Redefining the Institutional Repository : webinar note

Note: These are my notes for the webinar. As such they are written in an informal, oral style. It is not written in the style of an article. Not everything included in the webinar is included here, while some points mentioned here were left out of the webinar.

http://www.uksg.org/institutionalrepository

Institutional Repositories have grown in importance over the last 10 years to offer a core University and Library service, however, their role is developing faster now than it has ever done. Funder Open Access requirements, internal reporting, research data. Ref2020 and more are increasing the demands on the traditional repository, putting pressure on staff resources and challenging the underlying software.

This webinar will outline these issues as well as look at how the needs and use of repositories may change in the future.

Date: Thursday 27 November 2014

Time: 1300 GMT

Duration: 45 minutes including Q&A (up to 60 minutes maximum if there is sufficient demand for an extended Q&A)

SPEAKER:

Chris Keene, Technical Development Manager, University of Sussex Library

Chapter 1 : a very brief history of the institutional repository

What is an Institutional Repository?

First: it's a stupid name. Putting research online for the benefit of the world should be something exciting and enticing, instead it stirs the same emotions as a system for filing your tax returns.

So… what is an IR? My definition, "*Web based free access to a university's research outputs*" As simple as that. The fact there's software and processes behind the scenes is not really relevant, that's what it does.

Though, as we will see, for many the original role has been joined by other requirements, and at times this has taken a backseat.

What does it look like?

I won't dwell on the technical elements, they're not particularly important. The issue instead was one of awareness and engagement.

However I will give a brief outline of what is involved in adding items to an IR, particularly for those who have never used one.
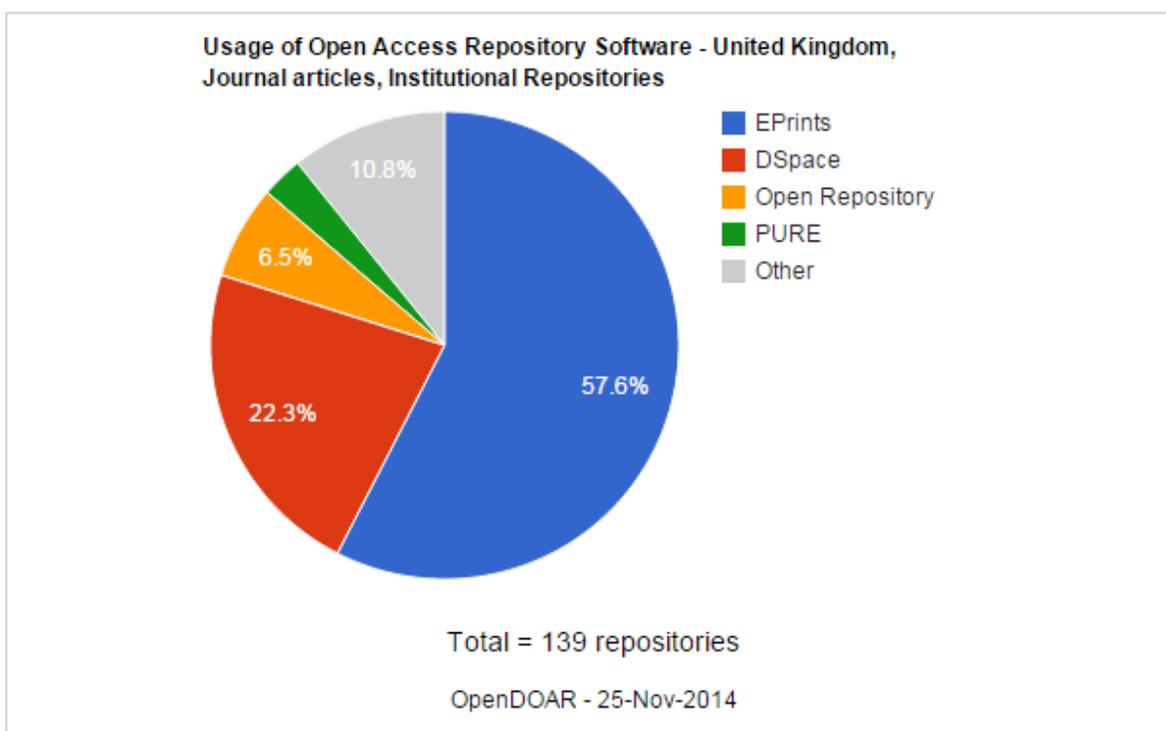
Items are normally added by the academic, or their assistant, or someone in the library. It will vary from University to University. For example, here at Sussex we don't add records in the Library, except for theses. So it's down to the researcher, or someone in their School who can support them.

