

Presentation :: The Fascinator: A lightweight, modular contribution to the Fedora-commons world

[Peter Sefton](#)

University of Southern Queensland
sefton@usq.edu.au

Oliver Lucido

University of Southern Queensland
lucido@usq.edu.au

2009-05-20

Abstract

The ARROW project sponsored a hybrid commercial/open-source approach to building vendor-supported repository infrastructure with open-source underpinnings. The Fascinator was conceived initially as a way to prove a point in an ongoing dialogue within the ARROW project about repository architecture. The goal was to test the hypothesis that it would be possible to build a useful, fast, flexible web front end for a repository using a single fast indexing system to handle browsing via facets, full-text search, multiple 'portal' views of subsets of a large corpus, and most importantly, easy-to-administer security that could handle the most common uses cases seen in the ARROW community. This contrasted with the approach taken by ARROW's commercial partner, which used several different indices to achieve only some of the same functionality in an environment which was much more complex to manage and configure. The Fascinator has also been used experimentally to index research data on desktop computers, with a view to allowing academics to classify their data and have it routed to downstream repositories via protocols such as ATOM and OAI-PMH.

Background

The Australian government has supported the development of repository infrastructure for several years now. One product of this support was the [ARROW project](#) (Australian Research Repositories Online to the World). The ARROW project sponsored a hybrid commercial/open-source approach to building vendor-supported repository infrastructure with open-source underpinnings.

Acknowledgements & Credits

The Fascinator was built with [ARROW](#) money on [Apache Solr](#) and [Fedora Commons](#). (Using code from the [FRED](#) project.

At USQ, there was a team involved in this work: Oliver Lucido, Ron Ward, Linda Octalina, Bronwyn Chandler Caroline Drury and Duncan Dickinson all assisted in programming and/or project management.

Alison Dellit at the National Library of Australia and Neil Dickson from Monash were on the project team as stakeholder/drivers.

Summary

- Show multiple web views (portals) from a single repository in two modes: Server/Desktop.
- Demonstrate the benefits of a [Jython](#)-scriptable indexer for Fedora.
- Show workable access control using limit queries to constrain what user-groups can see.
- Used in: Humanities repository @ [USQ](#). [Australian IR census](#).

The Fascinator was conceived initially as a way to prove a point in an ongoing dialogue within the ARROW project about repository architecture. The goal was to test the hypothesis that it would be possible to build a useful, fast, flexible web front end for a repository using a single fast indexing system to handle browsing via facets, full-text search, multiple 'portal' views of subsets of a large corpus, and most importantly, easy-to-administer security that could handle the most common use cases seen in the ARROW community. This contrasted with the approach taken by ARROW's commercial partner, which used several different indices to achieve only some of the same functionality in an environment which was much more complex to manage and configure.

We will give an overview of the product in both functional and technical terms, in the two modes that the product is used in: The Fascinator Server and The Fascinator Desktop.

What does it do?

Functionally, The Fascinator offers:

- Click-to-create portals.
- Easy to configure security based on a filter system, the repository owner can express security in terms of saved-searches that define what a user or group is allowed to see.
- Highly flexible indexing of a Fedora repository for administrators (and by extension anything the harvesting module can scrape-up).
- Runs on a server or (experimentally) the desktop.

- [journal article](#) (626)
- [thesis](#) (119)
- [conference paper](#) (117)
- [report](#) (65)
- [book chapter](#) (47)
- [more](#)

However, this view is only available because the normalising rules in the NLA harvester, which pulls data from repositories and then normalises it for inclusion in the repository. A similar search on an incomplete dataset without normalisation looks like this:

Type

- [Journal Article](#) (184)
- [PeerReviewed](#) (105)
- [Article](#) (75)
- [Thesis](#) (66)
- [Book chapter](#) (65)
- [NonPeerReviewed](#) (62)
- [Conference Paper](#) (35)
- [Journal Articles \(Refereed Article\)](#) (30)
- [c1](#) (28)
- [techreport](#) (27)
- [Full-text link or file](#) (26)
- [Conference or Workshop Item \(DEST Category E\)](#) (21)
- [PhD Doctorate](#) (20)
- [Article \(DEST Category C\)](#) (19)
- [journal article](#) (18)
- [Book Chapter](#) (17)
- [text](#) (14)
- [Book Section](#) (10)
- [Report](#) (9)
- [Conference Publications \(Full Written Paper - Refereed\)](#) (8)
- [Conference or Workshop Item](#) (8)
- [e1](#) (8)
- [Book Chapters](#) (7)
- [b1](#) (7)
- [Book Chapter \(DEST Category B\)](#) (5)

