

Open Access Open Library

Thomas Krichel^{1,2,3}

¹Long Island University ²Novosibirsk State University ³Open
Library Society

December 1, 2010

my life

- From house out am I economist.
- I published the first bibliographic dataset in economics on Internet in 1993.
- I published the first online paper in economics in 1993.
- This became the RePEc digital library. It's the work I'm most known for.

my times

- We live between two worlds.
 - In the old, information was scarce and attention plentiful.
 - In the new, information is plentiful and attention is scarce.
- This nowhere else as important as in academic information.
- I understood this in the early 90s. I was not alone.

my Harnad story

- Harnad apologies to me in a public meeting with our funders in 1997, for him having advocated a centralized repository infrastructure á la arXiv.org.
- RePEc was already a repository system before OAI come along.

RePEc today in numbers

- 375000 working papers
- 575000 journal articles
- 2000 software components
- 21000 book and chapter listings
- 26000 author contact and publication listings
- 11900 institutional contact listings

RePEc today in brief

- There are close 1300 archive that furnish data to RePEc.
- They give metadata about documents. They may contain full-text links.
- Some archives serve other special purposes.

RePEc and Harnad

- RePEc is living proof that a decentralized system levies large contents *without* mandates.
- It's the only one ;-(that I am aware.

Learning from RePEc is learning how to win!

- OK, works only in economics!
- Economics had a working paper culture before.
- But compare to computer science. Without a RePEc equivalent, the working paper culture there is gone.

first thing we can we learn?

- First: we need a library. But it's not your grandpa's library.
- We build a comprehensive description of what is published. That's traditional library fare.
- But we also also do non-traditional work: we measure performance of academic actors. This presumed to be the main driver behind the growth of RePEc.

second thing we can learn

- Second: it's the services stupid.
- But it's not your grandpa's library.

libraries and services in RePEc

- What we need to learn from RePEc is that libraries and services are in sequence.
- RePEc we have feedbacks from the library to services and back.
- I will give some examples.

example 1: “NEP: New Economics Papers”

- service
 - This is service by human editors who filter all new additions to the working paper stock.
 - I set it up in 1998.
- library
 - NEP groups papers by themes, like subject headings.

example 2: “CitEc”

- library
 - It is a set of citations data that is extracted citations from freely downloadable papers in RePEc
 - José Manuel Barrueco Cruz set it up in 2000.
- service
 - The data appears in all RePEc services that have a citation component.

example 3: “RePEc Author Service”

- service
 - This service allows authors to create a profile of their works.
 - I (with Markus J.R. Klink) set it up in 1999.
- library
 - This allows for access control.

example 4: “LogEc”

- library
 - This service records usage of RePEc papers across user service.
 - Sune Karlsson set it up in 1999(?).
- service
 - With RAS data, allows us to tell scholars how popular their works are.

the crowning achievement

- We have a bunch of ranking of both individuals and their institutions.
- They are keen to improve their position in the ranking.

conclusion

- Learning from RePEc is learning how to win!
- We need a super-RePEc for all disciplines.
- Like RePEc it must be an open library.
 - something that everybody can use,
 - something that everybody can contribute to.

probing goes over studying

- As the motto of Saarland goes, I have already started with it.
- but I am not completely on my own...

Open Library Society, Inc.

- It's a small 503(c) not-for-profit organization.
- It mainly provides an umbrella for my work.
- It runs four projects, that I will discuss here.
- It now also is a legal representation for RePEc.

3lib

- It's an attempt to group paper data from various sources.
- To have a large capture area, we need to keep the common elements simple.
- We still need to avoid junk data.

common elements

- handle
- title
- author names
- link to abstract HTML page

current contents

- PubMed
- DBLP
- RePEc
- CiteSeer (problematic)
- Bielefeld/Driver OAI data
- others..., about 30 million documents

purposes

- author claiming (now)
- search/detection of full text (later)

author claiming: AuthorClaim

- Author create a profile.
- The profile contains name variations.
- The system searches for the papers containing the name variations as author names.
- more on this tomorrow...

institutional data

- When authors register, they can also claim affiliation with an institution.
- This data comes from Academic and Research Institutions in the World.
(ARIW)

ARIW

- ARIW confuses a dataset and a service.
- The dataset contains close to 30,000 minimal institutional records.
- It does not have much maintenance these days ;-(

minimal data contents

- handle
- name expressions (possibly in different languages)
- url/domain

ARIW service

- It is a model for an open service.
- All data *and code* can be downloaded.
- You can access the code that is “life” on the machine.
- On <http://ariw.org/internal> we enable a design like the Centre Pompidou.

conclusions 1

- We are not alone. There is a growing movement for open, re-usable bibliographic data.
- Repositories are naturally aligned with this movement.
 - It is important to integrate repository items to the wider scholarly environment.
 - It is even more important to do this in a way that allows for performance evaluation.

conclusions 2

- Is this completely crazily optimistic?
Yes, but so did a structure like RePEc appear
- It will never take off?!? Well, repositories have not taken off much either. Are they the right thing to do? Yes.
- We will fight on, together!