Once submitted, someone in the library will check the metadata and copyright and make the record live. For a journal article, checking the copyright involves checking SHERPA Romeo. Technically, the source of truth is the CTA (copyright transfer agreement) the academic signs with the publisher, but as the University almost certainly doesn't have access to it we rely on SHERPA Romeo to provide a generic version of the CTA for that publisher. For conferences and

book chapters, it is a whole lot more complex. Those with the luxury of staff time may choose to write to the publisher.

Once the metadata is checked and any uploaded file is given restricted access if copyright dictates, the record can become live. Most software by default locks the record at this point, which can be a source of irritation to academics who take offense that they can't edit the records of their own outputs, something which I can understand.

In the UK, the most popular software package is Eprints, followed by Dspace, Fedora and a handful of others.

Usage of Open Access Repository Software - United Kingdom,
Journal articles, Institutional Repositories

- EPrints 57.6%
- DSpace 22.3%
- Open Repository 6.5%
- PURE
- Other 10.8%

Total = 139 repositories

OpenDOAR - 25-Nov-2014

1

Timeline

I'm not going to give a full blown history, but will focus on two dates. The first is 2001, the date of the *Budapest Open Access Initiative,* which resulted in a statement of OA principles. In

---

1

http://www.opendoar.org/onechart.php?cID=224&ctID=1&rtID=2&clID=&lID=&potID=&rSoftWareName=&search=&groupby=r.rSoftWareName&orderby=Tally%20DESC&charttype=pie&width=600&height=300&caption=Usage%20of%20Open%20Access%20Repository%20Software%20-%20United%20Kingdom,%20Journal%20articles,%20Institutional%20Repositories

2002 this was opened up for others to sign. Software was developed and made available around this (Eprints 1999, Dspace 2002, Fedora 2003) and early adopters launched a repository.

By 2007 many Universities had an IR, including most which describe themselves as Research Intensive.

By this time, the IR was part of the lingo in an academic library. The problem was it wasn't, for all but a few academics?

Engagement
On the face of it, adding a research output to an IR was a no brainer. A win win for all.

2

http://www.opendoar.org/onechart.php?cID=224&ctID=1&rtID=2&clID=&lID=&potID=&rSoft
WareName=&search=&groupby=r.rDateAdded&orderby=&charttype=growth&width=600&heig
ht=350&caption=Growth%20of%20the%20OpenDOAR%20Database%20-%20United%20King
dom,%20Journal%20articles,%20Institutional%20Repositories

Academics got increased exposure of their work, a greater chance of being read (cited?) by other researchers and by the wider public, such as schools and those in developing countries - the warm feeling that doing that provides; statistics; being able to access their own research in the future. Plus - access to research by others in an IR which they would otherwise not have access to.

For the University: linking the research to their name (something easily missed once the research is in a journal, the media will often report 'the research published in The Lancet today says that …'), it will help with strategic goals of knowledge exchange and sharing with the community and may even help foster new relationships. Universities can also see in one place the research they are producing.

And of course for the wider, internet connected, world there are benefits in terms of access.

All this from ten minutes work adding the item to a IR. What's not to love? Especially considering the amount of admin required to get an item published in the first place, relatively speaking this isn't much at all.

In some ways, I can see why take up was not good. Academics were and are under pressure regarding student admissions, the experience of current students, the NSS, research funding etc. Doing something 'nice' such as putting their research online was not in the forefront of their mind. And even when it was, it meant learning a new system, finding files, asking for help (and the fear of looking stupid), and did they really want their draft versions visible to all?

So a key issue was getting academics to use it. Another was resourcing (though it would probably vary for different universities). Certainly we found, in a world where academics wanted more access to journals and our budget was going up by less than journal inflation, and students wanted more books for their course, it was not just hard to argue for resources for the IR, it would be politically unwise. 'Why are you spending money on this when you want more on the core services you should deliver?' Showing we spent money on marketing and supporting an IR may not help our case.

