LIBSENSE-Connect - Establishing a National Repository Service: Workshop Report

For attention of: Omo Oaiya, WACREN

Compiled by: Paul Walk, Antleaf

2024-10-10

Introduction

This brief report summarises observations made in preparation prior to, and during, the *LIBSENSE-Connect: Establishing a National Repository Service Workshop* held in Abuja in August 2024[1].

The general goal of the workshop was to consider requirements for a shared repository platform for Nigeria, and/or the broader West African region. Participants were invited to comment on pre-stated requirements, and to articulate any new requirements that they might have.

The workshop was facilitated by Paul Walk from *Antleaf*.

The growing importance of repositories

Repositories are, again, becoming important in scholarly communication. Some modern requirements for repositories are outlined here.

Meeting anticipated national open-access policy requirements

Funders of academic research in many parts of the world now require that work which they have funded be made *Open-Access*. In general (and especially where no other affordable publishing route exists) this can be achieved through deposit into an open-access repository. The draft *Model Open Science Policy of Nigeria* (March 2024) follows this same trajectory and makes frequent reference to repositories as important components in satisfying the policy requirements [2].
Furthermore, funding policies will sometimes impose rules on aspects such as limiting embargo periods, mandating certain classes of licensing for re-use, the use of particular metadata schemas or identifiers etc.

Support for new publishing paradigms

In some domains there is a growing interest in and appetite for alternative publishing arrangements - notably the emerging Publish Review Curate (PRC) paradigm, established by eLife in 2023 [3]. PRC favours the notion of the Reviewed Preprint, and opens up the possibility of repositories becoming first-order components in the scholarly communication environment. This, coupled with increasing interest in "overlay journals" presents a good opportunity for repositories to take a greater role in scholarly publishing [4].

The need for a shared repository platform

The general need for repositories in the region can be answered in a number of ways. The focus of the workshop - and of this report - is on one approach: the deployment of a shared repository platform.

The current repository landscape in Nigeria

The repositories "landscape" in Nigeria (and in the wider West-African region) is a mixed one.
According to data collected for the COAR IRD service (in development), there are 41 recorded repository systems operating in Nigerian institutions. This data is based on harvesting well-known registries such as OpenDOAR [5] and re3data [6], so it may be that there are others not known to these registries. In any case, this seems like a low number when compared with the number of "research organizations" (320) recorded in ROR [7].
Of these 41 recorded repositories:

- 25 were found to be online and responsive
- 4 were judged to be probably still "live", although offline and non-responsive at the time of checking
- 12 were found to be apparently defunct [8].

Approximately 30% of the repository systems recorded for Nigeria are almost certainly defunct. This is somewhat higher than the global average of approximately 20%.
Of the 29 live instances, the great majority are running on locally deployed DSpace instances, ranging from version 3 through to version 7 (the current/latest version of DSpace is 8). The fact that there some rather out-of-date installations is not, in itself, especially concerning: the situation in most of the rest of the world is much the same.
Of more concern is the number of repositories in Nigeria which do not respond to an OAI-PMH request. Of

the 29 known deployments, only 9 (31%) responded correctly to an OAI-PMH request. This compares quite poorly with the global average, which is approximately 52%.
These figures suggest a repository landscape which is constrained by a lack of technical resources.

With a shared repository platform, limited resources can be pooled and applied more effectively.

Integration with national, regional and international infrastructure

For repositories to function as first-order components in scholarly infrastructure, they will need to support machine-to-machine interactions, using standard and well understood interfaces (APIs).
Some existing standards are close to ubiquitous., For example, all significant repository software platforms support the OAI-PMH protocol which allows external services to access and gather metadata records from a repository. It is therefore assumed by services in the wider scholarly infrastructure that a given repository will respond correctly to an OAI-PMH request.
New standards and protocols are developing quickly (for example the COAR Notify Protocol [9] and Signposting [10], to allow repositories to interact with emergent services. Examples might include:

- services providing "persistent identifiers" (PIDs) - e.g. ARKs [11], ORCIDs [12], ROR [13] etc.
- indexers and aggregators - e.g. CORE [14]
- open peer-review and endorsement services. - e.g. PCI [15]

However, the evidence suggests that repositories in the local region have a fairly low level of support even for OAI-PMH. In most cases, the software does have an OAI-PMH capability, but it has not been correctly implemented (or, sometimes, even enabled). It seems likely that it would be a slow process to establish support for existing and new standards and protocols in repository systems across a majority of research organisations across Nigeria.

