samvera, Haplo etc.) but these lack the community we are looking for, and which would be important in moving to a new platform (See Appendix 3). WRL also checked the platforms against our list of requirements, and - while some areas would need greater investigation, - both platforms seemed satisfactory.

The Repositories Board therefore focused on DSpace and EPrints as the best viable open source options for the next phase of WRL repositories moving forward.

EPrints

Hosted EPrints providers did not engage with the PIN, and WRL discovered this was because they had not received the information about the process either direct from Leeds (despite being on the dissemination list) or through seeing it on the government site.

WRL met with other EPrints users and with CoSector (an EPrints provider), and also discussed the situation with them at length at OR2024. The EPrints position is stronger than perhaps the sector realises. There has not been a significant move away from EPrints, which remains particularly strong in the UK. There is continued engagement with EPrints, and the support for the platform is ongoing, with a new release imminent. WRL also spoke to other institutional EPrints users. Informal discussions indicated that a key driver where a site is considering moving away from EPrints is often a lack of local support/development resources, and not due to dissatisfaction with the EPrints platform itself.

At OR2024, there were discussions about the EPrints governance model and community engagement, specifically as compared to the DSpace approach (detailed below). EPrints Services and Co-Sector (who both support the platform development and offer hosted EPrints options) agreed current structures needed review and, in direct response to this feedback, a Steering Group is being formed. The release model is not as transparent as the DSpace release model, and this can mask that the same functionality is offered, even though this is present. Revitalising the governance and community comms would be beneficial to the user community, which remains strong but is possibly less active than in the past.

DSpace

OR2024 was really useful in giving a detailed overview of DSpace, in terms of global use, governance and technical approach, as many sessions were focused on DSpace. WRL colleagues also spoke to UK-based DSpace users about their experiences. The community is international with significant user numbers in North and South America, and fewer users in Europe and the UK. This community is now focussed on DSpace 7 and later. DSpace 7 saw a change in the underlying system structure, which makes the move from earlier DSpace versions to DSpace 7 a significant undertaking. This doesn’t bring any real change in functionality for users, however, and that is expected to come with DSpace 8, due imminently. If WRL would move to DSpace now, this would likely mean looking at DSpace 8, though it's not 100% clear what user functionality that will offer yet. DSpace 8 could offer support for student journals and

diamond publishing, for example, but these are not significant requirements for WRUP. With DSpace 8, DSpace plans to implement a pattern of annual releases, alongside the regular meetings of its community-led governance groups.

From a technical delivery perspective, advice at OR2024 on implementation suggested that a reasonable developer without current in-depth knowledge of Java would need 2+ years fully devoted to this to become familiar enough to properly support and develop a self-hosted DSpace instance. This might indicate an extra development resource would be needed, in the short term at least, to cover implementation and the bedding in period. There was also some discussion about whether the community supply supported the direction of some aspects of the DSpace core development. From an external perspective, it's hard to judge if these are well-founded concerns or down to individual views/preferences.

A key consideration is that the DSpace 'support' community has limited capacity at the moment and, with a major release due, this would only get worse in the short/medium term. This would impact on a potential move to DSpace by a newcomer to the platform. There were also indicators that DSpace is resource intensive on the servers/databases it runs on, which could see higher costs to run in the cloud.

On the whole, DSpace is not an unattractive option. There are pros in terms of renewed infrastructure and well-structured community and governance etc. There are also challenges for WRL to move to DSpace within the current project timeframe. Given this, the Repositories Board conclusion is that DSpace does not currently offer functionality that justifies moving to it from EPrints. The recommendation is to remain with EPrints at this time. This is not through a desire to stay with the status quo, but a genuine belief that EPrints is the best choice for us at this time. To confirm this, WRL has also sanity checked EPrints against the requirements generated by this project, to ensure that these were met.

Having decided which platform to recommend, the Repositories Board then considered the hosting model. Given the recommendation to remain with EPrints, the Board also felt that there were not significant drivers to move to a hosted instance, and WRL should focus on developing the self hosted instances to maximise the opportunities offered by newer versions of the software and through e.g. moving to cloud hosting.

The Repositories Board also made other recommendations as to how to move forward if these recommendations were supported by WRL Exec.

RECOMMENDATIONS

After full consideration of the options identified through this project, from WRL own investigation and through the work of Ken Chad Consulting, the Repositories Board have reached the following recommendations for WRL Exec to consider.

