Handling Repository-Related Interoperability Issues: the SONEX Workgroup

Peter Burnhill[1], Pablo de Castro[2] *, Jim Downing[3], Richard Jones[4], and
Mogens Sandfær[5]

[1] EDINA National Data Centre, University of Edinburgh, Edinburgh, UK
[2] Carlos III University Madrid, Leganés (Madrid), Spain
[3] University of Cambridge, Cambridge, UK
[4] Symplectic Ltd, London, UK
[5] Technical University of Denmark (DTU), Copenhagen, Denmark
http://sonexworkgroup.blogspot.com/

Abstract. Sharing of scholarly content through a network of Open Access repositories is becoming commonplace but there is still need for systematic attention into ways to increase the rate of deposit into, and transfer of content across, the OA repository space. This is a report of the work of a small international group, supported by JISC, with remit to describe, analyse and make recommendations on deposit opportunities and use cases that might provide a framework for project activity geared to the ingest of research papers and other scholarly works. The multi-authored, multi-institutional work is put forward as the default, and nine use case actors are listed, as deposit agents, with four main use case scenarios. There is also some comment and pointers to projects in Europe which address some of these use case scenarios.

Keywords: repositories, interoperability, deposit, research output, CRIS, metadata

1 Introduction

The SONEX Group [1] has its origins in a workshop held in Amsterdam in March 2009 that was held to "identify essential components of international repository infrastructure". SONEX is an acronym for Scholarly Output Notification and Exchange, and reflects a focus on interoperability between repositories of all sorts and across many different countries. SONEX has remit to analyze opportunities for deposit of new content into repositories, as well as ways of assisting transfer of content across repositories. The graphic in figure 1 illustrates the variety of repository into which authors deposit/issue their content.

Our purpose is to share comment and recommendations about the implied use case actors (stakeholders) and hence some use case scenarios that might be

---

* Correspondence: Pablo de Castro, Servicio de Recursos Electrónicos, Universidad Carlos III de Madrid, Avda. de la Universidad 30; E-28911 Leganés (Madrid), Spain, tel: +34 916249081, e-mail: pcastro@db.uc3m.es

addressed and tested by other initiatives or projects in their implementation of technology and solutions. We report on work of others whom we know are doing just that.
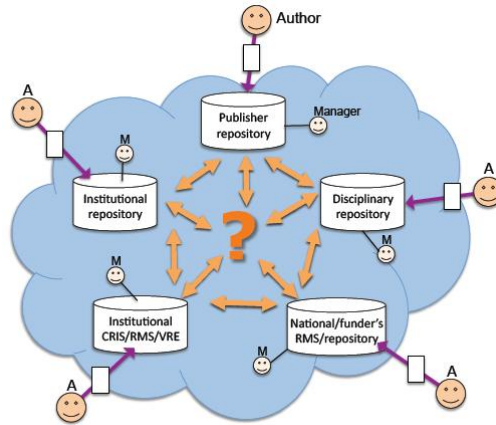
Fig. 1. Interoperability needs within repository workspace

By publishing our analysis and recommendations, via blog, presentations at meetings and through this article, our intention is to assist project managers and developers in their investigation and implementation of enabling (repository) facilities, as well as the Joint Information Systems Committee (JISC) and related funding agencies in implementation of their strategic objectives.

2   Origin and objectives

The International Repositories Workshop in Amsterdam on March 19-20th, 2009 was organised by JISC, the SURF Foundation and the DRIVER Project [2]. The workshop brought together representatives from many of the main projects and initiatives involved worldwide in developing repositories and repository-related services. A series of requirements for building repository infrastructure had been grouped into four main strands to be discussed within seminar groups. There were 'brainstorming' sessions on repository interoperability, citation services, identifiers for authors and institutions, and international repository organisation. Action Plans for workgroups were subsequently established following each of those four sessions and published on a wiki [3].

There were representatives from Denmark, India, Ireland, Israel, Japan, Netherlands, Spain, UK, USA and beyond among the twenty people who took part in the 'Repository Handshake' (RH) session. The consensus reached was the need

to work to a strategy that addressed the low level of content, such as research papers, in most Institutional Repositories (IRs), rather than have detailed technical debate on protocols, although there was general support for using the SWORD (Simple Web-service Offering Repository Deposit) protocol [4]. Key to success in populating repositories was an understanding of the full range of possible opportunities for the deposit of content into repositories, and to characterise what needed to be done at these 'deposit opportunities', by individuals and by machine-as-user, including automation of content ingest.

