Chapter Nine

Information Management Solutions for Modern Research and Innovation: Interoperability & Why It Matters

Matthew Buys

ORCID Regional Director
Africa & Middle East
Email: m.buys@orcid.org
http://orcid.org/0000-0001-7234-3684

Abstract

Global trends and research information systems development are extremely dynamic and a continuous changing landscape. In this context, where more and more institutions and researchers manage information, the channels increase such that the research information needs to be exchanged between systems. We see this as a digital ecosystem of research information. It is pertinent to advocate the use of standards to facilitate the exchange of data to ensure that the information can flow seamlessly through the ecosystem and be reused to its maximum capacity. In addition to this streamlining the process will reduce the administrative burden faced by researchers and administrators. The paper provides personal insights whilst exploring the current landscape and methodologies.

Introduction

Exploring trends and developments in the changing landscape as well as various standards and platforms such as CERIF-XML, VIVO, ORCID and CASRAI is essential in modern day research management. The research information management landscape has shifted from home-grown stand-alone systems to a complexity of systems and adaptations to allow for the seamless flow of information (Moreira, 2013).

Many institutions take time to make such decisions, as the cost-benefit analysis at the onset seems very high. Institutions or countries showing rapid adoption of standards and systems with a clear set of criteria have seen much better results (Moreira et al., 2015).

This synchronization framework used in the Portugal PTCRIS project has recently been prototyped. The clear project objectives and criteria for outputs ensured that the system achieved excellent results. The first stable and detailed specification of the synchronization framework will be made available soon as an open-access report. (Moreira, 2013) (Moreira et al., 2015)

A recent report produced by JISC and Association of Research Managers and Administrators (ARMA) in the United Kingdom considered various factors in Institutional ORCID implementation and Cost-Benefit Analysis. The report consulted with a number of different stakeholders in the research and scholarly communications process. It was found that the adoption of ORCID would be greater based on funder mandates. Eight pilot institutions participated in the project and early engagement with senior management was key to the success. The velocity of decision-making within an institution becomes a critical success factor. As stated by Henderson et al. (2015) perhaps surprisingly, technical issues were not the major issue for most pilot institutions. A range of technical solutions to the storage of researchers' ORCID iDs were utilised during the pilots. Four institutions used their institutional research information system (CRIS): two used Pure; one Symplectic; and one Converis. Two other institutions developed in-house systems, one used Agresso Business World and one the student portal of SITS e:Vision. Of the eight pilot institutions, only one chose to bulk create ORCID iDs for their researchers, the others opted for the 'facilitate' approach to ORCID registration. The institutions found it relatively easy to convey the benefits to senior management; however, researchers and staff seemed to see this as another level of bureaucracy. In summary, the project unveiled that the cost of implementation was negligible and the potential benefits far exceed the cost. (Henderson et al., 2015).

The skills and systems that are required by the modern researcher are extremely diverse. Institutions need to evaluate the following key functional areas whilst providing the environment for the researchers to grow and position themselves as leaders in their fields. Common challenges for institutions are as follows:

- Research and Innovation Performance of Universities: Africa and the rest of the World;
- Competitive Intelligence for strategic research management, encouraging collaboration, capacity building and training
- Enabling technology transfer and commercialization;
- Measuring and Benchmarking (including ranking for modern university management), using the right tools to measure;
- Using the right research tools, creating a personal brand, measuring performance;
- Finding possible collaborators;
- Where to publish – and how to innovate.

Whilst the list may seem limited, these are critical pain points that the research digital ecosystem should address (Mouton, 2014). In bringing this together, the key functional aspects of a digital eco system should be scalable and interoperable.

The silo effect

One point that I have found resonates with all institutions currently is the hindrance to research administration presented by siloed systems and

Figure 1: Lack of interoperability across the research digital ecosystem
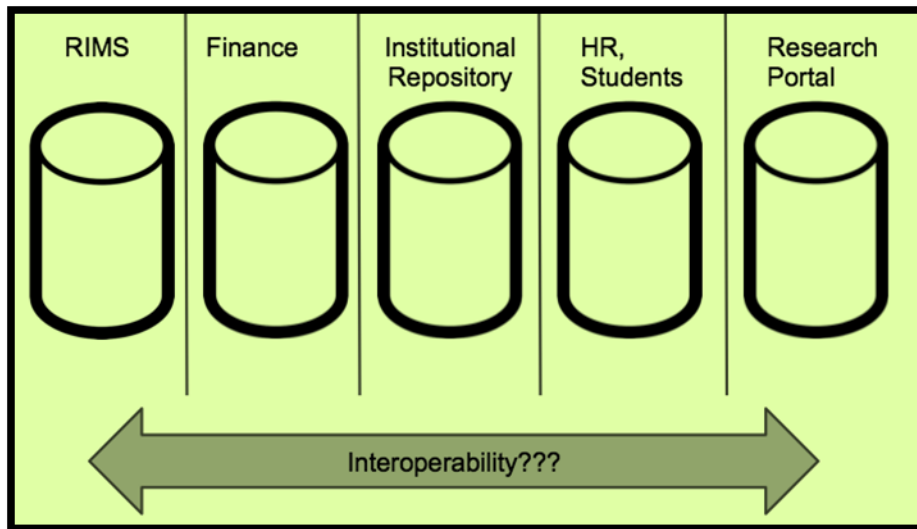
data sources. This takes a large amount of the researchers' time in data capturing due to lack of interoperability.

