

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
 - Kennison, R, Shreeves, SL, Harnad, S. (2013). Point & Counterpoint: The Purpose of Institutional Repositories: Green OA or Beyond?. *Journal of Librarianship and Scholarly Communication* 1(4):eP1105. <http://dx.doi.org/10.7710/2162-3309.1105>

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)



The screenshot shows a digital repository interface. At the top, there is a navigation bar with the JYX logo and text: "Jyväskylän yliopiston julkaisutietokeskus" and "Jyväskylä University Digital Archive". Below this, there are links for "Login", "Send Feedback", "User's Guide", and language options "en" and "fi". The main content area displays the title "Ranking lists and European Framework Programmes: Does university status matter for performance in Framework Programmes?". To the left of the title is a PDF icon. Below the title, there is a citation box containing the following text: "Nokkala, T., Heller-Schuh, B., & Paier, M. (2011). Ranking lists and European Framework Programmes: Does university status matter for performance in Framework Programmes?. In P. Teixeira, & D. Dill (Eds.), Public Vices, Private Virtues? Assessing the Effects of Marketization in Higher education. (pp. 111-139). Rotterdam: Sense Publishers." To the right of the citation box, there is a list of "Related publications" and "More from authors". The "Related publications" list includes: "The use of metadata in network management system as a social action (2000)", "Hengen hehkusta tietostrategioihin : Jyväskylän yliopiston humanistisen tiedekunnan viisi vuosikymmentä (2008)", "Evaluation of the university sport services for international students : a case study from the university of Jyväskylä (2007)", "Puheviestintä - taitoaineesta tieteenalaksi (2012)", and "Mitt i brytpunkten: Finska universitetsstudenters åsikter om svenskan (2009)". The "More from authors" list includes: "Institutional Autonomy and the Attractiveness of the European Higher Education Area Facts or Tokenistic Discourse? (2012)".

Name: nokkalaetalcherbo ...
Size: 626.8Kb
Format: PDF
Description: Author's Final draft
Downloads: 108

Citation:

Nokkala, T., Heller-Schuh, B., & Paier, M. (2011). Ranking lists and European Framework Programmes: Does university status matter for performance in Framework Programmes?. In P. Teixeira, & D. Dill (Eds.), Public Vices, Private Virtues? Assessing the Effects of Marketization in Higher education. (pp. 111-139). Rotterdam: Sense Publishers.

Title: Ranking lists and European Framework Programmes: Does university status matter for performance in Framework Programmes?

Author: Nokkala, Terhi; Heller-Schuh, Barbara; Paier, Manfred

Abstract: The operational context for higher education institutions has become increasingly competitive: universities have to compete on national and international markets for students, staff, funding and prestige. In this context, universities have increasingly become to think of themselves as actors who are in direct competition with others, and adapt their strategies to increase their status and survive in the new environment. The possibly best-known approach to measuring the status of universities is ...

ISBN: 978-94-6091-464-5

Publisher: Sense Publishers

Date: 2011

Belongs to series: Public Vices, Private Virtues? Assessing the Effects of Marketization in Higher education/P. Teixeira & D. Dill (eds.)

Subjects: yliopisto ranking-lista puiteohjelma suoriutuminen university ranking Framework Programme performance

Rights: © Sense Publishers. This is an electronic final draft version of an article whose final and definitive form has been published in the book "Public Vices, Private Virtues?" by Sense Publishers.