IRs were a "nice to have", and so never on the top of a University's or academic's agenda.

Dorothea Salo's '*Innkeeper at the Roach Motel*' is a great overview of this period[3].

Chapter two: The publication store

This chapter starts with what may seem like a bit of a tangent.

It seems a fairly basic need for an organisation to be able to quantify what it does; what its outputs are.

Universities do two core things: teaching and research. For teaching, the outputs are quite clear, crudely put: a university will know exactly how many students graduated with degrees.

For research, it surprises me that most Universities really didn't have a way to track the research outputs they were creating. For Research funding, of course, they had a much clearer idea of the money coming in, and out, to fund their research activity, but not what was actually being produced.

Meanwhile, during the 2000s, academics were increasingly having to produce lists of their research outputs: for new web profiles, for funding bids, for the RAE and internal reports for their department or faculty.

This was an opportunity for the IRs.

I said earlier that the key concept of the IR was uploading a Research Output to the web. A file uploaded on its own isn't that useful, so some details about the research output would be added alongside it.

This information, bibliographic data, could be used as one central place to hold information about a University's Research outputs.

For a number of repositories (the proportion may be larger in the UK than other countries) the metadata become of equal or even greater importance than the Open Access. IRs could end up with a majority of 'metadata only' records. OA was still encouraged, and welcomed, but the focus had moved.

---

[3] http://minds.wisconsin.edu/handle/1793/22088

Just like OA, the benefits seems obvious and a win win for all. Academics wouldn't need to fill out their research activity again and again. Deans and PVCs could see what their School and University was producing. They could also see which researchers were producing high levels of outputs, and which ones less so. All from a system that was designed to hold bibliographic data.

However, again, take up and engagement mostly did not really happen. It varied from University to University, but the outputs in the IR nowhere near represented the actual number of outputs from the University. Or at least we had to presume so, there being no single source to find the actual number of outputs a University produces (Web of Science, and in particular Scopus, help in this area, but neither tracks all the journals our researchers publish in).

Finally, a good example, just a year ago was the REF 2014. Many Universities used their IR for collecting research output information, including information specific to the REF itself (REF sub-panel, group, etc). However it was not uncommon for researchers to enter just those items they planned to submit for the REF and no more.

Chapter three: The haves and have nots.

Around 2010 a new product started to appear at some Universities: The CRIS. (Common Research Information System).

Broadly speaking A CRIS manages much of the administrative research process. They can often pull in publication output information for external sources such as WoS and Scopus, this helps make adding outputs as easy as possible, though doesn't include all outputs, especially in some subject areas.

As shown by the diagram, a system such as a CRIS will often be connected to the IR. Normally the CRIS feeds publication information into the IR, for items with a full text item. Hence the IR becomes something as it was originally envisaged: open access research on the web.

We shall see in a moment that there is an increasing need to capture and report on information around publication outputs. Universities can be divided into two groups, those with a research system such as a CRIS or otherwise, and those who do not, and hence have to rely on the IR to adapt to the various requirements. It's important to note these two groups as they will often had different demands on the software and future development.

Chapter four: Funders; Mandates and Compliance

In June 2012 the Finch Report was published. From the HEFCE website[4]:

---

[4] http://www.hefce.ac.uk/whatwedo/rsrch/rinfrastruct/oa/oa/

> **The 'Finch' report** - Dame Janet Finch chaired an independent working group on open access. The group's report, published in June 2012, supported the case for open access publishing through a balanced programme of action.
> The report recognised the need for different channels to communicate research results, but recommended support for the 'gold' route in particular.
>
> **Government** - The Government accepted all recommendations in the Finch report. In its formal response it has asked the four UK higher education funding bodies and the Research Councils to put the recommendations into practice by working with universities, the research and publishing communities.

The was the start of quite a fast process the led to a Government response[5] a month later fully backing the report and tasking RCUK and HEFCE with adopting it. In the same month RCUK announced its new Open Access policy[6].