Platform and model

- WRL should commit to using the open source EPrints platform as the shared WRL repository infrastructure until after the next REF (so until 2030).
- WRL should continue with the self-hosted model, and restart all the upgrade and improvement work previously paused (including identifying the best Cloud solution to move to, and implementing this).

Other recommendations

- WRL should be clear and open about this decision, reinforcing that this is not accepting the status quo, but a researched and informed conclusion to a structured process.
- WRL should implement a structured approach to upgrading the EPrints platform as new version releases are available.
- WRL should take the opportunity to engage with Eprints (e.g. as part of the steering group etc.) and contribute to making this the best solution for the specified period, in strategic as well as technical ways.
- WRL should use the period to 2030 to engage in a deeper way with other potential solutions, to fully understand current and emerging options.
- During the same period, WRL should increase strategic engagement with the repositories sector, to drive the conversation around next generation repositories and work with partners if appropriate to deliver progress on these.
- From 2027, WRL should actively review its repository provision, and position across the sector, and begin the process to secure the next solution to take the repositories forward, with the expectation to implement directly after the next REF (approx 2030).

Acting on these recommendations would best position WRL in terms of current provision, but also enable WRL to lead future discussions from a strengthened position.

NEXT STEPS

Outcome:

WRL Exec accepted the recommendations. The focus therefore moved to the second phase of the Repositories Renewal Project. This was updated to:

- 2022-2023: scoping and confirming requirements;
- Jan 2024 - July 2025: identifying and securing the next platform;
    - Jan-July 2024: Market testing/PIN; recommendations and decision on next steps based on this;
    - Aug 2024-July 2025: identifying and securing the best Cloud platform to support an updated EPrints platform potentially via formal procurement;
- Aug 2024 - Dec 2025: implementation and transfer to the new platform, to be fully complete by Dec 2025.
- Dec 2025 - implement a structured approach to updating the platform and maximising the benefit of new releases.

This would effectively end the exploration phase of the project so the focus would turn immediately to updating the EPrints software and our support infrastructure. That would include the planned move to the Cloud. Timelines around this element may change further.

APPENDIX 1: PRINCIPLES OF THE WHITE ROSE REPOSITORIES (2022)

The WRL Repositories Board, working with colleagues across the three White Rose institutions, outlined a set of Principles as part of the Repositories Renewal Project. These were felt to be at a level higher than the system requirements, and shape the underlying WRL approach to repository provision:

1. The White Rose Libraries have agreed to continue to take a consortial approach to repository provision for research and etheses, while reviewing the detail and level of this shared approach. Working as a collaboration brings many benefits: working with a larger data set in terms of system service development; a more significant voice in the national debate; shared costs and resources; and a local community of practitioners.
2. Whilst evaluating all potential solutions, preference should be given to open source solutions. This fits with our desire for the contents of the repositories to be open where possible.
3. The repository should use PIDs - Persistent Identifiers (with provenance) - wherever these are available to increase the interoperability of the repository data and to ensure we align with current sector standards and best practice
4. The repository content and metadata should be “Open by default, and restricted only as necessary”.
5. The repository should support ingestion/deposit, storage and delivery of all file types and relevant metadata.
6. The repository should be compliant with major funder requirements, evolving as these evolve.
7. The solution reached should enhance the reputation of WRL in the repository space, building on current strengths and recognitions.
8. The solution should be sustainable in terms of: technology, community, financing, staff resourcing, and sustainability of the files and metadata for long term reuse. It is anticipated that the solution needs to be sustainable for 10 years, but will be initially reviewed 5 years after implementation to assess the ongoing suitability of the solution from that point.
9. The repository must be indexable, harvestable and aggregated by the principal global discovery tools such as Google Scholar.
10. The repository should be able to integrate with relevant institutional systems and support the information security policies of the institutions.

It was noted that the WRL repository solution is not intended to be a CRIS, but must integrate with each Institution’s CRIS, as implied by Principle 10. The repository solution needs to be focused on discovery and deposit. The repository needs to be able to handle research data but is not envisaged as the primary data repository and is not expected to handle “big” or commercial data sets.

APPENDIX 2: REQUIREMENTS FOR THE WRL REPOSITORIES (JAN 2024)

Stand-alone system

NB: CRIS is a current research information system

Please confirm the solution can be supplied as a stand-alone product and does not require purchase of any other system (such as a CRIS).

