

Repository content – lost in discovery?

Open Repositories, Helsinki, June 12, 2014

Jyrki Ilva (jyrki.ilva@helsinki.fi)

Discovery systems

- The basic idea: giving the patrons a single search interface into content (both local and non-local) provided by the library

Repositories and discovery

- How does repository content show up in discovery systems, or does it show up at all?

Discovery in Finland

- A national discovery system, Finna (www.finna.fi) for Finnish archives, libraries and museums
- Part of the National Digital Library Project (KDK, www.kdk.fi/en) funded by the Ministry of Education and Culture

Search Organisations About Finna Instructions Contact us Login Suomi Svenska

FINNA

Find the treasures of Finnish archives, libraries and museums with a single search
Currently 8 791 473 entries included!

Find... Whole Index [Search] [Advanced search]

For seekers of information and inspiration
The Finna information search service brings together the collections of Finnish archives, libraries and museums.
New content is continuously added to the service, which can also be used to browse and read from thousands of electronic resources.

With the search you can find...

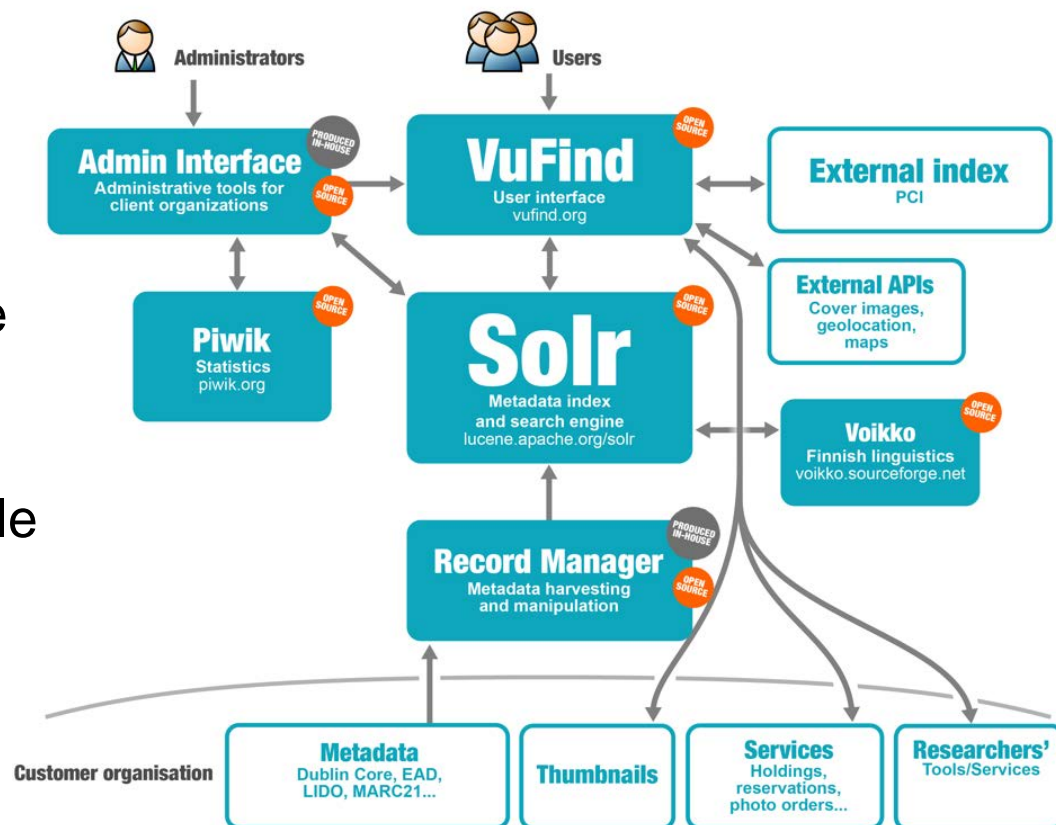
- Journals and articles
- Books
- Images
- Works of art
- Physical objects
- Maps
- Documents
- Theses
- Sound recordings
- Musical scores
- Videos
- Places

Collection highlights

- Hvitträsk – a monument of Finnish romantic nationalism
- Pekka Halonen: Tomatoes (1913). Ateneum Art Museum
- Constantine of Africa: Theorica Pantegni
- Archive of the State Treasury office of compensation affairs

Development of Finna

- Developed by the National Library of Finland
- Based on VuFind and a number of other tools
- Lots of good work done, see <https://github.com/KDK-Alli>
- Provides both a national interface for publicly available content and also separate instances for individual organizations



Using Finna on an organizational level (1)

- Helsinki City Museum, <https://hkm.finna.fi/>

Home About Search Help Contact us Login Suomi Svenska

Helsingin kaupunginmuseo

FINNA

Find... Whole index Advanced search

Anything (no search term)

Search results 1 - 100 / 906 Sort Relevance Results per page 100

Narrow search

Available online

Content type: Image

Author: Brander Signe HKM

Content type

Image (906)

Author

Brander Signe HKM

Kaisaniemen lammikko.

Kaisaniemi.