Some of the basic ideas upon which the SONEX group would later build its analysis are summarized below:

- A focus on automatic mechanism for populating IRs with researchers' scholarly output: PUT (rather than KEEP and GET) functionality through machine-to-machine (m2m) interoperability, whenever possible.
- Consensus is needed on what sufficiently-good-metadata is required at ingest and for onward transfer of object+metadata into repositories; attention should be paid to leveraging existing accurate metadata.
- The technological means for interoperability largely existed: the SWORD protocol was the accepted means to support negotiation between depositing agents and target repositories.
- Dissatisfaction that projects and IRs had their focus on single-author, single-institution research papers; regret that IRs and subject repositories were often viewed as competing alternatives.
- Recognition that CRIS (Current Research Information Systems) and desktop software were sources of content but, as with bibliographic authority files, were presently disconnected from repositories.

The Action Plan [5] drawn up at the end of the workshop sessions called for real-life use cases as exemplars: their analysis on what was common or specific to each; a gap analysis of tools and mechanisms needed; outreach to willing and able partners to test one or two preferred use cases. JISC decided to convene a small group to work on this Action Plan. Intent on retaining the international character of the workshop session in the group, the four individuals[1] invited came from the Danish Technical University (DTU), the Spanish National Research Council (CSIC), and the Universities of Cambridge and Edinburgh (EDINA) [6]. No specific timeline was set for the Action Plan, other than recommending an initial and brief gap analysis in order to prevent risk of conclusions becoming rapidly outdated.

3 SONEX analysis

The main objective of the 'Repository Handshake' workgroup after the Amsterdam workshop was to meet and agree upon a conceptual model and vocabulary.

---

[1] As of Nov'09 the workgroup coordinator changed affiliation from CSIC to Carlos III University Madrid, and a new member with affiliation Symplectic Ltd became part of the group in Apr'10 [7].

We noted that the initial focus in the workshop session on journal articles was fine but concluded that models and infrastructure enabling repositories should be tested for the deposit and sharing of research data and learning materials (and perhaps also open software geared for academic purpose). The acronym SONEX was chosen to reflect that: *Scholarly Output Notification and Exchange.*

We were free from obligation to carry out project work ourselves, although we could prompt and facilitate others to do project activity. Sharing the burden of travel, the SONEX Group held their meetings since 2009 in Cambridge, Copenhagen, Edinburgh, Geneva and Madrid, taking economic advantage of being able to tag onto other repository meetings.

Having defined our problem space as 'metadata+digital object deposit, notification and transfer', we set about examination of system verbs/operations, followed by some further study on the possibilities of communication between repositories. We then reviewed the initial RH use-case scenarios. There were a growing number of projects, funded by JISC or other partners, focusing on populating repositories with some overlap of scope, see table 1 below.

Table 1. Some repository-related projects running at the start of SONEX

Project	*Institution*
SWORD	UKOLN-JISC
Open Access Repository Junction (OA-RJ)	EDINA-JISC
PEER Project	STM, ESF, Göttingen Univ.
CRIS/OAR Interoperability Project	DTU-KE
JournalTOCsAPI project	Heriot-Watt University-JISC
The Deposit Plait	Aberystwyth University-JISC
EM-Loader	EDINA-JISC
EIDeR	King's College London-JISC

An updated version of this listing is available on the SONEX Blog [8], from which the URLs may be actioned.

4 Identifying use-cases from deposit opportunities

The significance of the author is evident in figure 1 above in which one or more authors (A) deposit into an environment of repositories. This array of different kinds of interacting repository systems includes institutional (IRs) and subject-specific and disciplinary repositories (SRs). The managers of each different types of repository have potential interest in any given submission; a given author might be required to make multiple submissions for the same work.

Repositories managed as part of research information systems (CRIS) are alternative sources for ingest or transfer of content into institutional and subject repositories. The inclusion of publishers in the picture is recognition of their significance for peer-reviewed journal articles and for authors, and both necessary

for some types of research output, and also special with respect to interests and rights for related versions of a work, typically the author's final copy and the publisher's version.

