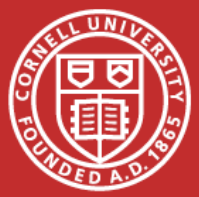


Cornell University
Library

Author identifiers:
1) Services at arXiv, and
2) ORCID and repositories

Simeon Warner

DLIT, Cornell University Library
(simeon.warner@cornell.edu)

Open Repositories 2010, Madrid, Spain, 7 July 2010

1) Author id services at arXiv

Update from last year

This work in collaboration with Thorsten Schwander, Nathaniel A. Woody, and Paul Ginsparg

Funded by a Microsoft TCI grant

Recap

- Since 2009-03 arXiv has offered the option to create a “public author id” linked to arXiv authorship account information (still opt-in)
- Ids look like
  - http://arxiv.org/a/warner_s_1
  - HTTP conneg to HTML, Atom with 303
- Encourage users to curate authorship data by providing services to motivate participation
- Experiment with connections to social networking tools

Salvatore Mele's articles on arXiv

[1] **arXiv:0906.5418** [pdf]

Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories

Anne Gentil-Beccot, Salvatore Mele, Travis Brooks

Comments: Version to be published in Scientometrics

Subjects: **Digital Libraries (cs.DL)**

[2] **arXiv:0907.2658** [pdf]

Study of hadronic event shape in flavour tagged events in e+e-annihilation at s = 197 GeV

L3 Collaboration

Comments: 29 pages, 9 figures

Journal-ref: PMC Phys.A2:6,2008

Subjects: **High Energy Physics - Experiment (hep-ex)**

[3] **arXiv:0906.0485** [pdf]

First results from the PARSE.Insight project: HEP survey on data preservation, re-use and (open) access

Andre Holzner, Peter Igo-Kemenes, Salvatore Mele

Comments: Contribution to the First Workshop on Data Preservation and Long-Term Analysis in High-Energy Physics, DESY, Hamburg, Germany, January 26th-28th 2009

Subjects: **Digital Libraries (cs.DL)**; High Energy Physics - Experiment (hep-ex); Data Analysis, Statistics and Probability (physics.data-an)

[4] **arXiv:0805.2739** [pdf]

Innovation in Scholarly Communication: Vision and Projects from High-Energy Physics

Rolf-Dieter Heuer, Annette Holtkamp, Salvatore Mele

Comments: Based on a keynote talk of Rolf-Dieter Heuer at APE2008, International Conference on "Academic Publishing in Europe", Berlin, January 22nd-23rd 2008

Journal-ref: Inform.Serv.Use 28:83-96,2008

Subjects: **Digital Libraries (cs.DL)**

[5] **arXiv:cs/0611130** [pdf, ps, other]

Quantitative Analysis of the Publishing Landscape in High-Energy Physics

Salvatore Mele, David Dallman, Jens Vigen, Joanne Yeomans

Salvatore Mele's articles on arXiv

Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories
11/25/2009 04:58 AM

Contemporary scholarly discourse follows many alternative routes in addition to the three-century old tradition of publication in peer-reviewed journals. The field of High- Energy Physics (HEP) has explored alternative communication strategies for decades, initially via the mass mailing of paper copies of preliminary manuscripts, then via the inception of the first online repositories and digital libraries. This field is uniquely placed to answer recurrent questions raised by the current trends in scholarly communication: is there an advantage for scientists to make their work available through repositories, often in preliminary form? Is there an advantage to publishing in Open Access journals? Do scientists still read journals or do they use digital repositories? The analysis of citation data demonstrates that free and immediate online dissemination of preprints creates an immense citation advantage in HEP, whereas publication in Open Access journals presents no discernible advantage. In addition, the analysis of clickstreams in the leading digital library of the field shows that HEP scientists seldom read journals, preferring preprints instead.

Study of hadronic event shape in flavour tagged events in e+e-annihilation at s = 197 GeV
07/15/2009 01:10 PM

Results are presented from a study of the structure of hadronic events in high-energy e+e- interactions detected by the L3 detector at LEP. Various event shape distributions and their moments are measured at several energy points at and above the Z-boson mass. The event flavour is tagged by using the decay characteristics of b-hadrons. Measurements of distributions of event shape variables for all hadronic events, for light (u, d, s, c) and heavy (b) quark flavours are compared to several QCD models with improved leading log approximation: JETSET, HERWIG and ARIADNE. A good description of the data is provided by the models.