Original source: <https://www.sensepublishers.com/files/9789460914669PR.pdf>

Permanent link to this item: <http://urn.fi/URN:NBN:fi:jvu-201205171676>

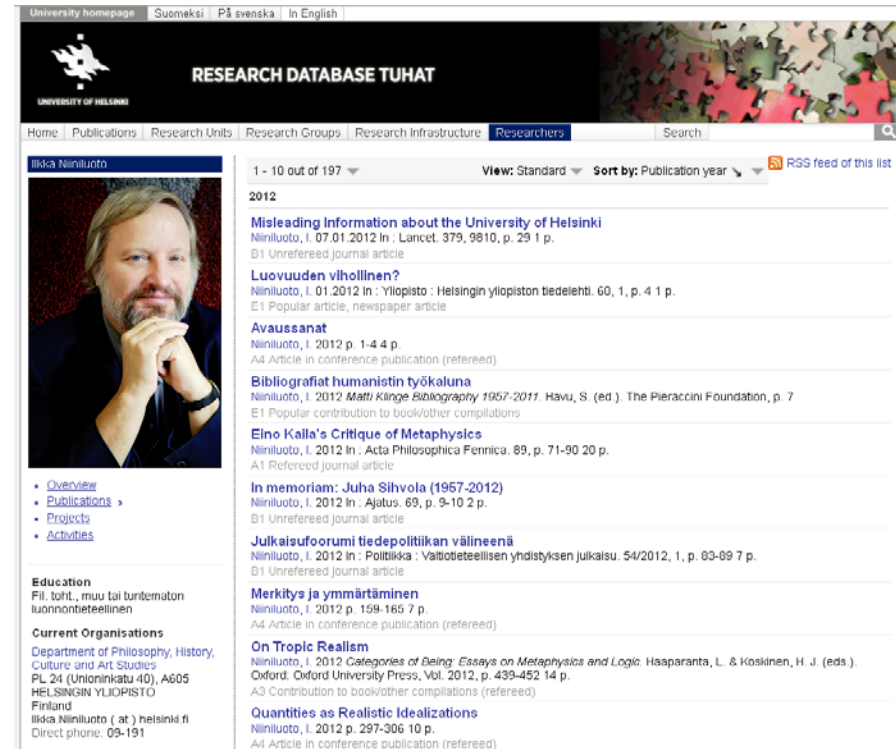
[Show full item record](#)

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



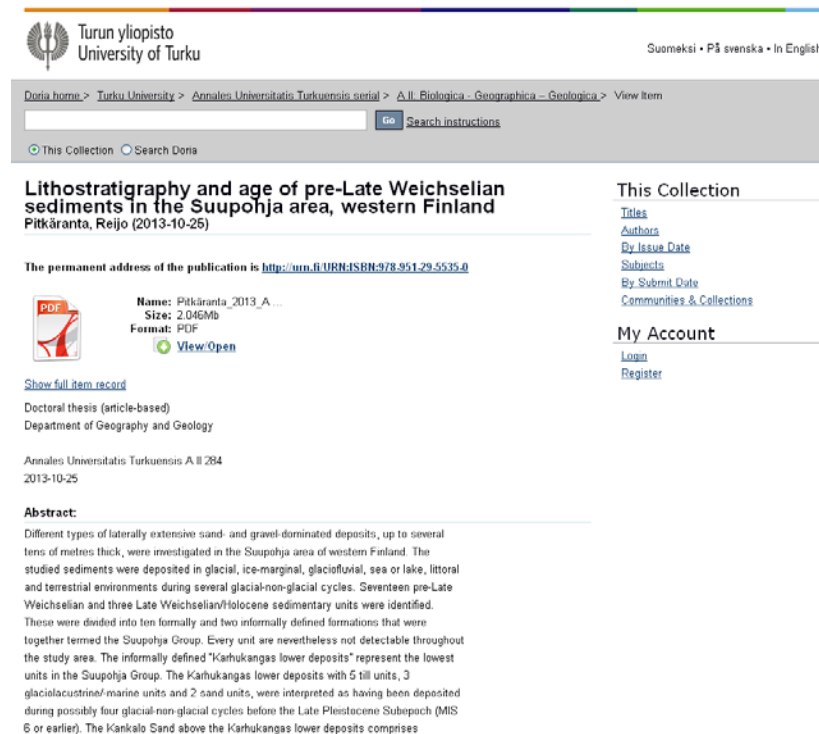
The screenshot displays the 'RESEARCH DATABASE TUHAT' interface for the University of Helsinki. It features a navigation bar with links for Home, Publications, Research Units, Research Groups, Research Infrastructure, and Researchers. A search bar is located on the right. The main content area shows a profile for Ilkka Niiniluoto, including a portrait photo and a list of publications. The publications list includes titles such as 'Misleading Information about the University of Helsinki', 'Luovuuden vihollinen?', 'Avaussanat', 'Bibliografiat humanistin työkaluna', 'Eino Kaila's Critique of Metaphysics', 'In memoriam: Juha Sihvola (1957-2012)', 'Julkaisufoorumi tiedepolitiikan välineenä', and 'Merkitys ja ymmärtäminen'. Each entry includes the journal name, date, and page numbers.