The RCUK policy[7] mandated that all peer-reviewed research papers funded by the Research Councils should be published in a OA-compliant journal, with requirements around length of embargo for Green OA and a CC-BY licence for Gold OA. This would apply to any outputs submitted after April 2013.

To support this, RCUK provided funds to a number of Universities (those which receive a certain level of RCUK funding) to pay for the Gold publication.

Universities had a number of issues to grapple with, and with a relatively tight timeline relative to the speed of change in Higher Education.
How to turn RCUK policy into a University Policy and procedure?
How to allocate the funds? (if any!)
How to track and monitor, in particular so we can report our compliance back?
And how to disseminate all those to Researchers?

And how to endure that decisions around where to publish were kept as academic decisions, not those dictated by support services (Library, Research Finance Office) and budgets.

---

[5] https://www.gov.uk/government/news/government-to-open-up-publicly-funded-research
[6] http://www.rcuk.ac.uk/media/news/120716/
[7] http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/RCUKOpenAccessPolicy.pdf

It required the Library and Research Office to work closer together than they had likely ever done before. Libraries had little understanding of Research Financing, Grants and Funders. And Research Officers had less of an understanding around the scholarly publishing process and Open Access.

As well as the larger policy questions for a University, there was the nitty gritty of exactly what was required, developing detailed workflow and looking into how we would record this information for compliance and reporting.

How would we even know if an output has been funded by RCUK? A researcher may have added it to the IR and made it compliant, but how would we know to even include it in our reporting.

And what *would* we need to report to RCUK? At the time it hadn't been stated. You can't collect data that you don't know you need.

At Sussex we added a number of fields to our IR to capture information that would be useful to report to RCUK. Funder information tracks the funders and projects connected to a project. We worked with our IT Services to develop a feed in data into the IR, so that you just have to select a project name and the other fields will be completed. We also had some fields around OA, though these are currently being reviewed.

The funder fields were essential as we need to identify, track and report on outputs which were RCUK funded (and increasingly for other funders as well).

Funder Information

	Project Name	Sussex Project Number	Funder	Funder Ref	
1.	Single Molecule Imaging of th	G0250	EUROPEAN UNION	268788	▼
2.	Replication fork stability and	G0745	MRC-MEDICAL RESEARCH	G1100074-E01/1	▼▲
3.					▼▲
4.					▼▲

[More input rows]

Open Access

Open Access Status:
Unset
Gold: Item freely available on publishers website
Green: Item freely available from SRO
Green: Item freely available from SRO after embargo period
Not Open Access: Item is not freely available and not RCUK-funding compliant ▾

Embargo Length (months): [   ]

Licence Type:
Unset
All rights reserved
Creative Commons with Attribution (CC-BY)
Creative Commons with Attribution non-Commercial (CC-BY-NC) ▾

The funder information is displayed publicly on the output's IR record:

mTOR kinase-dependent, but raptor-independent regulation of downstream signaling is important for cell cycle exit and myogenic differentiation

⊞ Tools

Pollard, Hilary J, Willett, Mark and Morley, Simon J (2014) *mTOR kinase-dependent, but raptor-independent regulation of downstream signaling is important for cell cycle exit and myogenic differentiation*. Cell Cycle, 13 (16). pp. 2517-2525. ISSN 1538-4101

PDF - Published Version
Available under License Creative Commons Attribution.
Download (502kB) | Preview

Official URL: http://dx.doi.org/10.4161/15384101.2014.941747

Abstract

Myogenic differentiation in the C2C12 myoblast model system reflects a concerted and controlled activation of transcription and translation following the exit of cells from the cell cycle. Previously we have shown that the mTORC1 signaling inhibitor, RAD001, decreased protein synthesis rates, delayed C2C12 myoblast differentiation, decreased p70S6K activity but did not affect the hypermodification of 4E-BP1. Here we have further investigated the modification of 4E-BP1 during the early phase of differentiation as cells exit the cell cycle, using inhibitors to target mTOR kinase and siRNAs to ablate the expression of raptor and rictor. As predicted, inhibition of mTOR kinase activity prevented p70S6K, 4E-BP1 phosphorylation and was associated with an inhibition of myogenic differentiation. Surprisingly, extensive depletion of raptor did not affect p70S6K or 4E-BP1 phosphorylation, but promoted an increase in mTORC2 activity (as evidenced by increased Akt Ser473 phosphorylation). These data suggest that an mTOR kinase-dependent, but raptor-independent regulation of downstream signaling is important for myogenic differentiation.