First results from the PARSE.Insight project: HEP survey on data preservation, re-use and (open) access
06/02/2009 09:55 AM

There is growing interest in the issues of preservation and re-use of the records of science, in the "digital era". The aim of the PARSE.Insight project, partly financed by the European Commission under the Seventh Framework

Using the `myarticles` widget

We have created a JavaScript widget that allows you to easily display your articles on your website, using the magic of JSON and Atom. To the right of this paragraph you should see an example of what this provides for arXiv user Simeon Warner. If you would like to use this functionality, you will need to create an arXiv public author identifier. The quickstart section below will get you up and running rapidly, complete details follow the quickstart.

Quickstart

First, you will need to figure out where on the webpage you would like the arXiv atom feed information to be placed and add a `<div id="arxivfeed"></div>` element. The arXiv data will be inserted into this element (you may include HTML inside this element which will be replaced when the data loads, for example this page has `[Loading myarticles...]`). Next you will need to include the arXiv JavaScript that will access arXiv, retrieve the list of your articles, and then format and display the data on your page. Simply insert this:

```
<script type="text/javascript">
<!--
var arxiv_authorid = "yourauthorid";
//--></script>
<script type="text/javascript" src="http://arxiv.org/js/myarticles.js"></script>
```

either in between the `<head>` and `</head>` tags or soon after the `<body>` tag in your web page. The first script entry provides configuration variables that controls what information is rendered on your web page. The only required

Simeon Warner's articles on arXiv

Author Identifiers in Scholarly Repositories
Simeon Warner

A Web-Based Resource Model for eScience: Object Reuse & Exchange
Carl Lagoze, Herbert Van de Sompel, Michael Nelson, Simeon Warner, Robert Sanderson, Pete Johnston

Object Re-Use & Exchange: A Resource-Centric Approach
Carl Lagoze, Herbert Van de Sompel, Michael L. Nelson, Simeon Warner, Robert Sanderson, Pete Johnston

Plagiarism Detection in arXiv
Daria Sorokina, Johannes Gehrke, Simeon Warner, Paul Ginsparg

Pathways: Augmenting interoperability across scholarly repositories
Simeon Warner, Jeroen Bekaert, Carl Lagoze, Xiaoming Liu, Sandy Payette, Herbert Van de Sompel

E-prints and Journal Articles in Astronomy: a Productive Co-existence
Edwin A. Henneken, Michael J. Kurtz, Simeon Warner, Paul Ginsparg, Guenther Eichhorn, Alberto Accomazzi, Carolyn S. Grant, Donna Thompson, Elizabeth Bohlen, Stephen S. Murray

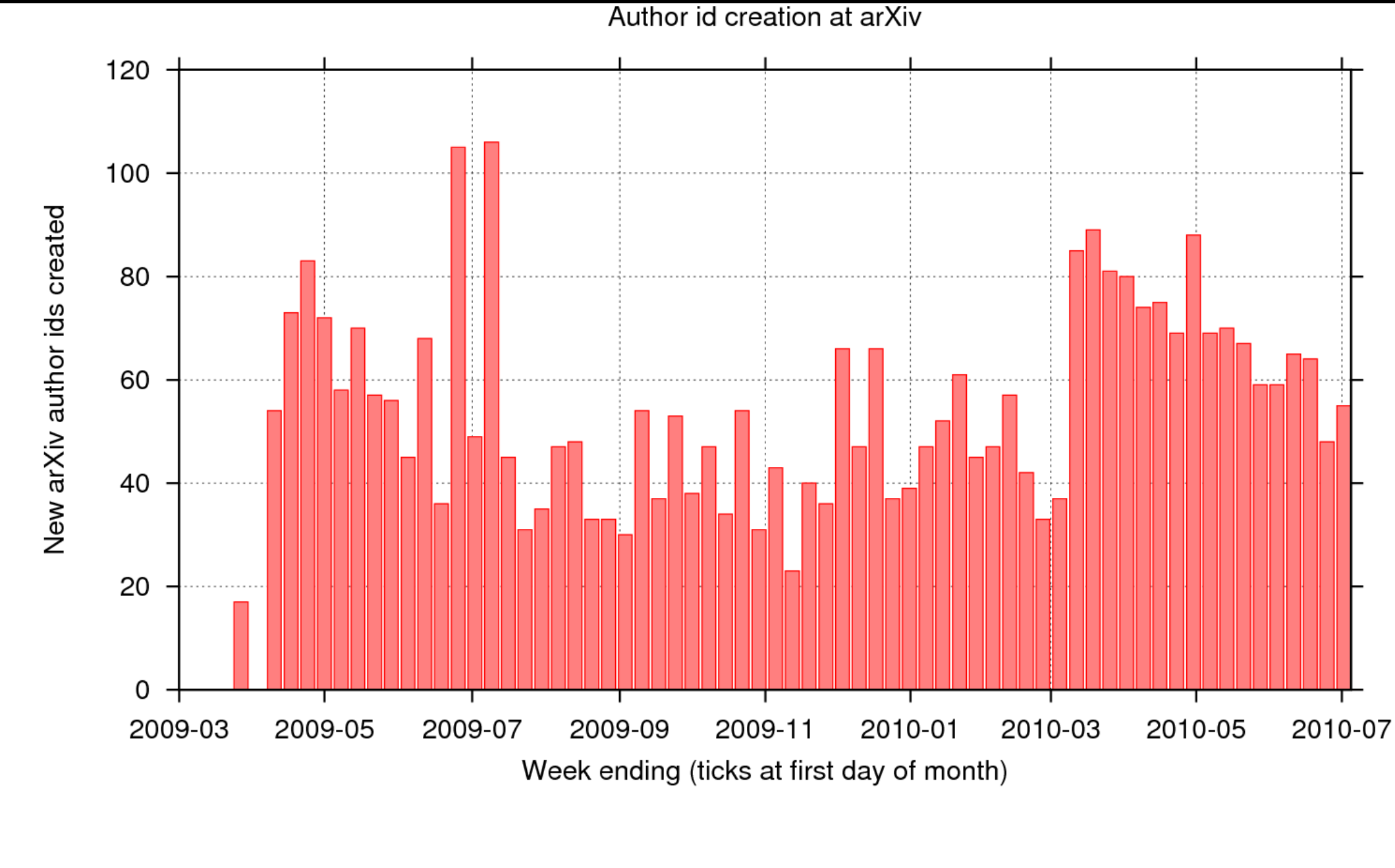
The OAI Data-Provider Registration and Validation Service
Simeon Warner