Note the inconsistency of what repositories publish in their OAI-PMH feeds – there are things labeled as *Journal Article*, *Article*, *journal article* which may or may not be the same as the NLA's *journal article* type, And what is a 'b1'? Only those familiar with Australian Government repository requirements for university research are going to understand that one.

<http://cairss.caul.edu.au/www/aust-repos-census/aust-repos-census.htm>

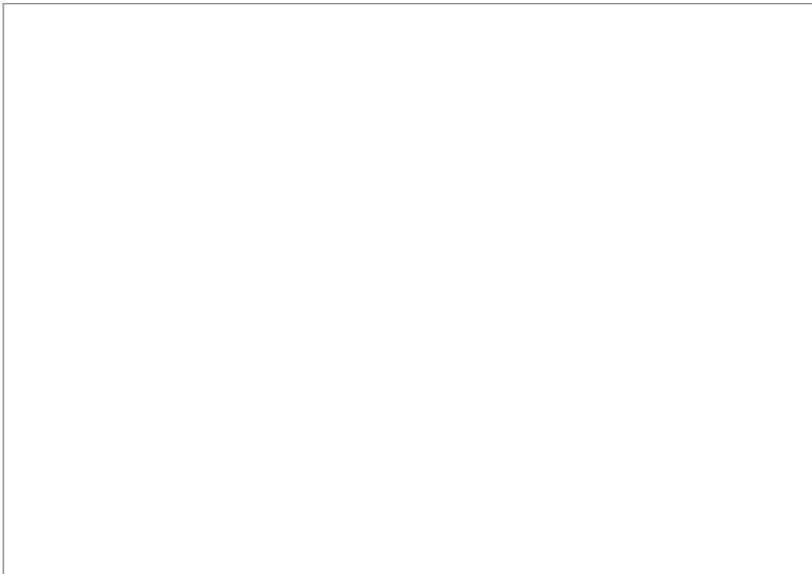
The Fascinator Desktop

The Fascinator Desktop is a new software project which was inspired by a perceived gap in the available free eResearch software for small-scale research. How does one get 'ordinary' files from a hard disk into a data grid, or even backed up?

The concept was introduced in a blog post:

We'll start with a local installation of The Fascinator – that puts Fedora 3 and Apache Solr on your desktop. Don't worry, we have a simple installer. It's all Java, so it might be painful for the programmers at times but it should install pretty much anywhere.

Then we will add a file-system indexer for The Fascinator – pretty much like what Picasa does, it will index *all* of your stuff. It will grab whatever metadata it can, including properties from office documents, EXIF metadata and tags from images . We will also treat the file system as a source of metadata so you will be able to explore using metadata facets and file system facets using the same interface. This should be a very straightforward addition to the existing software, it's just a matter of bolting together some standard software libraries.



