Repositories of past, present and future

Nordic Perspectives on Open Access and Open Science, Helsinki 15.10.2013

Jyrki Ilva (jyrki.ilva@helsinki.fi)
The purpose of repositories

- Wikipedia: “An institutional repository is an online locus for collecting, preserving, and disseminating - in digital form - the intellectual output of an institution, particularly a research institution”
- The repositories are often seen as synonymous with Green OA
  - “They are primarily about Green OA”
  - “Even if they are not, they should be about Green OA”
- However, it’s not quite as simple as that
Is there more to repositories than Green OA?

- Kennison and Shreeves point out that there are thriving institutional repositories built on other kinds of content
- Having a successful institutional repository makes it easier to work towards a mandate for scholarly articles as well
- Harnad: Green OA to refereed articles is urgent; everything else is secondary
- Instead of going after "low-hanging fruit", energies should be concentrated on achieving the right kind of mandate for refereed articles (and its enforcement)
Looking back: Where did it all come from?

- Although "institutional repository" as a distinct concept may have its origins in the Green OA movement, the real-world repositories have a much more complicated background.
- The repository as we know it now is a hybrid born over time out of or in contact with different motivations:
  - CRIS – current research information systems
  - DOMS – management of digital objects
  - ETDs – electronic theses and dissertations
  - OA books and serial publications
  - Green OA – self-archiving of scientific articles/publications
  - Open research data
CRIS – Current research information system

- A system designed to serve the needs of research administration and research evaluation
  - Publications key criteria in evaluation
- Basic difference: CRIS contains information on all research publications, IR contains only Open Access full-text publications
- Although the mission of CRIS differs from that of IR, there has been a lot of effort spent on integrating these two types of systems and their workflows
Doms – the management of digital objects

- In the late 1990s and early 2000s research libraries were looking for a way to manage their digital collections
  - Digital Object Management System (Doms): a technical platform to handle the growing number of digital materials generated and/or maintained by the libraries
- Connected to a larger library infrastructure, including discovery systems and library catalogs
  - Possibly connected to a digital preservation system and its workflows
ETDs – Electronic theses and dissertations

- The first ETD projects in the 1990s
  - An international movement, but the Nordic tradition of publishing printed dissertations ensured that we were among the pioneers
- Providing access to collections of student theses also a primary concern in many university libraries
- The starting points generally practical, not ideological
  - Open access sometimes chosen mainly because there seemed to be no way to build viable commercial services for ETDs
- Contrast: the role of UMI/ProQuest in the US
OA books and serial publications

- Open access books and serial publications published by the organization itself
  - Important category for research institutes as well
- As with ETDs, open access started to replace the traditional system of exchanging printed publications internationally with other research organizations
Green OA: Self-archiving and repositories

- Solving several problems at once
  - The traditional distribution model restricts access to scientific information
  - The publishing process slows down the spread of information
  - The Big Deal: rising cost of licensing for libraries
- Subject-based repositories (incl. ArXiv) and pre-prints
  - Did not (and do not) cover all of the scientific fields
- The concept of institutional repository
  - Early belief: ”Build it and they will come”
  - After a few years it was noted that it isn’t quite that easy: OA generally not first priority for scholars
  - Organizational and funder mandates needed to encourage them
Repository platforms: open source

- One of the surprising achievements of the OA movement was the introduction of Open Source software in many libraries
  - The traditional library software vendors were not able to deliver credible Digital Object Management Systems
  - The new generation of proprietary CRIS platforms (with IR-like functions) was still several years away
- The leading repository platforms (DSpace, Fedora, Eprints) were adopted for repository use and often for other purposes as well
  - Based on work done by international developer communities
  - Gradually replaced most of the older self-built systems
Aggregation and discovery

- The user interface of the institutional repository has mainly local importance (for the users from the own organization)

- Discovery on global level was planned to be based on metadata harvesting
  - Harvesting protocols: OAI-PMH, later OAI-ORE & ResourceSync

- OAI-PMH-based search engines specializing on scholarly content (Oaister, BASE) have not really taken over the world
  - Google and Google Scholar dominate in the discovery of repository content
  - Repository content has had very good visibility in search results especially after Google started indexing PDFs in 2001
Shared repository services?

- The original vision: institutional repositories built on a local level and their metadata harvested to portals and search engines
  - As a result we have a large global network of mostly separately-hosted repository software instances
- Nordic countries one of the partial exceptions
  - Finland: nearly 40 customer organizations use the centralized repository infrastructure provided by the National Library
  - Norway: Bibsys provides repository services for many organizations
  - Sweden: University of Uppsala (DiVA) with a consortium of 30+ organizations
- In each of the countries there are also organizations (often big universities) that are using locally-hosted repositories
Current situation

- Just about all of the Nordic universities already have a repository, but there is still room for expansion on other sectors
- The number of items in repositories is growing steadily
  - Repositories have a strong position in providing open access to some content types, including ETDs and serial publications
  - In some other types (including refereed articles) there is still a lot of work to do
- Much of the repository content is heavily used, largely thanks to search engine visibility
Repository and CRIS – more on integration

- Repository is not an island: integration with other systems and workflows essential for its success
- Integration of IRs with CRIS is growing ever more important
  - Denmark: many of the universities are using the same platform (Pure) for both purposes
  - Sweden: DiVA is used both as an IR and as a CRIS for publications; information from IRs harvested to Swepub portal
  - Norway: Integration on a national level in CRIStin
  - Finland: Repositories and CRIS are for the most part built separately on different platforms, the universities are currently investing on the renewal of the local CRIS infrastructure
New content types: open research data

- Focus seems to be shifting from publications to open research data
- Funding available and new services being developed specifically for research data
- To what extent can this be integrated with or connected to the existing repository infrastructures?
- Possibilities for co-operation in advocacy and linking publications and data into each other
Repositories as a library service?

- Discussion on the merits of Green and Gold OA has intensified
  - New policies and recommendations on the EU level
  - UK: Finch report, RCUK policy
  - US: Share vs. Chorus – two competing initiatives
- Will the future of Open Access be based on the repository services provided by the libraries, or on an infrastructure developed and run by the publishers? Or both?
- The role of the repository as part of future library services?
  - Providing open access to documents produced within the organization (instead of providing local users access to content produced elsewhere)
- Relationship with cloud-based social networking services (e.g. Academia.edu, Mendeley) aimed at scholars?
Open repositories 2014

- The next Open Repositories conference will be in Helsinki, June 9-13, 2014
  - A five-day conference hosted by the National Library of Finland and Helsinki University Library
  - [http://or2014.helsinki.fi](http://or2014.helsinki.fi)
- The leading international conference on repositories, for the first time in the Nordic countries
  - Excellent opportunity to learn what is going on and what is being discussed in the global repository community!