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

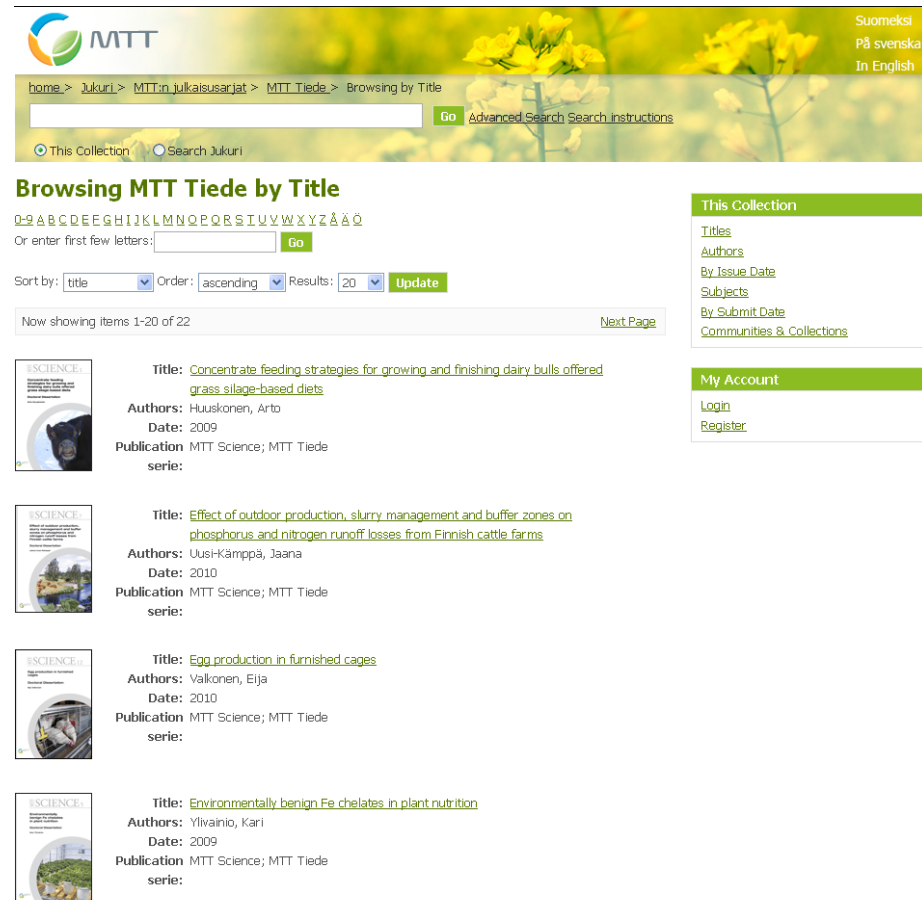
- The first ETD projects in the 1990s
 - An international movement, but the Nordic tradition of publishing printed dissertations ensured that we were among the pioneers
- Providing access to collections of student theses also a primary concern in many university libraries
- The starting points generally 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