Item Type:	Article
Schools and Departments:	School of Life Sciences > Biochemistry
Subjects:	Q Science > QD Chemistry > QD0241 Organic chemistry > QD0415 Biochemistry
Depositing User:	Catrina Hey
Date Deposited:	10 Nov 2014 15:36
Last Modified:	10 Nov 2014 15:36
URI:	http://sro.sussex.ac.uk/id/eprint/51138

## Actions (login required)

🔍 View Item

Document Downloads

More statistics for this item...

⊟ Funder and Project Information

Project Name	Sussex Project Number	Funder	Funder Ref
mTOR signalling, hyperphosphorylation of 4E-BP1 and translational control during myogenic differentiation	G0193	BBSRC-BIOTECHNOLOGY & BIOLOGICAL SCIENCES RESEARCH COUNCIL	BB/H009728/1

University of Sussex: Open Access publishing workflow for RCUK-funded publications
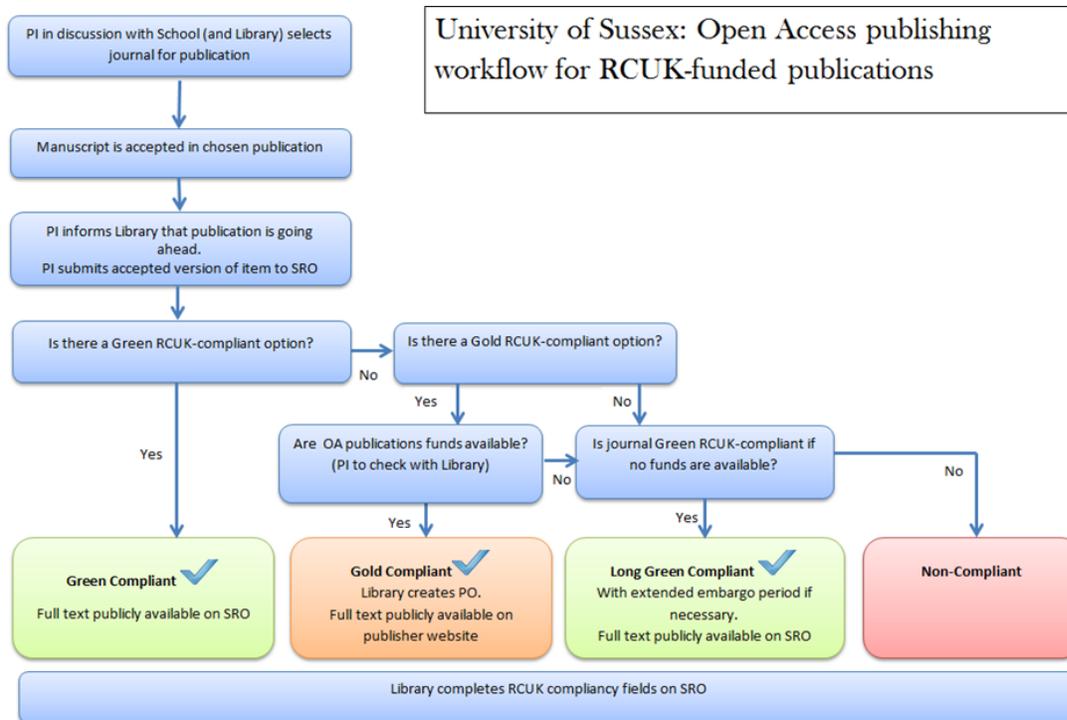
The diagram shows a first attempt at a decision flow chart for deciding green or gold. In practice, of course, things are much more fluid, and expecting researchers to contact us was somewhat more of an aspiration. It's more typical that our first contact (if any) with an academic is often at the point that they have committed to pay for Gold and simply want to know how to pay for the thing.

Meanwhile, the process of raising a Purchase Order (and if need be request a new supplier on the finance system), and receiving authorisation can take days. And add to the mix this is also new to publishers, who have at times struggled to adapt to the new requirements, and who are not used to dealing with libraries in the editorial process (and sending invoices to academics rather than the correct person).