Table 2. Deposit agents (SONEX use-case scenarios)

---

*Default content: multi-authored, multi-institutional work (*eg *journal article)*

Use-case Actor 1: Individual (co-)author/researcher
Variant 1a: where the author making the deposit is the PI of funded research project (need for compliance with mandate from funder to deposit)
Variant 1b: the author making the deposit is not the PI of funded research project but the work being deposited is associated with one or more funded research projects (and one or more PIs)

Use-case Actor 2: Depositor (not author) - Delegated deposit
Variant 2a: Mediated by an actor directly reporting to the author
Variant 2b: Mediated by another institutional agent - *eg* library.

Use case Actor 3: Institutional/Departmental Research Support Systems (CRIS/RMS/VRE systems)

Use case Actor 4: Publisher
a) OA deposit of the author's final copy
b) Supply of authoritative metadata and identifiers (DOIs and pointer to published copy)

Use case Actor 5: Funding bodies/Policy bodies

Use case Actor 6: Repository Manager (RM) of an IR, with co-authored work wishing to notify RM(s) of the other IR(s); RM of subject (SR) wishing to do similarly; RMs of IRs wishing to know of and obtain copy of work by author (now) at institution.

Use case Actor 7: Developer/vendor of authoring software

Use case Actor 8: Repository software developer/vendor

Use case Actor 9: Libraries
Potential depositors of their own resources and document collections

---

From the start we wished to promote the SONEX default of the 'multi-authored, multi-institution' work. We would urge projects and systems not to plan on the basis that research papers are generally from a single author, nor that a given research paper is relevant to only a single institution. By default any deposit might be of interest to more than one institution. Whatever a repository

manager holds is potentially of interest to another - researchers and authors move, and for assessment purposes past publications count. One the other side of the coin, there is prospect of multiple issue of the same work by its several co-authors.

All this provides real motivation for exchange of information, and content, across repository space: across institutional and national boundaries as well as across repositories of different types.

Our task was to complete the list of actor-based use-cases produced during the Repository Handshake workshop sessions. The most important persons in discussions about Open Access, and also for plans of repository infrastructure and enabling projects, are the researchers/authors and the readers: both wish for prompt and ready release or the author's work, for attention by her/his peers and the world beyond the academy. It is for all intermediaries to assist that goal. The extended list, based on the deposit opportunities illustrated in figure 1, is given above in table 2.

Reflection on these use cases helped identify four principal use case scenarios for further analysis. Use case scenarios would also serve as basis on which to review which projects could test them through implementation:

I **CRIS/RMS/VRE systems** (at institution or national levels) [use case actor 3] where transfer of objects plus agreed metadata into all relevant IRs should be automated. [Associated Projects: Trinity College Dublin (TCD), Technical University of Denmark (DTU)]

II **Publisher** [use case actor 4] as publisher wishes to deliver service to author, by depositing the (co-)author's final copy in appropriate IR, complete with DOI and pointer to the published version. [Associated Projects: Nature Publishing Group (NPG) + Repository Junction (OA-RJ) Project; many publishers in PEER Project]

III **Funder-mandated deposit** [use case actor 5] wishes assurance of compliance that output from funded projects is deposited under OA with relevant SR and/or IRs. [May also involve use case actors 1, 2 and 3] [Associated Projects: Repository Junction (OA-RJ) Project with UKPMC].

IV **Deposit via personal software**, by researcher/co-author [use case actor 1] e.g. desktop software [use case actor 7] or bibliography (web or desktop).

5   Use case scenario I: CRIS systems as source of content

An increasing number of universities in Europe are investing in some kind of Current Research Information System (CRIS) in order to support research groups and to keep track of their research output. These systems can hold information

on a variety of research-related activity such as projects, grants, PhDs, patents, often relating to obligation to funders and (national) systems of research assessment [9]. They include metadata on research publication but rarely the full text of the published work.

There was first-hand knowledge of the *CRIS/OAR Interoperability Project* within the SONEX Group, developed by Knowledge Exchange and the Danish Technical University (DTU). This project had the objective to "increase the practical interoperability between Current Research Information Systems (CRIS) and Open Access Repository (OAR) systems by defining and proposing a metadata exchange format for publication information with an associated common vocabulary". A CRIS/OAR interoperability workshop was held last June at the CRIS2010 conference in Aalborg, Denmark, in which the KE group presented its results [10]. Several CERIF-based initiatives, both institutional and national, were also presented on metadata exchange and CRIS/IR integration [11]. The high number of projects and a growing commercial engagement in the area indicates that the CRIS/IR interoperability use-case scenario is being successfully addressed.