The screenshot displays the digital library interface of the University of Turku. At the top, the university's logo and name are visible, along with language options: 'Suomeksi • På svenska • In English'. Below this is a navigation breadcrumb: 'Doria home > Turku University > Annales Universitatis Turkuensis serial > A II. Biologica - Geographica - Geologia > View Item'. A search bar with a 'Go' button and 'Search instructions' link is present. The main content area features the title 'Lithostratigraphy and age of pre-Late Weichselian sediments in the Suopohja area, western Finland Pitkäranta, Reijo (2013-10-25)'. A permanent address is provided: 'http://urn.fi/URN:ISBN:978-951-29-5535-0'. A PDF icon is shown with details: 'Name: Pitkäranta_2013_A...', 'Size: 2.046Mb', and 'Format: PDF'. A 'View/Open' button is available. Below the document information, it is identified as a 'Doctoral thesis (article-based)' from the 'Department of Geography and Geology'. The publication is cited as 'Annales Universitatis Turkuensis A II 284 2013-10-25'. An 'Abstract:' section follows, describing the study of laterally extensive sand- and gravel-dominated deposits in the Suopohja area of western Finland, detailing the glacial and terrestrial environments and the identification of pre-Late Weichselian and Late Weichselian/Pliocene sedimentary units.

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



The screenshot shows the MTT Tiede website interface. At the top, there is a navigation bar with the MTT logo and language options (Suomeksi, På svenska, In English). Below the navigation bar is a search bar with a 'Go' button and a link to 'Advanced Search Search instructions'. The main content area is titled 'Browsing MTT Tiede by Title' and features a search bar with a 'Go' button. Below the search bar are sorting options: 'Sort by: title', 'Order: ascending', and 'Results: 20', with an 'Update' button. The page indicates 'Now showing items 1-20 of 22' and has a 'Next Page' link. The list of publications includes:

- Title:** [Concentrate feeding strategies for growing and finishing dairy bulls offered grass silage-based diets](#)
- Authors:** Huuskonen, Arto
- Date:** 2009
- Publication:** MTT Science; MTT Tiede serie:

- Title:** [Effect of outdoor production, slurry management and buffer zones on phosphorus and nitrogen runoff losses from Finnish cattle farms](#)
- Authors:** Uus-Kämpö, Jaana
- Date:** 2010
- Publication:** MTT Science; MTT Tiede serie:

- Title:** [Egg production in furnished cages](#)
- Authors:** Valkonen, Eija
- Date:** 2010
- Publication:** MTT Science; MTT Tiede serie:

- Title:** [Environmentally benign Fe chelates in plant nutrition](#)
- Authors:** Ylivainio, Kari
- Date:** 2009
- Publication:** MTT Science; MTT Tiede serie:

On the right side of the page, there are two green navigation menus. The first is 'This Collection' with links for 'Titles', 'Authors', 'By Issue Date', 'Subjects', 'By Submit Date', and 'Communities & Collections'. The second is 'My Account' with links for 'Login' and 'Register'.

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

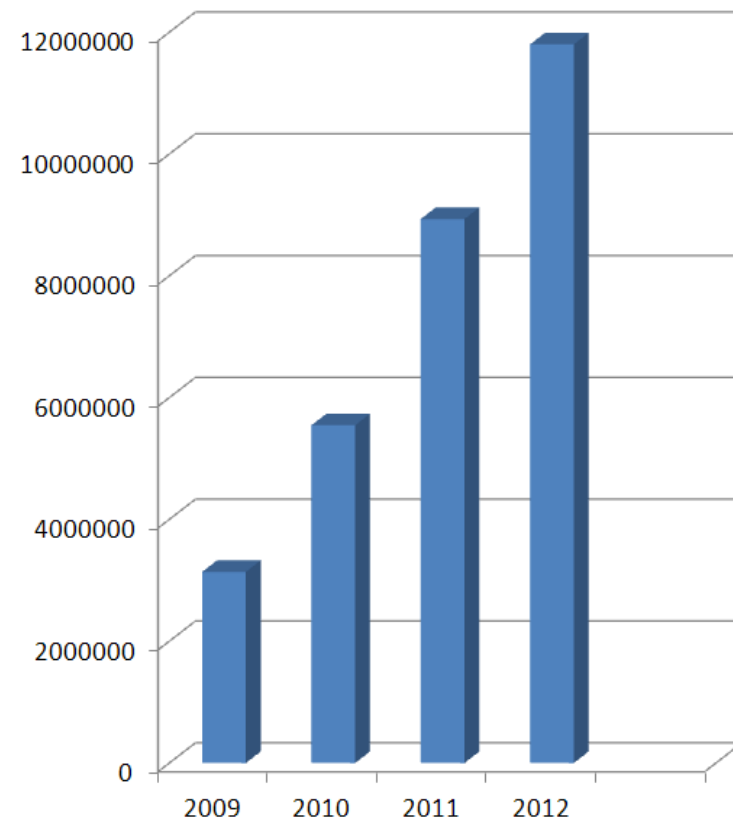
Shared repository services?