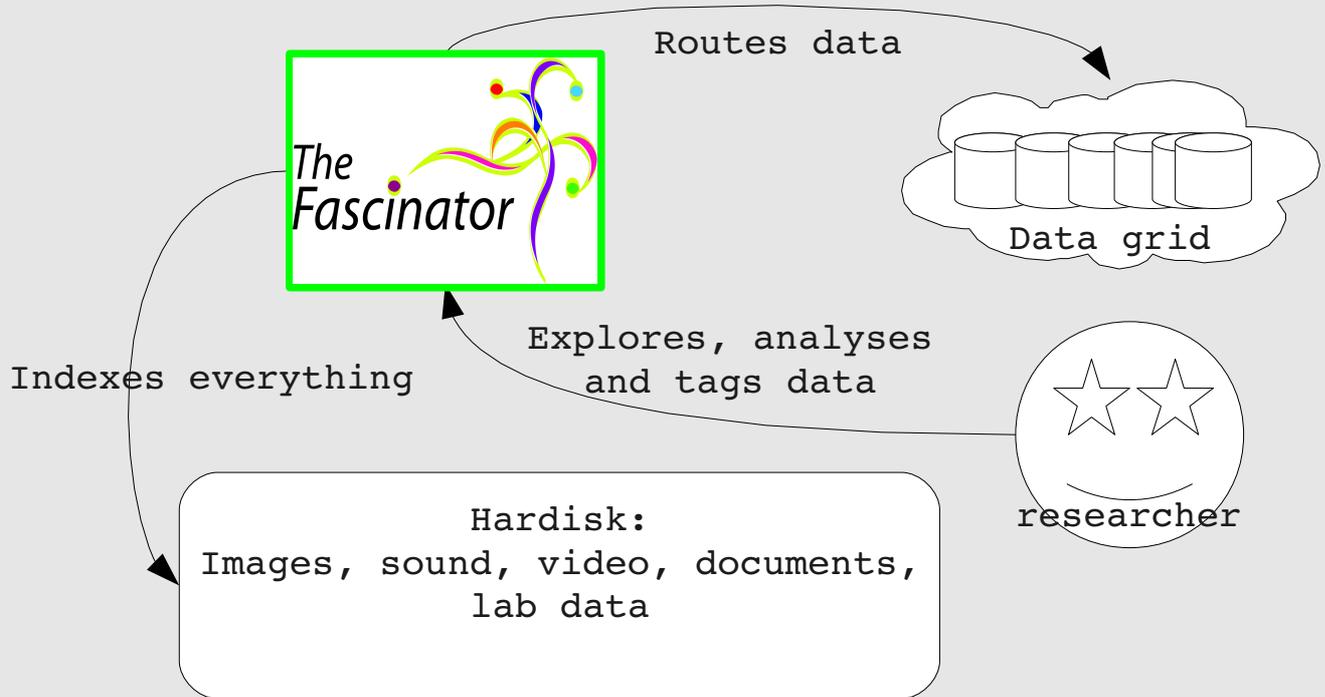
Next comes the taxonomy/tagging bit: we need a way to import tag-sets and taxonomies that you might want to apply to your content and then let you tag it. I think it will be important to support both formal metadata and informal tagging. For example, you might want to set up your own tag hierarchy with home/work at the root, and with work broken down into teaching/research and research broken up by project.

Desktop sucker-upper

The Fascinator desktop finds all your stuff:

- Treats directories/folders as metadata in their own right.
- Folksy tagging and formal ontology/taxonomy support.
- Route data to departmental, institutional or global repositories.
- Configurable to deal with ANY data type – automatic conversion.
- Configurable to deal with the way you work.

What does it do?



The Fascinator Desktop was inspired by a number of other software packages.

Inspirations / alternatives

- Similar ideas: [Lensfield](#), [MediaFlux](#).
- Inspiration: [Picasa](#), [DigiKam](#) and [FieldHelper](#) for data management, tagging and upload, MP3 players for browsing, [Yahoo Pipes](#) for pipiness, [VuFind](#) for faceting.

But going beyond these, the software is designed to allow researchers to 'slice and dice' their own data files, and to work with eResearch analysts on customised indexers which can deal with data and relationships between data. For example, for a researcher using a lot of video the indexer might be configured to make web-deliverable and preservation quality versions of video and route then to a repository and backup respectively. If there are time coded transcripts of video interviews, the the indexer will be able to provide full-text search into the video. These indexing rules will take time an effort to develop for each discipline and workgroup, but over time should be able to be re-used.

On a simpler level, researchers will be able to tag items, then create 'portal' views based on tags and move data via Atom to other services.

Summary

The Fascinator is not yet widely deployed in either of its manifestations, but it is freely available to download under a GPL license and at the Australian Digital Futures institute we are working on building a sustainable approach to maintaining the software.

Getting involved

- See the [demo site](#) or our [repository census](#).
- Download from [The Fascinator](#) website.
- Download the [demonstration virtual machine](#) with The Fascinator (plus some other stuff)
- Join the [Google Group](#).