6   Use case scenario II: Publishers as use communities

Publishers are the recipients of the author's final copy, prior to the publication of the publisher's version, a principal source for deposit of that copy into the institutional repository. Nominally, there may appear to be a conflict of business with publishers, which are seen as wishing to have paid-for attention to the copy that they publish. However, publishers rely on the free supply of content from authors, and wish to do service to authors by exploring ways in which they can assist authors comply with mandates from funders that the author's final copy should be deposited in Open Access repositories. They can act as an appointed proxy for the co-authors of an article. Several projects exist, two of which are reported here: the EU-funded PEER Project and JISC-funded Open Access Repository Junction (OA-RJ) Project.

The **PEER Project** (Publishing and the Ecology of European Research) [12], began in October 2008 with support from the European Commission eContent+ Programme. PEER aims to investigate the effects of the large-scale, systematic deposit of authors' final peer-reviewed manuscripts (also known as post-print versions) upon reader access, author visibility, and journal viability, as well as on the broader ecology of European research. The project is a collaboration between publishers in the International Association of Scientific, Technical and Medical Publishers (STM) and a number of repositories belonging to partner academic and research institutions (Göttingen State and University Library, the Max Planck Society, University of Bielefeld, INRIA, Kaunas University of Technology, University Library of Debrecen and Koninklijke Bibliotheek Amsterdam). In carrying out the publication transfer, PEER intends to unify the ingestion services based either on format used or protocols such as OAI-PMH or SWORD.

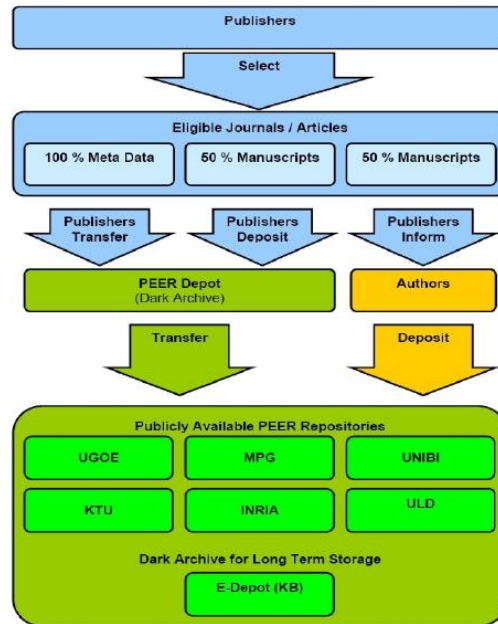The PEER workflow is represented in figure 2 below.

Fig. 2. PEER workflow for journal articles, from publisher to OA repository

The JISC-supported **Open Access Repository Junction** (OA-RJ) Project, which started in August 2009 has a generic approach to automatic ingest into a broker which then manages notification and transfer of metadata, with or without content, to target repositories. This generic deposit broker, which deploys SWORD and makes use of OpenDOAR and ROAR, is designed to work for all types of content and in one strand is being applied to research papers in work with Nature Publishing Group and a group of test target repositories as part of a test implementation for use case scenario II. By interacting with the business needs of the publisher, an understanding and a mechanism has been achieved for a two-stage workflow, as shown in figure 3 below.

This shows a two-stage engagement which is sympathetic to the opportunity that publishers have to enhance the author's own copy with additional metadata derived from the publisher's own version (e.g. DOI and URL to the publisher's copy), and to do so with bibliographic accuracy. The intention is to generalise this so that any journal publisher might initiate deposit of the co-authors' final copy into each of the respective institutional repositories. More about this is reported on the OA-RJ Blog [13].
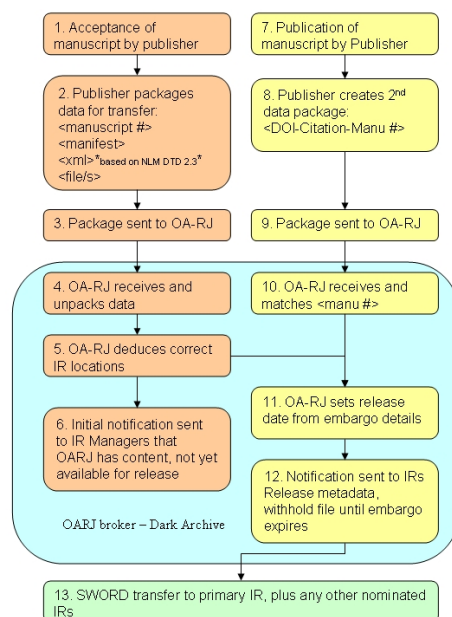
Fig. 3. OA-RJ workflow for journal articles, from publisher to OA repository

We have also noted other more recent initiatives, such as the BioMed Central partnership with MIT Libraries to deposit open access articles using Sword [14], via a paid for service. These follow similar deposit patterns and workflow for research output transfer into repositories, so some common guidelines for publisher-driven deposit arising from PEER and OA-RJ could be very useful to the repository community. This is one of the lines for SONEX current work.