With a shared repository platform, these standards, protocols, technologies and integrations only need to be implemented in a single system.

Costs associated with repository services

There are many costs associated with providing a repository service for an institution. These can be characterised in terms of:

- deploying, configuring and and running the system
- rolling out security-updates/new features/bug-fixes to the software
- maintaining auxiliary services - e.g. content/data storage
- user support

There is little available information quantifying these various cost categories, but one study from the US in 2013 provides some data:

Costs	n	Min	Mdn	M	Max
Implementation	17	$1,200	$25,000	$52,100	$300,000
Annual	20	$500	$31,500	$77,300	$275,000
Personnel, Annual	17	$100	$70,000	$86,186	$235,200
Software, Annual	10	$2,500	$23,000	$22,350	$40,000
Hardware, Annual	6	$500	$5,500	$13,250	$50,000

Data from *Institutional Repositories: Exploration of Costs and Value* (Burns CS, lana A, Budd J, 2013) [16].

Clearly, in the case of a repository system deployed by an institution, these costs must be met by the institution. Some of these costs are unavoidable, regardless of the deployment strategy. or example, some personnel costs (e.g. support for local users) are inevitable. However, some other costs are associated with software deployment and are only incurred by an institution if it is deploying its own software.

With a shared repository platform, many of these costs can be defrayed by sharing them.

Selection of a software platform

The selection of an appropriate software platform for the shared repository service is critical. Fortunately, there is a good range of mature, open-source software to choose from.
Most widely-available, open-source repository software platforms are designed to be deployed and used for a *single* repository. Typically they will be deployed by an organisation to support its members or users. It is not yet common for a single repository platform to be deployed to host repositories for multiple institutions.

Multi-repository platforms

In terms of widely-deployed, open-source repository systems, two might be considered ready for use as a multi-repository platform: Samvera Hyku [17] and WEKO 3 [18].

Samvera Hyku

Hyku is a multi-tenanted version of an open-source repository platform called Hyrax. It is a relatively recently developed system, and has a few deployments in the US. It has recently been deployed by the UK's British Library [19]. Hyku could be a candidate for consideration. However, Hyku is written in the Ruby programming language. Expertise with Ruby is somewhat limited worldwide: it is common in Japan, the US, UK and Germany, but is not used much in the rest of Europe. In Africa, expertise with Ruby is rare.

WEKO 3

WEKO 3 is a version of CERN's Invenio RDM platform [20]. It has been developed by Japan's National Informatics Institute (NII) [21] to allow them to host the great majority of Japan's institutional repositories in a single, managed service.
WEKO 3 is a good choice for the Nigerian context for several reasons. It is already proven as a solution to provide a national repository service, and is supported by NII, with whom WACREN already has a working

relationship. It is written in Python, which is the *de facto* programming language of open science. And it is closely related to Invenio RDM, which WACREN has already deployed in a different context.

WEKO 3 appears to be a very good choice of platform to implement a shared repository service for Nigeria.

Rioxx

Rioxx has been developed to address a number of issues not commonly addressed by other metadata profiles. These include:

- Attaching and relating (persistent) identifiers to individual resources, and expressing clear relationships between these
- Clear licensing of individual resources
- Capturing funding information
- Making all of the above available in a machine readable (and even machine-"understandable") way

The philosophy behind Rioxx is that repositories are "infrastructure" and, as such, should present their metadata in as machine-readable way as possible. Thus, if a remote service (e.g. an indexer, aggregator etc.) has a machine-
readable Rioxx record, then it can automatically resolve the following:

- how to access the primary resource(s) of interest - e.g. for a pre-print repository this would often be a PDF
- how to access directly related/supplementary resources - e.g. for a pre-print this might be a related data-set or some software source-code
- how to access other related resources - e.g. for a pre-print this might be a published version of the same paper

Rioxx can also be optimised for particular contexts, by adding further constraints to properties, for example by constraining values of properties to controlled lists. This means that an optimised "Nigerian Rioxx" is possible, should this be desired. This would continue to be interoperable with other flavours of Rioxx, but could be optimised to the Nigerian national policy context. A national repository service would make this easier, since all repositories hosted by that service could support Rioxx "out of the box". Rioxx could sit comfortably alongside the Common Metadata Schema for Nigerian Repositories, with the latter providing richer bibliographic metadata.

A shared repository service could support both the Common Metadata Schema for Nigerian Repositories and Rioxx, to provide both rich bibliographic and machine-actionable metadata.

Common Metadata

Metadata lies at the heart of any repository system. It facilitates the management, discovery and sharing of the repository's content.
For metadata to be useful beyond the repository, it needs to be arranged according to a standard. We call such arrangements *metadata profiles*.
The workshop in Abuja revealed a certain amount of uncertainty about what metadata was necessary, and which profile or profiles might be used or developed. Some previous work has led to the development of a draft profile - the *Common Metadata Schema for Nigerian Repositories_v1* [22]. This is based in part on

JPCOAR [23] - the default profile used by NII to share metadata from Japanese repositories. This profile takes a traditional approach, focussing on bibliographic metadata to describe publications held in repositories. As such, it is a reasonable solution for the shared repository service.
However, repositories are capable to employing more than one metadata profile simultaneously. For example, it is possible to share metadata that is optimised to describe the bibliographic properties of publications in great detail, while at the same time offering a completely different metadata profile which is optimised to make it easier for external systems to navigate the repository's content to find resources. One example of this second type of metadata profile is Rioxx [24].

Conclusions

1. Repositories continue to be important in any national open-access and open-science context, and their importance appears to be growing.
2. The deployment of repositories in Nigeria is patchy, with a some evidence suggesting in general a relatively low-level of support or maintenance for the repositories which have been deployed by institutions.
3. A shared repository service is a sensible and potentially very effective solution to addressing the problem of limited resources for development, support and maintenance.
4. WEKO3 is a reasonable choice as a platform to provide such a shared service.
5. The adoption of Rioxx as a commonly available metadata profile in the shared repository service will increase the ability of the individual repositories to become part of the wider scholarly infrastructure. This can be deployed alongside, and in a way which is complementary to, the Common Metadata Schema for Nigerian Repositories.

End Notes

1. https://indico.wacren.net/event/213/ ↩
2. https://bit.ly/3YyJn39 ↩
3. https://elifesciences.org/inside-elife/54d63486/elife-s-new-model-changing-the-way-you-share-your-research ↩
4. https://en.wikipedia.org/wiki/Overlay_journal ↩
5. https://v2.sherpa.ac.uk/opendoar/ ↩
6. https://www.re3data.org ↩
7. https://ror.org ↩
8. Data verified by Joseph Walk for Antleaf, July 2024 ↩
9. https://coar-notify.net ↩
10. https://signposting.org ↩
11. https://pidslink.wacren.net ↩
12. https://orcid.org ↩
13. https://ror.org ↩
14. https://core.ac.uk ↩
15. https://peercommunityin.org ↩
16. http://www.dlib.org/dlib/january13/burns/01burns.html ↩
17. https://hyku.samvera.org ↩

18. https://github.com/RCOSDP/weko ↵
19. https://iro.bl.uk ↵
20. https://inveniosoftware.org/products/rdm/ ↵
21. https://www.nii.ac.jp/en/ ↵
22. https://docs.google.com/spreadsheets/d/14DftpDbWuhe81Vr0VUZCpWLfR6kwCzjl/edit?gid=1808480562#gid=1808480562 ↵
23. https://schema.irdb.nii.ac.jp/en ↵
24. https://www.rioxx.net ↵