Institutions are exploring the increasing need for data sharing, and the opportunities presented in using standard data models and unique identifiers in building bridges between systems. This occurs on both a macro and micro level across the digital ecosystem. Persistent identifiers such as ORCID, allow for the seamless flow of information across this system whilst using existing standards. By building persistent identifiers into the workflows there will be the notion of entering data once and using it many times (Otjacques et al., 2007).

Increasingly is interoperability required across all enterprise systems and not the standalone research information management system. The key to interoperability is in identifying vendors that make use of persistent identifiers and standards. This will ensure that the silo is broken down and that information can flow seamlessly across the digital ecosystem.

Scalability is extremely important to avoid the silo effect in the future. Similarly without interoperability, the system renders itself standalone and the information lost within a digital vortex.

Research information management

There are many trends driving the needs for increased professionalization of research management.

Governments are increasing the frequency of national research performance exercises. This not only allows for accurate resource allocation but also increased visibility for the country and the region. Complex research information management systems, allow for government bodies to manipulate the data into meaningful charts and tables.

Funders around the globe have complex application processes which creates ongoing challenges for the researcher when applying for grants. In addition to the complex application process, reporting back on projects and linking research outputs to projects is not always simple. There is often a disjuncture at this point.

Universities are looking to conduct benchmarking both internally and externally. These exercises allow for internal performance reviews, faculty activity reports. In addition to other key functions such as accreditations, compliance (ethics reviews), attracting talent or collaborations and generally increased efficiency in the research workflows.

Finally from the researcher point of view, the research information management system should allow for less administration, more research. In this sense, it should break down the silo and create an interoperable research ecosystem.

Skill sets in research management: Africa

There is an increasing need for skilled individuals within research management. Many research information management systems (RIMS) vendors provide comprehensive technical support, however, there is still a need to have in-house skill sets. These skillsets vary by stakeholder but essential skills are basic programming and API knowledge. This is one of the key challenges, which I have found when discussing these systems in Africa.

The evolution of information systems has meant that most programming code has become open source. As an example, ORCID who serve as a hub to connect digital information post all the code openly to the community. In the same sense, many members of ORCID create codes that either read or write information from the hub and in turn post this openly to the community. Sharing of use cases and web services code certainly is a positive step, however, there is still the need to for someone in-house to be able to make use of this code.

In addition to hard skills, research management is also moving towards dissemination of knowledge and skills. Researchers are demanding specific training on systems and the creation of guides to navigate this now, complex digital ecosystem. The skillset in research management should be able to breakdown the complexity for the individual users and showcase the benefits.

Interoperability and persistent identifiers

Interoperability only becomes possible when information systems use a common language (or data dictionary as it is commonly referred as). In this sense, one does not need to change the source code but rather build adaptations on the periphery of these systems to create a common language web services interface.

According to (Zhao and Xia, 2014) their literature review indicates that interoperability has never been formally examined in prior empirical studies of interorganisational systems. It is unclear how interoperability should be conceptualized and operationalised in the context of digital value networks. Also under researched is how interoperability is formed and whether it can lead to organizational performance gains.

If one can make use of standards and a common language, interoperability becomes possible despite the heterogeneity in software. These common languages and standards in the research digital landscape are as follows:

(i) CASRAI

The Consortia Advancing Standards in Research Administration Information (CASRAI) is a non-profit organization that is dedicated to reducing the administrative burden on researchers and improving business intelligence capacity of research institutions and funders (CASRAI, 2015).

Their approach is simply to improve the flow of information and the various stakeholders in the digital research ecosystem. CASRAI serves to change the source code of systems but rather provide adaptations through the CASRAI common data dictionary. This enables system-to-system interoperability and seamless flow of information (CASRAI, 2015).

There have been several organisations that have adopted CASRAI common vocabulary (from the data dictionary) and used it to produce compliant CVs for researchers. This aids the funders in receiving data in a compliant format and streamlines the process.