General solution characteristics

We require:

A solution partner with relevant expertise, experience and reputation in repository solutions for research outputs including theses and data

a) How many UK Higher Education client Institutions currently use your solution as their institutional level research repository? Give up to 5 relevant examples and implementation dates.

b) How many Higher Education client Institutions outside the UK currently use your solution as their institutional level research repository? Give up to 5 relevant examples and implementation dates.

c) How many Higher Education Institutions use your solution as a consortium shared system across 3 or more institutions?

d) Give examples of when mature institutional repositories have migrated into your solution?

e) Do any of your clients in Higher Education Institutions use your solution simultaneously with multiple, different CRIS systems? Please list the institutions and the CRIS systems used.

f) Do any of your clients qualify as a trusted repository (e.g. complying with or accredited by CoreTrustSeal).

A transparent system development programme/roadmap, shared with, and influenced by, clients

How does the solution:

a) Share the roadmap with clients and how regularly?

b) Engage with UK/global client communities to:
   - prioritise developments accordingly?
   - accommodate individual client development requirements?
   - provide and sustain strong developer and/or user communities to share best practice, code, processes etc.

c) Develop to deliver essential functionality so the UK/global client communities can:
   - Embrace emerging system and technology developments such as AI?
   - Meet the requirements set out by the Coalition S 'PlanS' and COAR's 'Next Generation Repository' recommendations and those from any other relevant new initiatives and directives?
   - Satisfy the evolving requirements of the 2028 Research Excellence Framework (REF)[1] ?
   - Satisfy changing UK/national legal and policy compliance requirements?

[1] The REF 2028 is still being discussed but it is already clear that 'the work of all researchers and research-enabling staff will be eligible for submission.

A documented adoption and exit strategy

a) Confirm that an exit strategy is included in any contract for the consortium and for individual institutions
b) What support and mechanism is provided to enable all metadata, system data, and content to be easily imported?
c) What support and mechanism is provided to enable all metadata, system data, and content to be easily extracted?
d) Confirm that the contract will ensure rights and ownership of data and content remain with the depositing institution.

Overall Technology and accessibility

We require:

A solution that will work securely across a wide-ranging IT environment including a variety of browsers, desktop and mobile devices.

a) Does the solution provide a simple, accessible web interface in compliance with current web accessibility standards, e.g. Web Content Accessibility Guidelines (WCAG) 2.1 AA standard?
b) Are all solution functions available via a range of web browsers and which ones do you currently support?
c) How does the solution evolve to keep up to date with web standards and browser developments?
d) Do all solution functions work via browsers on desktop, mobile devices, and tablets?
e) Describe support for open methods of authentication for machine and human access (e.g. UKAMF, OpenAthens, Active Directory, OAuth, OpenID, SAML).
f) Please provide details of technical capacity (e.g. concurrent usage) and any relevant standards your platform meets.
g) How can the solution interface be customised to meet branding requirements?

Technical provision of the solution

a) Does the solution allow multiple bespoke and configurable workflows to support different institutional requirements for different content types?
b) What are the options for curating the content as distinct collections?
c) What are the options for sandbox/test environments?
d) Does the solution allow multiple configurable workflows based on content source (e.g. CRIS, manual entry)?
e) What are the hosting options for system and storage e.g. SaaS; customer-hosted? If vendor hosted, where geographically are the solution (system and storage) hosted?
f) Is it possible to link to storage that is external to the system?
g) What level of solution/system management is provided, e.g. software updates, protection against hacking etc.
h) How is the software updated and how often?
i) How can the solution be configured and what are the configuration options?
j) Supports authentication of repository user accounts via multiple institutional identity management systems and ORCiD.

A solution that meets security and data governance standards

a) Are you Cyber Essentials or ISO27001 compliant prior to commencement of work and be able to provide copies of certification?
b) Is the solution UK GDPR compliant?

Support
We require:

Continuous high level availability

a) If vendor hosted:
    - Describe how the solution provides high levels of reliability and availability?
    - What level of uptime, excluding scheduled downtime for maintenance or upgrades, is guaranteed in the service level agreement (SLA)? (Plan S requirement is 99.7% uptime, excluding scheduled maintenance)
    - What is the schedule for maintenance or upgrades? How is this managed?

Support for clients using the solution

- Describe your helpdesk provision:
    - What hours is helpdesk provision available?
    - What is the response time to tickets/queries stated in the SLA?
    - What channels is the help desk available via?
- Describe the "account manager" support provided to clients.
- Describe the support and development resource (FTE) provided to support the solution post implementation.