- 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

Full-text downloads at the National Library repository infrastructure



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 CRISin
 - 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

The screenshot displays the Dryad Digital Repository interface. At the top, the Dryad logo is visible alongside navigation links for 'About', 'For researchers', 'For organizations', 'Contact us', 'Login', and 'Sign Up'. A featured data package is highlighted with the title 'Data from: Using genomic tools to maintain diversity and fitness in conservation programmes' and a thumbnail image from the journal 'MOLECULAR ECOLOGY'. A green button labeled 'Submit data now' and a link 'How and why?' are present. Below this, a search bar is available with a 'Go' button and a link to 'Advanced search'. A section titled 'Be part of Dryad' includes links for 'Membership', 'Submission integration', and 'Pricing plans'. The main content area, 'Files in this package', contains a table with the following details:

| Title | Codes and data for "Using genomic tools to maintain diversity and fitness in conservation programmes" |
|-------------|---|
| Downloaded | 11 times |
| Description | Files included in file DryaData.tar.gz : BigPopMSD.f90, genotiposCGD.dat, OCgen_ranmat.f90, OCroh_ranmat.f90, DistFreqsSeg_overReps.f90, genotiposMukai.dat, OCmol_ranmat.f90. Genotype data: genotiposCGD.dat and genotiposMukai.dat are obtained after 10000 generations of mutation-selection-drift with BigPopMSD.f90 for the CGD and the Mukai scenarios, respectively. Population size is kept constant at 1000 diploid individuals (500 females and 500 males), and the genome is assumed to have 20 chromosomes of 1 Morgan each. Every chromosome includes 2000 neutral loci, 1000 selected loci and 1000 marker loci, all of them biallelic. For the CGD scenario, the parameters used are lambda is 0.03, beta is 2.3, mean s is 0.264, and mean h is 0.2. For the Mukai scenario, the parameters used are lambda equals 0.5, beta 1, mean s is 0.05, and mean h equals 0.3. BigPopMSD.f90 generates the base population. OCgen_ranmat.f90 performs 10 generations of optimal contributions for population management with random matings between the individuals who contribute. Contributions are calculated to optimise genealogical coancestry, which is calculated assuming the founder individuals are unrelated. The number of replicates is 100, but can be varied, as well as the population size during management. OCmol_ranmat.f90 performs 10 generations of optimal contributions for population management with random matings between the individuals who contribute. Contributions are calculated to optimise molecular coancestry, calculated as identity by state (see Toro et al 2002, Conservation Genetics). The number of replicates is 100, but can be varied, as well as the population size during management. OCroh_ranmat.f90 performs 10 generations of optimal contributions for population management with random matings between the individuals who contribute. Contributions are calculated to optimise IBD-based coancestry, calculated as the proportion of the genome that is identical by descent between individuals. It is calculated by looking only at marker loci, and it requires a minimum length of a segment for it to be |

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>
- 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!





THE NATIONAL LIBRARY OF FINLAND – Library Network Services