In August 2014 RCUK announced[8] that the key compliance reporting requirements for this period were four numbers:

•The number of peer-reviewed research papers arising from research council funded research that have been published by researchers within that institution.

---

[8] http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/ComplianceMonitoring.pdf

•Of these research council funded papers, the number that are compliant with the RCUK policy on Open Access by:

  • The gold route

  • The green route.

•And the number which have been published in a journal which is not compliant with the RCUK policy on Open Access.

The biggest challenge was reporting on non-compliancy. How do you report on those you do not know about? This required a large amount of time to manually search abstract databases (mainly Scopus) for our PIs to look for outputs which fall in this time period. Once found, we needed to check if they were RCUK funded research, and if they were *submitted* after the 1st April 2013.

Often, we could actually added them to our IR to make them compliant, if the journal's policy allowed and the author had the right version available. And if not we would count the output as non-compliant.

But it shows again the problem of Universities not knowing what they produce.

Like Buses

RCUK is not alone.

A number of other funders have also put mandates in place in the last 18 months.

Europe has an OA policy for their new Horizon 2020 stream of funding, which places strong emphasis on deposit into an IR.

The Wellcome Trust has a policy that encourages immediate OA or short embargos and CC-BY licensing. It has also, in partnership with a number of other charities, set up a fund (COAF) to help pay for the costs of Open Access.

and then we have the REF...

REF 2020 (or post-2014 REF)

Journal articles and conference papers submitted to the REF 2020 with an acceptance date of 1 April 2016 or later will have to be available as Open Access, with a maximum embargo date of 12 or 24 months (depending on subject) after acceptance.

The author's accepted version will need to be deposited into a repository within three months of acceptance for publication. That means choosing a journal that complies, and uploading the article onto the IR, or other repository service, within three months.[9]

This is a massive undertaking. The REF potentially applies to all academics, not just those funded by a particular funder.

Ensuring as many as possible comply will be a real challenge, and processes put into place to support this real need to scale.

The repercussions of not doing so are large: the REF dictates the block Research Grant for a University; it can make or break a University in terms of Research.

Universities can't start to prepare for the REF a couple of years before its deadline, as for some has been the case in the past. By that time, it will be too late to deposit the research into a repository within three months of acceptance.

Universities are really only just starting to address how to respond to the REF Open Access requirements. It will require a change of culture across the institution. It will also require a notable increase in staffing resource to support and inform academics, and to supplement and check the required extra metadata.

Research data

Another Funder requirement is around Research Data. Typically, funders are requiring researchers and sometimes institutions to have research data management plans, and are encouraging the sharing of data and making it open.

There are a number of different approaches here. A University can use its existing IR, can implement a separate repository just for data, or use an external service. There are also

---

[9] http://www.hefce.ac.uk/media/hefce/content/pubs/2014/201407/HEFCE2014_07.pdf

different storage options, such as traditional local storage, 'cloud' based storage (Amazon S3) and specialist archival systems. One example is Loughborough, who have recently announced [10] they will use Figshare as a frontend and as a backend for storage.

The Issues

Many of the issues are apparent above, but to summarise some
- Policy (with others); develop procedures; create websites and support materials; engage with academics; report to managers; look into changes in to the IR; support; check and add metadata; search for publications we don't know about
- Staffing has (mostly) not adapted to these new requirements, let alone those in the pipeline (REF) – staff still doing the same jobs before this supporting researchers.
- Software hasn't adapted. UK specific issues; software is global. Don't want to reinvent the wheel, especially with risk of getting it wrong.
- Doesn't fit naturally into University structure. Requires all Schools and researchers to comply

Also we are facing increasing demands to provide management reporting, and the need to integrate with the systems - the latter partly due to the IR not containing all the information we need to report on.