7   Use case scenario III: Funders and subject repositories as use communities

As research councils and other research funders turn to mandates for the issue of research work into Open Access repositories, so there is need for facility such as that being developed as the OA-RJ broker[2]. This is being tested by OA-RJ with UKPubMed Central, which is a subject repository supported by multiple funders in the bio-medical area. This will engage with both the repository manager for a subject repository, use case actor 6, but also the funder, use case actor 5, together with routing of notification to repository managers of the respective IRs.

---

[2] An earlier version of the OA-RJ broker is used for the Depot [15], which helps authors self-deposit by re-directing the author to her/his institutional repository; it will be used to support OpenDepot.org [16] as a global facility for both affiliated and non-affiliated researchers who wish to release their research work as Open Access.

The OA-RJ broker architecture lends itself very well to networking and SONEX is investigating how this may be tested in practice. Scalability seems well served by a distributed approach (see figure 4 below) where national/regional brokers on the one hand are responsible for a select group of publishers/subject repositories and on the other hand for their national/regional network of IRs. Whenever such a networked OA-RJ broker receives information on publications relevant to IRs of another nation/region it transfers the information to the respective broker. Thus the networked OA-RJ brokers may share the work of establishing relationships and technical interfaces with the world's publishers and subject repositories - and these will only have to deal with one broker to the world of institutional repositories. Likewise each broker only needs to maintain detailed information on and relations with its own national or regional group of IRs.
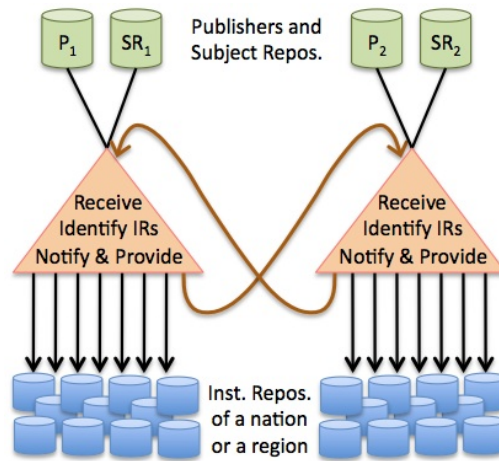

Fig. 4. An array of networked national/regional brokers

8   Looking to the future: JISC deposit call

We have so far presented much of what we set out to do as deposit use-cases review of project activity. We now look to outreach and to helping create an international space for exchange of information about interoperability issues with regard to repositories.

Strand A of a recent funding call by JISC (Funding Call 2/10, March 2010) had the specific objective of "ensuring take-up of solutions that enable and encourage author deposit of Open Access research outputs into repositories by embedding deposit into research or related practice" [17]. The outcome of this

funding call represents an opportunity for the SONEX Group to assist the selected projects by (i) providing a use case scenario framework and (ii) providing webspace/poster-space for deposit initiatives[3].

Three projects [18] were announced in July 2010 at the Open Repositories Conference held in Madrid, and representatives of each attended a SONEX Deposit meeting along OR10 to give a briefing and examine how they relate to the SONEX use case scenarios.

- **DepositMO: Modus Operandi for Repository Deposits** [19]. Led by the University of Southampton, in close liaison with Microsoft and the University of Edinburgh, the DepositMO project aims to create a workflow connecting the user's computer desktop: part of use case scenario IV. This involves interoperability between MS Office and the EPrints and DSpace repository software [use case actor 7]. The DepositMO project further tests the Sonex use-case analysis with content such as datasets and software, as well as a real-life exemplar for software vendor-driven deposit.

- **RePosit: positing a new kind of deposit** [20]. The RePosit Project seeks to increase uptake of a web-based repository deposit tool embedded in a researcher-facing publications management system. Led by the University of Leeds, it involves several additional universities and Symplectic Ltd as commercial partner. This project further tests SONEX use case scenario I, with a variety of starting points and institutional strategies for research information system.