Eprints and the Open Archives Initiative
Simeon Warner

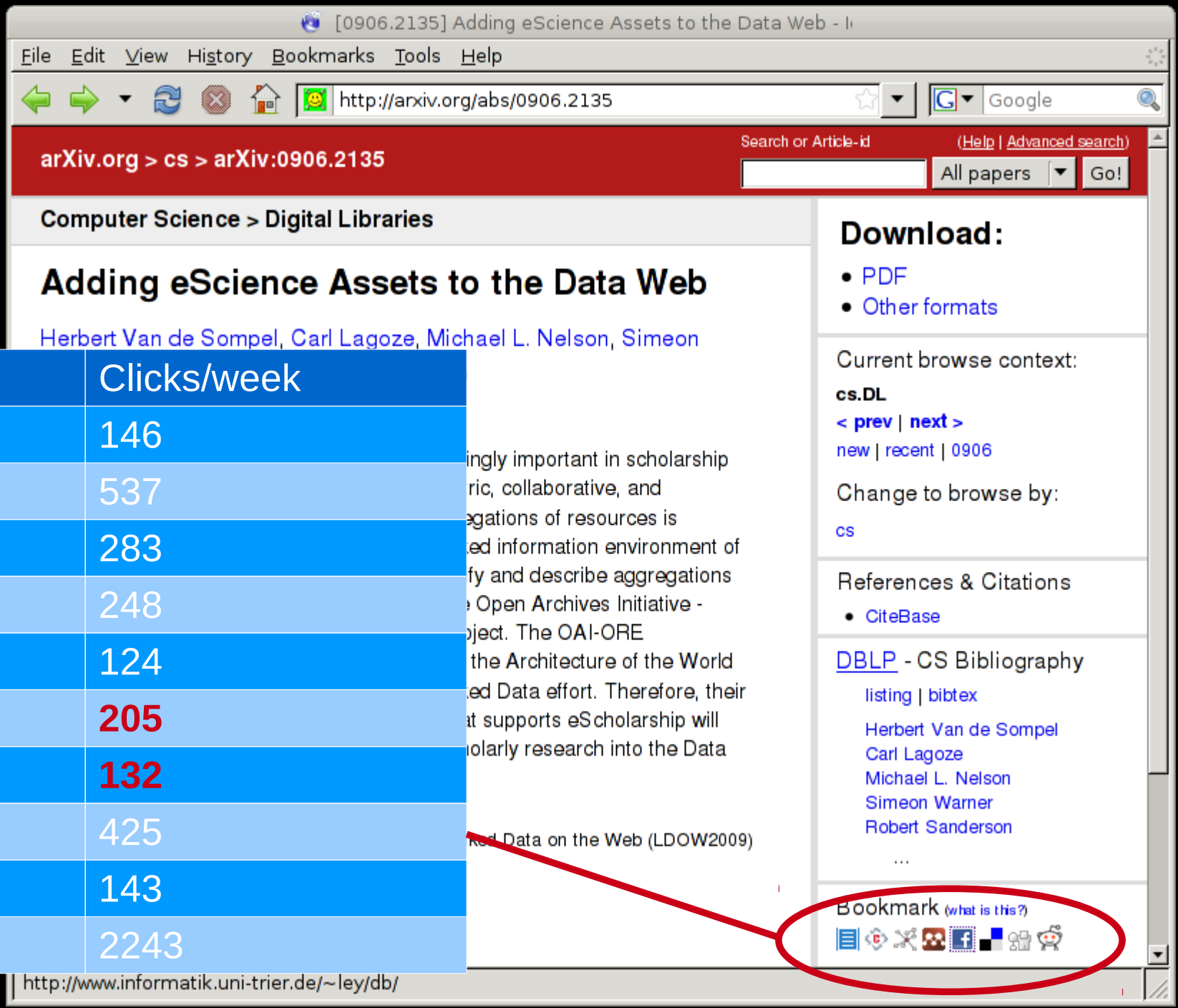
Exposing and harvesting metadata using the OAI