Chapter five : help on the horizon

A number of projects are currently in progress which may help us address these issues
- Rioxx - Rioxx is a metadata profile. The plan is, I think, for IRs to expose data about their records according to this profile, and funders will harvest this information to track compliance and for their own reporting. Part of this work will include plugins for the major repositories. Essentially, by complying with Rioxx and its requirements, we are complying with the funder's requirements. However, we need to make sure the data in the IR is stored and entered in such a way that it can be exposed in the correct manner. http://www.rioxx.net/
- Jisc pathfinder projects : e2e: this project specifies the fields that need to be added to Eprints to help comply with the different funders. It should also result in a plugin or similar to the eprints software which will create the required fields. At the moment there are 31 fields that have been identified to ensure we can record all the information


10

http://figshare.com/blog/Loughborough_University_figshare_Arkivum_and_Symplectic_announce_pioneering_rese/136

needed to comply and report for the different funders. That's a lot of fields and raises serious questions around how we can support so many. http://e2eoa.org/

- JISC Monitor : "Jisc Monitor is a 12-month project exploring whether a user-centred, shared national service could potentially help institutions to manage their OA activity effectively. It complements UK projects such as Open Mirror, and others by HEFCE and the research councils, attempting to scope and understand the issues around OA reporting and work up some practical solutions."[11] - should help with managing APC payments and identifying research outputs. http://jiscmonitor.jiscinvolve.org/wp/

- IRUS : Provides COUNTER style reports of downloads of full text. Anyone who has worked with web statistics will know the importance of reliable numbers that filter much of the 'noise' from web reports. And IRUS reports not just on our own data but also of other contributing IRs, allowing us to compare with our peers. http://irus.mimas.ac.uk/
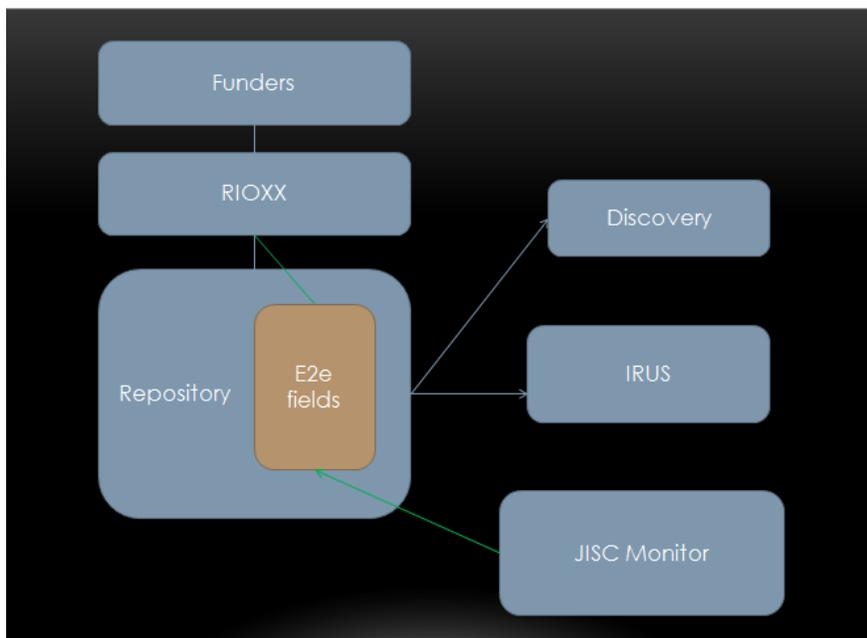
This is how I model the different projects interacting with each other. I see the fields created by e2e as key, as they will hold the information that RIOXX will make available for funders, and which will be populated in part by information found (I think) by JISC Monitor.

11

http://www.jisc.ac.uk/blog/monitoring-and-shaping-the-transition-to-open-access-05-nov-2014

Summary

IRs need to adapt to a new world of funder mandates and reporting and Universities need to adapt and increase their staffing to support the IR, as it takes on a larger and more important role. Though this will be less true for HEI's with other systems.

The REF in particular will require a change in how Universities operate, something most Universities have barely begun to address.

The Library finds itself in a new role. While the Library has always played an important part in campus life, it now finds itself providing critical reporting and processes which the financial (and research reputation) future of the institution may depend on. Making a mistake in a catalogue record is very different to making a mistake in a REF submission.

Meanwhile, Universities are looking at how to support open data and the sharing of data and there are services being developed to help us in this new world.

Finally, funder mandates, especially in combination of CRIS and other Research systems, may actually bring the IR in full circle back to its OA roots.

Chris Keene
University of Sussex
November 2014.