Kaisaniemen lammikko. Taustalla

Kaisaniemen lammikko.Taustalla

Kaisaniemi. Taustalla
Kaisaniemenlahti.

Kaisaniemen joutsenlampi.

Kaisaniemen joutsenlampi.

Munkkisaari ja Munkkisaaren salmi

Using Finna on an organizational level (2)

- University of Jyväskylä, <https://jyu.finna.fi/jykdok/>
- Designed to replace the university level OPAC and meta-search tool

The screenshot displays the JYKDOK search interface. At the top, there is a navigation bar with links for 'About', 'Search', 'Help', 'Contact us', 'Jyväskylä University Library', and 'Login'. The main header features the 'JYKDOK' logo and the text 'JYVÄSKYLÄN YLIOPISTON KIRIASTO'. Below the header, a search bar contains the query 'scientific publishing'. To the right of the search bar, there are filters for 'JYK's materials' and an 'Advanced search' option. The search results are displayed in a list format. The first result is a book titled 'Productivity in academia : scientific publishing at Norwegian universities' by Kyvik, Svein, published in 1991. The second result is an eBook titled 'Stealing into print : fraud, plagiarism, and misconduct in scientific publishing' by LaFollette, Marcel C., published in 1992. On the right side of the search results, there is a 'Narrow search' panel with various filters such as 'Available on line', 'Organisation', 'Content type', 'Topic', 'Author', 'Language', and 'Year'. The 'Content type' filter is expanded, showing options like 'Book material (159)', 'Journal/Article (71)', 'Other (7)', and 'Database (1)'. The search results are sorted by 'Relevance' and show 'Search results 1 - 20 / 241'.

Repository content in Discovery?

- Finland currently has about a dozen repositories, with 50+ organizations and 200.000+ OA full-text publications
- VuFind/Finna can harvest metadata from repositories
- However, it is not always clear how to deal with the repository content in discovery context

The screenshot displays the TamPub Institutional repository interface. At the top, it identifies the University of Tampere and the repository name. A search bar is present with options for 'This Collection' and 'Search TamPub'. Below the search bar, there are navigation links for 'Articles' and 'Browsing by Title'. A list of articles is shown, each with a PDF icon, title, author(s), year, and type. The first article is 'A Body Shape Index' by Cheung, Yin Bun (2014). The second is 'A brief intervention is sufficient for low threshold adolescent' by Laukkanen, Eila, Hintikka, J. Mauri (2010). The third is 'A defined and xeno-free cell clinical-grade human embryonic cells' by Rajala, Kristiina, Lindroos, Riikka S., Pekkanen-Mattila, Susanna, Narkilahti, Susan, Otonkoski, Timo, Suuronen (2010). To the right, there is a 'This Collection' sidebar with various filters. Below the article list, there is a section for 'MTT's publication series' with a search bar and a list of collections. A 'Recent Submissions' section follows, listing three articles with their titles, authors, dates, and publication series.

Discovery: Conceptual issues

- Finna has two separate indices – a local index and a central index
 - The local index contains harvested metadata for the content which may or may not have been previously catalogued into the library OPAC
 - The central index contains metadata for the licensed content acquired by the library
- Should the worldwide repository content – and other non-local OA content – be available either in the local index or in the central index, or both?

Repositories in a local index

- What should there actually be in the local index?
 - There seems to be a tendency to restrict the local index strictly to the local content
 - For example, only theses produced within the organization, not from elsewhere
 - There are fears that the local content will get lost in the search results, if there is too much outside stuff available
- So, we may end up having the content of the local repository harvested, but not the repository content from elsewhere

Repositories in a central index?

- There are several big commercial vendors providing a central index for licensed content
 - Can we trust them to include open access repository content from all around the world in their product ?
 - Currently the short answer is no: to a large extent, the vendors seem to harvest repository content only from their own customer organizations

