On the verge of success – or failure?
Repositories and the wider knowledge infrastructure, plus a bit about Hyku

Dr Torsten Reimer
Head of Research Services
Torsten.Reimer@bl.uk / @torstenreimer
http://orcid.org/0000-0001-8357-9422

Samvera Connect keynote (online), 23 October 2020
Previously: “buy it and they’ll come”

“other” content

library collection

LIBRARY/KNOWLEDGE INSTITUTION
Now: shrinking relevance?

Relative to Internet content our collections are shrinking

We can no longer hope to meet all user needs from our collections

Need to enhance our collections by connecting to relevant global content

Research is digital, are we?

Do we retain a role in discovery?
Survey of non-users about BL services

- **Online Catalogues**
  - 37% Happy with other sources
  - 33% Never considered it
  - 27% Didn’t know it existed

- **Reading Rooms**
  - 45% Too far away
  - 32% Never considered it
  - 30% Happy with other sources

- **Interlibrary loan**
  - 35% Didn’t know it existed
  - 31% Never considered it
  - 30% Too far away

- **Business & IP Centre**
  - 37% Didn’t know it existed
  - 33% Never considered it
  - 30% Too far away

- Online is first port of call for the majority, Google most common starting point
- Once beyond Google, there was an inherent trust in information from specialist websites
- Some are believing their area of interest unlikely to generate ‘untrustworthy’ online information
- Offline search, i.e. going to libraries, begins when online sources fail to meet research need
The repository in this (not so) new world

• An answer to the challenges could be to move from a collections to an open science service paradigm.

• Other industries are going down a similar route, by focusing on value-added services.

• For open access/science/research there is such a model already: move from acquiring (closed) content to enabling users to share research openly.

• The key infrastructure for this is the repository.
Repositories today
The success story of OA compliance?
Do readers, do authors really care?
Yet this is what we do…

Spare A Bullet by Wayne Stadler
https://www.flickr.com/photos/waynerd/6677201937/ CC BY NC ND 2.0
UUK repositories working group

- **Universities UK Open Access Coordination Group** with different stakeholders (government, funders, research organisations, publishers, libraries)

- A repositories working group set up to advise UK government.

- Series of meetings, a workshop at British Library and a survey of UK repository managers.

Challenges, (not just) for the UK

1. Concerns about sustainability of the underlying repository software package
2. Difficulties with CRIS system integration
3. Difficulties with integration with university systems (other than CRIS)
4. Difficulties with maintaining custom functionality
5. Issues with changing publisher and/or funder policies changing compliance status of articles
6. Lack of integration with identifiers (such as ORCID or DOIs)
7. Limitation of reuse through deposit licence (‘all rights reserved’), e.g. for text and data mining
8. Limited/no facilities (such as API) to support text and data mining
9. Linking publications to related datasets (and vice versa)
10. Linking publications to relevant funders
11. Management effort for journal embargoes
12. No or limited preservation functionality
13. Not enough resource to update from older/out-of-date version of repository software
14. Not enough staff resource for operational management
15. Reporting facilities not sufficient for funder reporting
16. Reporting facilities not sufficient for internal reporting
17. Technical support: lack of skills / capability
18. Technical support: not enough capacity
19. Tracking/integrating AAMs deposited in subject/other institutional repositories (REF OA policy)
20. Usability and user interface issues

https://doi.org/10.5281/zenodo.1136075
Selected WG recommendations

• [Recommendations on metadata, persistent identifiers etc.]

• HEIs, Jisc, subject repositories and other stakeholders take forward as a high priority improvements in the user experience.

• A study into the feasibility of a national preservation solution be undertaken, recognising that the British Library and Jisc are key stakeholders.

• A study be conducted to explore the need for national repository solutions or ‘hubs’ for one or all of the big challenges – discoverability, sustainability and preservation.
One system to rule them all?

• A single repository platform sounds tempting.

• However:
  – Can one monolith meet all needs?
  – Competition has benefits!
  – It is getting easier to exchange data between systems.
  – Multi-tenancy is hard.
  – Resilience vs. single point of failure…
  – Local resistance to giving up in-house system.

• With a government mandate unlikely and no obvious technology solution a gradual move to more shared services seems a more likely solution. If they are good enough, we might still get there.
The bit about Hyku: shared repository service at the BL
The British Library renews itself

• Living Knowledge articulates the vision of the British Library in 2023 as the most open, creative and innovative institution of its kind in the world.


• Everything Available is a strategic change management portfolio designed to transform the Library’s services to researchers and research organisations, on-site and online.
Developing a repository platform

- Consolidate BL repositories on a single platform
- Refresh preservation system for national collection (>5m items, petabyte-scale)
- Access layer with multiple repositories, shared service model
- Shared repository developed with partners
Shared Research Repository: approach

• Need for institutional repository and new home for data.bl.uk and EThOS, both sitting on outdated or inadequate platforms.

• Partner organisations interested to join shared service effort.

• Following a procurement process for a 15-month pilot project, we appointed Ubiquity Press as our technical partners.

• Selected Samvera Hyku, an open source, turnkey repository application that is inherently multi-tenancy:
  • “Based on Hyrax, Hyku is intended to be an easy-to-install, easy-to-configure, feature-rich multi-tenancy solution. Hyku is designed to reduce barriers to effective asset management and preservation for collections and content types of many kinds.” [Samvera Hyku Interest Group](#)

• BL and UP project team working in sprints, with partner feedback.
Shared Research Repository: pilot partners

National Museums Scotland
Tate
British Museum
Museum of London Archaeology (MOLA)
Royal Botanic Gardens, Kew
Shared Research Repository

Welcome to our pilot research repository where you'll find research produced by staff of the British Library and our current partner organisations: the British Museum, MOLA (Museum of London Archaeology), National Museums Scotland, Royal Botanic Gardens Kew, and Tate.