CASRAI develop and maintain a common and extensible dictionary of terms and exchangeable business objects that form bridges in our shared work processes. In addition CASRAI provide a forum and the mechanisms required to standardize the data that researchers, their institutions and their funders must produce, store, exchange and process throughout the lifecycle of research activity (CASRAI, 2015).

(ii) euroCRIS CERIF

euroCRIS is the European organisation responsible for publicising work on current research information systems (CRIS). The CERIF task group maintains a standard for CRIS systems to enable interoperability. Its focus is in Europe, however, it serves to address global challenge (euroCRIS, 2015).

The primary objective of euroCRIS is to serve as a common platform for dialog and discussion of common issues within research information management.

- Promote and improve communication and interaction between global CRIS;
- Maintain and publish the CERIF (Common European Research Information Format) recommendation and any standards endorsed by euroCRIS;
- Organize and run the CRIS series of conferences with associated workshops and other events;
- Provide a source of expertise in CRIS to members and to others under business arrangements made at the time;
- Develop euroCRIS guidelines;
- Nurture the CRIS community by events, a monthly newsletter, an online discussion forum and other appropriate mechanisms;
- Provide a forum for exploring and exploiting new and emerging concepts and technologies (including data quality, standards, etc.);
- Establish a one-stop portal / gateway to international CRIS resources.

(euroCRIS, 2015)

The premise for interoperability is that it requires a structures schema. CERIF serves to act as a model for a standalone (homogenous) system as well as adaptations for heterogeneous systems to facilitate data exchange and create a common data warehouse (Zhao and Xia, 2014).

(iii) VIVO

The VIVO project allows for researchers across institutions to be discovered and information to be shared. Institutions within the VIVO network set up local installations that will then allow for transfer of data amongst other institutions on the network. VIVO works with a range of stakeholders across the research lifecycle and include data such as researcher interests, activities, and accomplishments. This enables seamless discoverability of research information (VIVO, 2015).

Whilst standards are fundamental, it is as important to incorporate a persistent identifier into the research workflows. Name ambiguity has become a major challenge in research information management (Gilchrest and Blalock, 2014). It has been found that algorithms are simply not enough to create a profile. This is due to any number of the following issues when publishing:

- Different versions (full name versus initials)
- Shared names
- Transliteration
- Accents and other ALT characters
- Name changes
- Multiple family names

(Haak et al., 2012)

This creates a challenge across the entire research lifecycle, as metadata is either incomplete or incorrect. The use of a unique persistent identifier assists with flow of information between systems and improve the integrity of the data. This creates retrieval issues for users. *Name disambiguation, the process of identifying, merging, and making names accessible in one standard form, is a vital process repository staff should incorporate into their workflow to address these issues* (Walker and Armstrong, 2014).

ORCID is also works to build trust in research profiles. It serves to be a hub connecting information across the research digital ecosystem. In allowing this interoperability, the source of information can affirm the credibility and therefore reducing the self-claiming procedure. The source

as an example would be CrossRef. In addition this allows for data to be entered once and reduce the administrative burden of the researcher. ORCID is working closely with CrossRef and DataCite towards a metadata round up. This process will allow for the following:

1. ORCID persistent identifier to be capture through authentication at the publisher.
2. ORCID persistent identifier to be built into the production workflows and send with the metadata to either CrossRef or DataCite.
3. DOI and other metadata then pushed into the researchers ORCID profile.
4. ORCID will then push this data into various member integrations such as the institutional repository and institution profile system.

(Paglione, 2015)

The road ahead: it takes a village

The road ahead requires community involvement where all stakeholders work together. Publishers are beginning to use persistent identifiers in their workflows. The key is to ensure that these identifiers are pushed through to production and then associated with a digital object identifier (DOI). This allows for seamless tracking of the research output and almost zero administrative burden on the researchers. Through notifications and interoperable systems the institutions could pull data into their repositories. An example of such notification would be a new publication in which the institution then sources full text for their repository, effectively enhancing the integrity of the repository. Once the data enters the institutional digital ecosystem, systems should be able to push and pull data seamlessly through the use of standards and common data dictionaries. Finally, funders can associate research outputs to projects whilst also tracking the peer review process (Allen et al., 2014).

I see there to be three key pillars to building a scalable research digital ecosystem:

(i) Persistent identifiers

Persistent identifiers cannot simply exist to serve as an identity but also need to be built into the workflows of the research community. This will enable seamless tracking of information for the administrator and clear visibility for the researcher.

This process should be managed by systems in its entirety and avoid any manual entry whatsoever. Information should be self-created or by the system, however, ensuring compliance with privacy requires a digital "handshake" and authorisation from the researcher.