- **DURA: Direct User Repository Access** [21]. The DURA project, lead by the University of Cambridge with Mendeley Ltd and Symplectic Ltd as consultant firms, aims to embed institutional deposit into the academic workflow by using Mendeley and Symplectic tools to allow researchers to synchronise their personal research collections with institutional systems. This also addresses SONEX use case scenario IV, which had previously been represented as *Our Bibliography tools* as presented on the Repositories Infrastructure wiki [22].

It is time now to re-visit and update our use case framework in the light of these new projects and with more than just research papers in mind. For this we need feedback and criticism, and would welcome both.

Acknowledgements

---

[3] Suggestion for providing a SONEX webspace on deposit initiatives was followed by the creation of the SONEX blog in Apr'10 for publication of the workgroup files.

References

1. SONEX workgroup for Scholarly Output Notification and Exchange blog,
   `http://sonexworkgroup.blogspot.com/`
2. International Repositories Workshop (Amsterdam, The Netherlands, Mar 16-17,
   2009), `http://www.ukoln.ac.uk/events/ir-workshop-2009/`
3. International Repositories Infrastructure wiki, `http://repinf.pbworks.com/`
4. SWORD: Simple Web-service Offering Repository Deposit. SWORD2 Project
   Final Report, `http://www.jisc.ac.uk/media/documents/programmes/reppres/`
   `sword2finalreport.pdf`
5. Action plan for Repository Handshake at the International Repositories Infrastruc-
   ture wiki, `http://repinf.pbworks.com/Repository-handshake`
6. SONEX workgroup members and affiliations, `http://sonexworkgroup.blogspot.`
   `com/2010/04/workgroup-participants-and-affiliation.html`
7. "Richard Jones joins SONEX workgroup", `http://sonexworkgroup.blogspot.`
   `com/2010/04/richard-jones-joins-sonex-workgroup.html`
8. "A summary of ongoing deposit-related projects", `http://sonexworkgroup.`
   `blogspot.com/2010/05/summary-of-ongoing-deposit-related_9901.html`
9. Grace, S., Gartner, R.: Modelling national research assessments in CERIF. In:
   CRIS2010, Jun 2-5, Aalborg, Denmark,
   `http://www.cris2010.org/index.php?SID=ceacd8d1ab7c5e930f435b7991ca8fa5&action=`
   `showPaper&from=0&id=27`
10. Elbæk, M.K., Sandfær, M., Simmons, E.: CRIS/OAR interoperability workshop.
    In: CRIS2010, Jun 2-5, Aalborg, Denmark,
    `http://www.cris2010.org/index.php?SID=ceacd8d1ab7c5e930f435b7991ca8fa5&action=`
    `showPaper&from=0&id=27`
11. "CRIS2010 Aalborg: a brief report", `http://sonexworkgroup.blogspot.com/`
    `2010/06/cris2010-aalborg-brief-report.html`
12. The PEER Project, `http://www.peerproject.eu/`
13. Open Access Repository Junction (OA-RJ) Project blog,
    `http://oarepojunction.wordpress.com/`
14. "BioMed Central partners with Massachusetts Institute of Technology Libraries
    to deposit open access articles automatically using SWORD protocol", BMC Press
    Release,
    `http://www.biomedcentral.com/info/presscenter/pressreleases?pr=`
    `20100429`
15. The Depot, `http://depot.edina.ac.uk/about/`
16. OpenDepot.org, `http://opendepot.org/`
17. JISC Grant Funding Call 2/10: Deposit of research outputs and Exposing digital
    content for education and research,
    `http://www.jisc.ac.uk/fundingopportunities/funding_calls/2010/03/`
    `210depositexpose.aspx`
18. First projects selected at JISC Deposit Call, `http://sonexworkgroup.blogspot.`
    `com/2010/07/first-projects-selected-at-jiscdepo.html`
19. DepositMO: Modus Operandi for Repository Deposits, `http://blogs.ecs.soton.`
    `ac.uk/depositmo/`
20. RePosit: positing a new kind of deposit, `http://jiscreposit.blogspot.com/`
21. DURA: Direct User Repository Access, `http://jisc-dura.blogspot.com/`
22. SONEX Deposit Opportunities: a Poster,
    `https://repinf.pbworks.com/f/poster_SONEX_deposit_opportunities.pdf`