Implementation

a) Please provide an outline project plan with estimated timings, including the migration of data and content to the new solution, including system setup/configuration and training.
b) What project team support (roles and FTE) will be provided for implementation?
c) What are the expected resources required from White Rose Libraries to support implementation?

Test/QA system

a) What options are there for synchronisation of data from the live system to a test/QA instance?

Curation of content including workflows
We require:

A solution that enables efficient workflows for staff and end users
Describe how the system

a) Enables real time and customisable mechanisms for direct user deposit/input of research outputs from the relevant CRIS.
b) Supports the requirement for a human depositor to read and confirm a declaration.
c) Supports staff mediation of content deposited via CRIS system or direct researcher deposit.
d) Supports the disambiguation of individuals e.g. authors/creators.
e) Enables multiple levels of staff access to particular tasks, content and functions. For example to restrict editorial access to only that content from the staff member's institution.

f) Provides configurable and editable help text giving contextual support throughout workflows within the system.
g) Allows for multiple workflow configurations based on institution, user type, and content type.
h) Provides automated ingest methods using protocols e.g. REST API, SWORD, FTP.
i) Provides automated updates, reports and alerts to support staff workflows and processes.
j) Ensures direct depositors can only upload institution approved file types for that content workflow.
k) Provides audit logs for any changes made to system records and content.
l) Supports a variety of access restrictions to records and content including, temporary embargoes and permanent restrictions.
m) Supports the ability to apply different copyright and licence statements to different files on the same record.
n) Provides the ability to completely hide a record and attached content - A 'dark archive'.
o) Provides the ability to suppress records and content easily and quickly, either item by item or in bulk by e.g. institution or department.
p) Provides the ability to have multiple abstracts, e.g. a public abstract**,** alongside a full abstract which would be suppressed.
q) Provides the ability for a user to request updates to a record.
r) Provides a means to automatically generate (e.g. by AI) a lay summary of the content.
s) Enables links to other works by an author held in the repository and external systems
t) Enables user accounts to be updated to another authentication method
u) Handles customisable coversheets specific to institution, item type, licence etc
v) Enables authors to grant requests for access to embargoed content. ('request-a-copy').

Metadata, PIDs and file formats
We require:

Metadata

We require:

a) Support for open and interoperable metadata.
b) The ability to have metadata only records.
c) Support for established and emerging metadata schemas and formats. Please list those your solution currently supports.
d) The ability to support different object types (for example journal articles, books and book chapters, datasets, theses, analysis code, multimedia and non-traditional outputs).
e) Support for additional metadata types and schemas be defined/customised by the customer.
f) Support for records and items in different languages and non-roman scripts.
g) A solution that facilitates the use of authorities pulled in from other sources e.g. FAST (Faceted Application of Subject Terminology)
h) Metadata to be accessible to overlay journals and Text and Data Mining services.

Does the solution support rich metadata e.g. abstract including a subset of HTML elements?

Persistent Identifiers (PIDs)

a) How does the solution support and integrate with PIDs (e.g. (DOI, ORCiD) for a person, item, funder, institution, dataset etc.?
b) How does the solution 'mint' a persistent identifier (including DOIs) for a research output?

c) Does the landing page contain mandatory metadata elements required by the minting service (DataCite: title, author, date, publisher, DOI, item type)?
d) Does the solution provide a 'tombstone page' as resolution for a DOI if content/record has to be withdrawn/suppressed?

Support for a wide variety of file formats both in terms of ingest (automated and direct deposit) and making files available for access

a) Does your solution limit file deposit based on e.g. filesize, format? Is there specific consideration for very large datasets?
b) Is there a maximum number of files per record?
c) Can the solution generate and validate checksums for files during deposit and after deposit?
d) What file types can be viewed and previewed in browsers? Do some file types allow additional functionality e.g. thumbnails, full-text indexing, audio player, IIIF?
e) How does the system support 'complex' objects with multiple files potentially in different file formats - e.g. music, during deposit and re-use?
f) What are the system implications of managing large datasets comprising multiple files?
g) How does the solution support versions of documents in different language and non-roman scripts?
h) Does the system perform virus scanning on content?