Begin your search

Browse our repositories:

MOLA - TATE - National Museums Scotland - The British Museum - Kew
BL institutional repository
Next steps for BL repositories

• Finish procurement for next phase of hosting and development.

• Digital preservation capability for the shared repository service.

• Current shared repository partners are UK cultural sector bodies, but other sectors, organisations and countries are welcome.

• Options for standalone single repositories, and for other shared repository clusters, are also being explored – please do get in touch if you are interested.

• We are also happy to speak with individual organisations about developing shared services – get in touch! (openaccess@bl.uk)
Lessons learned: the pilot / shared service

• Working with a young team on both BL and UP sides was exciting but getting through puberty to maturity is always a challenge.

• The definition of multi-tenancy and shared service is critical for shaping success, as a clear vision, drive and planning.

• A business model that ensures sustainability and affordability is hard, especially working with partners with a small budget.

• Finding partners with similar challenges helps – in our case the PALNI/PALCI consortia in the US, who are also developing Hyku.

• The shared service means different things to different stakeholders and it has to deliver a lot to make everyone happy!
“Turnkey”? Enhancements made

- 11 new metadata work types and 55 resource types
- Visually appealing re-design
- DataCite minting integration
- DOI look-up & auto-population of fields
- ORCID & ISNI name identifiers
- ROR, Wikidata and GRID organisational identifiers
- Google Analytics
- Crossref Funder Registry lookup (searches 21k funders!)
- Universal Viewer
- File-level metadata & visibility
- Filters including file availability
Lessons learned: Hyku is growing up

- Hyku allowed us to achieve what we set out to do – it’s a promising product
- The Hyku community is growing but still very focused geographically
- Governance is strengthening, but needs to diversify with the growing community, and engage broadly with repository developments
- Multi-tenancy is easy – or hard: new repositories take seconds, a shared user interface between all tenants…
- Hyku uses Samvera/Hyrax developments, except where it doesn’t
- It should be easier to make Hyku look good and to provide an engaging user experience
- Make Hyku meet high demands and expectations to meet university scholarly communications and compliance requirements
Advancing Hyku Project

- $1M Arcadia grant (charitable fund of philanthropists Lisbet Rausing and Peter Baldwin)
- October 2019 - August 2021
- Supporting the growth of open access through institutional repositories by introducing significant structural improvements and new features to Samvera Hyku.
- [https://advancinghyku.io/](https://advancinghyku.io/)
Takeaways from Advancing Hyku project

- Community codebase contribution for Hyku improvements
- Multi tenancy service providing
- Hosted solution to ease institutional burden
- Joint effort to meet wider institutional, national and international requirements

A hypothetical 2030 library / knowledge environment
2030: the debate – is over: OA has won

https://www.flickr.com/photos/gforsythe/14270651836
…what will be the role of libraries though?
Open science – can we really support it?

https://about.hindawi.com/blog/a-radically-open-approach-to-developing-infrastructure-for-open-science/
Persistent identifiers: end of local data?
Our #1 UI will be for software not people

Algorithms will manage our content

https://commons.wikimedia.org/wiki/File:Motivator_HAL_9000_soory_dave.jpg CC BY 3.0
Big publishers reborn as data companies
How might the “new paradigm” work out?

- Our digital services will be provided by a few service providers
- Our role in technical open science services may be smaller than we expect
- We will procure, not run services and our information will be in the cloud
- Traditional role of discovery, access and managing content might be reduced
- Repositories could disappear behind commercial discovery layers
- We may need fewer but more highly skilled staff
The Digital Shift and community action
“In 2030, UK research libraries will be an integral part of both the local and global knowledge environment.

We will offer environmentally sustainable, inclusive services that enable a diverse set of users to identify and use trusted knowledge resources from all over the world.

Our open research and digital scholarship services will enable seamless and persistent sharing and reuse of research outputs.

Library staff will be recognised experts in (digital) research methods and valued partners in the research process, even leading in some fields.

Print collections will be as easy to discover as our digital ones, and we will provide digital and physical spaces that meet 21st century researchers’ needs.

Libraries will have mastered the use of technologies such as artificial intelligence, and we will deploy them on technology platforms that are open and transparent, built on sustainable and ethical principles.”

https://www.rluk.ac.uk/digital-shift/
Medium-term (2022-2024) objectives

1. Engage: with key stakeholders around the core challenges of the digital shift identified in phase 1.
   - 2.1.1. Establish a programme of knowledge exchange activities with experts from outside the library sector
   - 2.1.2. Contribute to a sustainability action plan for carbon-neutral digital research library services to help inform members in their decision making and those of their institutions

2. Build: a consensus around the steps to be taken to address skills gaps and the role to be played by RLUK, external stakeholders, and individual members.
   - 2.2.1. Create a strategic workforce plan for RLUK libraries, informed by the results of the skills audit and stakeholder mapping undertaken in phase 1. This needs to build on, and support, wider workforce planning.
   - 2.2.2. Working in conjunction with CSN, develop a manifesto/set of requirements for open information sources/infrastructure that will inform licensing and procurement for RLUK members and ideally beyond
Time to act

• Think in service and user value, not in systems terms
• Stop customising and forking
• Only develop/host if you can do it better than others
• Develop an engaging user experience
• Give credit to open sources throughout all workflows
• Rally around community infrastructures
• Follow open procurement principles
• Do not leave AI to closed (commercial) approaches
• …and try Hyku!
A Hyku Haiku

“Blustery autumn
open infrastructures
a community”