We are not making it easy, either

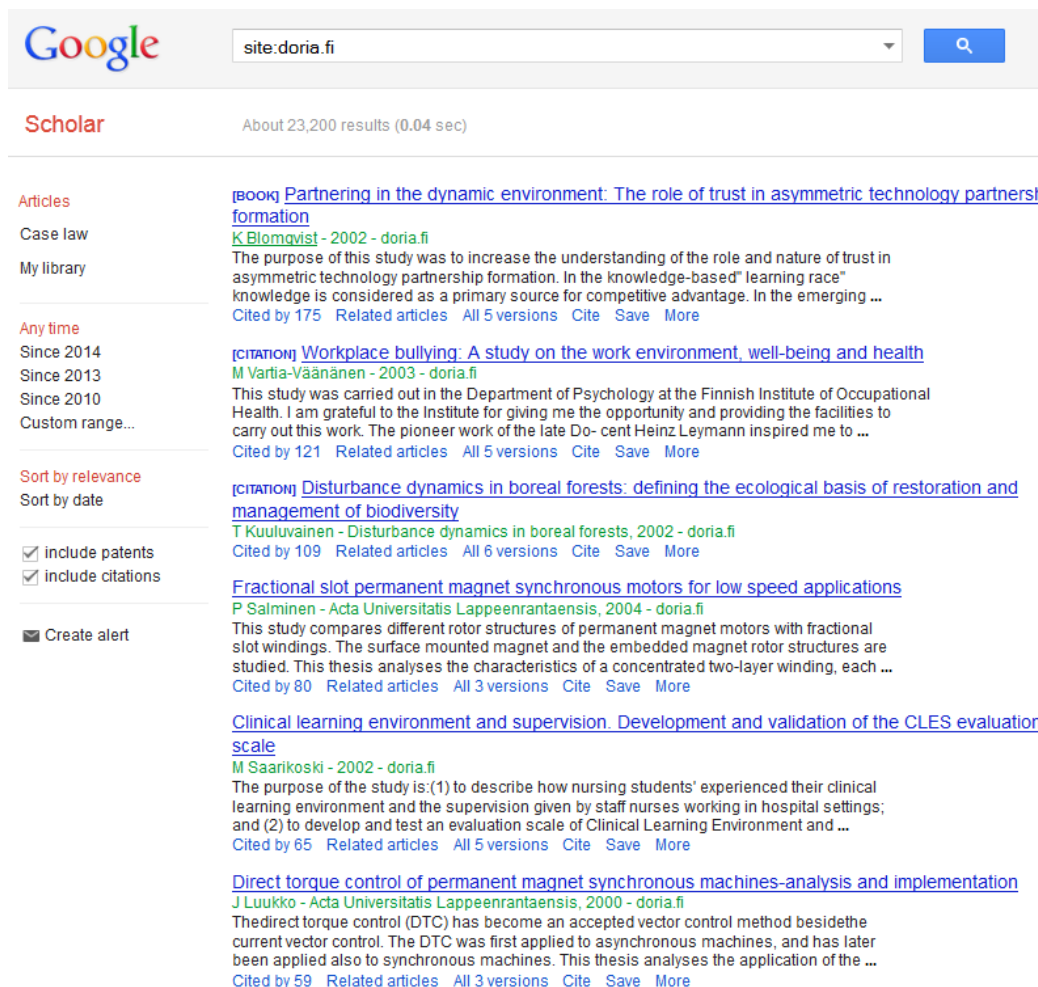
- The repositories contain very diverse stuff – should all of this be harvested?
- The repository metadata is far from uniform and has quality issues
- These problems have been well-known for a long time... but we haven't been able to solve them
 - OAIster wasn't much of a success
 - BASE (<http://www.base-search.net/about/en/>) is much better, but also provides excellent illustration for many of the problematic issues
 - Discovery systems like Finna must deal with the same issues, if we hope them to include the repository metadata as well

Challenges: Metadata formats

- Dublin Core
 - No clearly defined or easy-to-understand guidance on how to deal with metadata that goes beyond the basic 15 fields
 - Leads to different local interpretations, ad hoc solutions
 - The development of DC itself seems to have just about stalled

Challenges: Google is making us lazy?

- Despite these problems, we have been doing quite well, thanks to Google
- Most of us are getting most of our traffic via Google ... which doesn't really rely on the quality of our metadata



The screenshot shows a Google Scholar search interface. The search bar contains 'site:doria.fi' and the search button is visible. Below the search bar, it indicates 'About 23,200 results (0.04 sec)'. The left sidebar contains filters for 'Articles', 'Case law', 'My library', 'Any time' (with options for 'Since 2014', 'Since 2013', 'Since 2010', and 'Custom range...'), 'Sort by relevance' (with 'Sort by date' selected), and checkboxes for 'include patents', 'include citations', and 'Create alert'. The main content area displays a list of search results, each with a title, author, and a brief abstract. The first result is 'Partnering in the dynamic environment: The role of trust in asymmetric technology partnership formation' by K Blomqvist (2002). The second is 'Workplace bullying: A study on the work environment, well-being and health' by M Vartiainen (2003). The third is 'Disturbance dynamics in boreal forests: defining the ecological basis of restoration and management of biodiversity' by T Kuuluvainen (2002). The fourth is 'Fractional slot permanent magnet synchronous motors for low speed applications' by P Salminen (2004). The fifth is 'Clinical learning environment and supervision. Development and validation of the CLES evaluation scale' by M Saarikoski (2002). The sixth is 'Direct torque control of permanent magnet synchronous machines-analysis and implementation' by J Luukko (2000).

Challenges: Fragmented community

- Unfortunately, the repository community is fairly diverse and fragmented
- There are valiant efforts to solve metadata issues
 - Some may be platform-based
 - Some may aim for coherence on a geographical area
- But it's hard to come up with a global solution that would be adopted by everybody

Should we care? What should we do?

- We may continue to rely on Google even in the future...
 - But what if it really turns evil at some point?
- We should work with the discovery people to make sure that our content is included in a sensible way
- Even if it's not going to be easy, we should do support the efforts for metadata co-ordination

How to fix it for Finna?

- Of course, we just cannot wait for the issues to be solved on a global level (here we go again...)
- We must figure out how much normalization can be done after the metadata is already harvested from local systems
- And we must also figure out what can be done to improve the coherence of metadata in local repositories, at least on a national level

Thanks!

jyrki.ilva@helsinki.fi