Bi-directional flow of information through identifiers is the most important aspect as without this it renders it as a simple identity lost in a digital vortex. Building trust is enabled through the flow of information, allowing institutions to affirm an affiliation or research output in the researcher's profile (Gilchrest and Blalock, 2014, Walker and Armstrong, 2014).

Finally use cases for researchers such as populating their CV automatically and seamlessly applying for funding start to become key positive outcomes. This allows for simple and accurate benchmarking, linking funding to projects and research outputs, and improving the integrity of the data throughout the research digital ecosystem (Haak et al., 2012).

(ii) Research Information Management Systems (RIMS)

There are a number of RIMS providers that offer varying levels of service and functionality. It is important to clearly define the scope and objectives of the system. There are a couple key aspects that I would recommend an institution should evaluate:

- Functional workflows – the system should allow for customisation of workflows and approval processes without changing in source code;

- Standards and adaptability – the data model should be extensible and offer a degree of customisation, whilst still adhering to various standard employed;
- Support – there should be round the clock support and maintenance for such a system. It is pivotal in most research organisations and should something go wrong there must be a quick turnaround on resolving the issue;
- Scalability – this is one of the key aspects of any modern day system, as to how scalable is this for the future. There are so many legacy systems within the digital research ecosystem and it has become increasingly costly (and resource intensive) to maintain these systems. A truly scalable system has the three previous aspects; functional workflows, standards and adaptability, and extensive support.

Standards assist to break down the silo effect and allow for better interoperability and flow of information.

(iii) Standards

In this context, where more and more institutions manage research information; funders and national research performance assessment exercises, the channels increase where the research information needs to be exchanged between systems. This is a digital research ecosystem.

It is therefore important to advocate for the use of standards across all systems and hubs. This allows for future scalability and also importantly better interoperability across systems (Zhao and Xia, 2014). The benefits are ten-fold and not only realised by the institution but also the researcher due to the reduced administrative burden. Data can be entered once and re-used many times.

Summary

Researchers want to be read, acknowledged and quoted. The digital framework of the research ecosystem should be enabling of the basic researcher needs and reduce their administrative burden; allowing researchers to spend more time on their research and less on administration.

Scalability and interoperability for the future are two key terms that should be synonymous with research management. The key benefits of a truly interoperable research ecosystem will provide the following outcomes:

- Save time for researchers (applying and reporting);
- Improve access to quality data for institutions and funders;
- Simplify the measurement of research impacts on society;
- Provide peer networking opportunities for teams tackling admin data issues. (CASRAI, 2015)

Together the research community can build an open access framework for the digital research ecosystem. The technology is fast moving and stakeholders across the research lifecycle should attempt to share ideas for better interoperability in the future.

References

Allen, L., Brand, A., Scott, J., Altman, M. & Hlava, M. 2014. Credit where credit is due. Nature, 508, 312-313.

CASRAI. 2015. About | CASRAI [Online]. Available: http://casrai.org/about - .VZEKdhOqqko.

EUROCRIS. 2015. Main features of CERIF | euroCRIS [Online]. Available: http://eurocris.org/cerif/main-features-cerif.

Gilchrest, B. A. & Blalock, E. N. 2014. Who wrote that? J Invest Dermatol, 134, 3-4.

Haak, L. L., Fenner, M., Paglione, L., Pentz, E. & Ratner, H. 2012. ORCID: a system to uniquely identify researchers. Learned Publishing, 25, 259-264.

Henderson , H., Johnson, R. & Woodward, H. 2015. Institutional ORCID Implementation and Cost-Benefit Analysis Report.

Moreira, J. M. 2013. Eco-systems of research information systems: a global perspective.

Moreira, J. O. M., Cunha, A. & Macedo, N. 2015. An ORCID based synchronization framework for a national CRIS ecosystem. F1000 Research.

Mouton, J. 2014. Introduction to Bibliometrics. SARIMA Research Management for Research Librarians Workshop 26-27 August 2014.

Otjacques, B., Hitzelberger, P. & Feltz, F. 2007. Interoperability of E-Government Information Systems: Issues of Identification and Data Sharing. Journal of Management Information Systems, 23, 29-51.

Paglione, L. 2015. New Functionality Friday - Auto update your ORCID Record | ORCID [Online]. ORCID. Available: http://orcid.org/blog/2014/11/21/new-functionality-friday-auto-update-your-orcid-record.

VIVO. 2015. About | VIVO [Online]. Available: http://www.vivoweb.org/about.

Walker, L. & Armstrong, M. 2014. "I cannot tell what the dickens his name is": Name Disambiguation in Institutional Repositories.

Zhao, K. & Xia, M. 2014. Forming Interoperability Through Interorganizational Systems Standards. Journal of Management Information Systems, 30, 269-298.