Interoperability: Please provide an overview of the interoperability capabilities of your solution. Describe:

a) What open APIs does the solution support to allow access to metadata and content?
b) The ability to combine data from multiple different CRISs or systems for ingest.
c) What other systems does the solution have integrations with?
d) How does the solution support integrations with external systems, e.g. PubMed LinkOut, EThOS, Google Scholar, Google Datasets, CORE, IRUS.
e) How 'how-to-cite' data is exported to systems such as Refworks, Endnote, Reading List tools etc.
f) Does the system support CSL (Citation Style Language)?
g) What metadata 'crosswalks' are provided to map between systems?
h) How metadata is transferred and validated between systems?
i) How the solution links to associated internal and external content?
j) The ability for the repository UI to be enhanced with external data and services e.g. Altmetrics, CORE Recommender, Dimensions, ORCID, CORE Discovery
k) How metadata can be exportable in RDF triples or similar format, if supported.
l) The digital preservation systems you have integrations with. We require integrations with the different digital preservation systems used at each institution.

Reporting, staff enquiry and search.
The ability to provide rich, customisable enquiry and reporting to met both institutional and external requirements. This includes ad-hoc online enquiries of the repository by staff and end user searches.

Describe:

a) The overall approach to reporting and the key features. How are these defined, generated and

disseminated? Can these be scheduled? Is there a client dashboard?

b) The public search capabilities of the system.
c) How the solution enables immediate querying of internal data by repository staff, for example when an embargo date was lifted, and “in process” content.
d) How the solution provides usage statistics at item, unit, person, collection, institutional level etc. and whether this is COUNTER compliant.
e) How the solution supports compliance and reporting on funder mandates, automatically (such as through OpenAIRE validation) and through customisable reports, and auto completion of existing templates, such as the JISC APC data collection, Research Excellence Framework 2028.
f) The search capabilities for staff and end users that provides:
    1) Simple and advanced search capabilities for all metadata fields
    2) The ability to browse for items via author name, academic unit and date
    3) Deep searching of the full text
    4) The ability save search results with links, and create email alerts for updates.
    5) The ability to export search results with links.
g) Does the system allow ‘expert’ querying of data e.g. via SQL

Communications
We require:

a) A fully customisable range of system delivered and ‘contact us form’ emails with the facility for these to vary between institutions in terms of content, branding, workflow, frequency.
b) Ability to alert users of system maintenance.

APPENDIX 3: WHO IS CURRENTLY USING WHICH OPEN SOURCE PLATFORMS IN THE UK (Correct June 2024)

Platform(s)	Count	Universities
Eprints (with Pure as CRIS)	21	Aston University Lancaster University London School of Economics Teesside University Ulster University University of Bath University of Birmingham University of Bristol University of Dundee University of East Anglia University of Essex University of Huddersfield University of Manchester University of Northampton University of Portsmouth University of Southampton University of Strathclyde University of the West of Scotland University of Winchester University of Worcester University of York
Eprints (with Symplectic Elements as CRIS)	20	Anglia Ruskin University Bournemouth University City, University of London Institute of Cancer Research Keele University Leeds Beckett Liverpool John Moores University London School of Hygiene & Tropical Medicine Loughborough Manchester Metropolitan Natural History Museum Sheffield Hallam St George's University of London University College London University of Leeds University of Liverpool University of Sheffield University of Sussex University of the Arts London University of Wolverhampton
Eprints (other)	29	AECC University College Association for Learning Technology Bath Spa University Birkbeck British Antarctic Survey

		British Geological Survery British Geological Survey British Library Cardiff University Falmouth University Glasgow School of Art Goldsmiths, University of London Harper Adams University Kingston University Open University St Mary's University, Twickenham University for the Creative Arts University of Bradford University of Central Lancashire University of Glasgow University of Greenwich University of Kent University of Lincoln University of Nottingham University of Reading University of Stirling University of the West of England University of Warwick University of West London
DSpace (with Pure as CRIS)	10	University of Aberdeen Bangor University Coventry Edge Hill University Glasgow Caledonian University Heriot-Watt University University of Edinburgh University of Hertfordshire University of St Andrews Royal Holloway University of London
DSpace (with Symplectic Elements as CRIS)	6	Brunel University London Imperial College London Queen Mary University of London University of Cambridge University of Plymouth University of Exeter
DSpace (other)	10	De Montfort University Sussex University Queen Margaret University, Edinburgh Rothamsted Research Royal College of Music London Science and Technology Facilities Council University of Bedfordshire University of Chester University of Chichester University of Cumbria

Haplo	6	Institute of Development Studies London South Bank University University of East London University of Westminster Canterbury Christ Church Middlesex University
Fedora / Hyrax / Hyku (Samvera)	2	University of Oxford Leeds Arts University
VuFind	2	Swansea University University of Roehampton
OpenEquella	1	Oxford Brookes University