Use of author id services at arXiv

Venue	Service	Launched	Usage
arXiv	Author id creation	2009-03	>3600 created, ~60 new/week
Web/HTML	Author page	2009-03	3000 accesses to 600 ids/week
	widget	2009-03	2500 accesses to 120 ids/week
Facebook	MyArXiv	2009-03	2200 users, 780 “active”
	My papers		360 users, 20 new/month
	Share or alert	2009-08	275 users, 25 new/month; ~800 click-throughs/month
	Tagging authors	2009-10	Very little
LinkedIn	MyArXiv	2010-03	~500 click-throughs/month


Author id creation at arXiv
New arXiv author ids created
0
20
40
60
80
100
120
2009-03
2009-05
2009-07
2009-09
2009-11
2010-01
2010-03
2010-05
2010-07
Week ending (ticks at first day of month)

Site	Clicks/week
bibsonomy	146
citeulike	537
connotea	283
del.icio.us	248
digg	124
facebook	**205**
linkedin	**132**
mendeley	425
reddit	143
total	2243



Facebook – tidied UI and interaction, added thumbnails and tagging. Did not see significant increase in use. Steady use by modest number of users

MyArXiv on Facebook - Iceweasel

File Edit View History Bookmarks Tools Help

http://apps.facebook.com/myarxiv/?arxivpost=http%3A

Google

facebook Home Profile Friends Inbox 19 Simeon Warner Settings Logout

Welcome, Simeon!

Facebook arXiv Application

The arXiv
your frie
when ne
comment
basic ins

This appl
arXiv aut
/set_auth

When yo
as record
ownersh
/help/aut
comment

warner

arXiv Information

access to

Publish this story to your Facebook Wall and your friends' home pages?

What do you think of this paper?

This paper is awesome!

Adding eScience Assets to the Data Web

Herbert Van de Sompel,Carl Lagoze,Michael L. Nelson,Simeon Warner,Robert Sanderson,Pete Johnston

Not Published Yet via MyArXiv

Publish Skip

POST COMMENT

My Papers | Any arXiv paper | Profile List Papers

A Web-Based Resource Model for eScience: Object Reuse & Exchange

Carl Lagoze (Untag) , Herbert Van de Sompel (Tag), Michael L. Nelson (Untag) , Simeon Warner (Untag) , Rob Sanderson (Untag) , Pete Johnston (Tag)

Submitted:2008-11-04

Abstract: Work in the Open Archives Initiative - Object Reuse and Exchange (OAI-ORE) focuses on an important aspect of infrastructure for eScience: the specification of the data model and a suite of implementation standards to identify and describe compound objects. These are objects that aggregate

Acknowledgements

The arXiv e-print archive is a service of the Cornell University library. This facebook application is a project of Cornell Information Science and has been supported by Microsoft through a Technical

Applications Bookmark MyArXiv Brian Sakofsky Chat (Offline)

Done

Facebook | Simeon Warner - Iceweasel

File Edit View History Bookmarks Tools Help

http://www.facebook.com/simeon.warner?ref=profile

Google

facebook Home Profile Friends Inbox 19 Simeon Warner Settings Logout

View Photos of Me (80)

Edit My Profile

Write something about yourself.

Information

Networks:
Cornell Staff

Friends

Simeon Warner

Wall Info Photos MyArXiv +

What's on your mind?

Attach: Share

Options

Simeon Warner This paper is awesome!

Adding eScience Assets to the Data Web

Herbert Van de Sompel,Carl Lagoze,Michael L. Nelson,Simeon Warner,Robert Sanderson,Pete Johnston

7 seconds ago via MyArXiv · Comment · Like · View on arXiv

Alan Traine You must see this vide0 now! :) **Remove**

http://fmnmkswrnkx327.blogspot.com/

Singapore, Thailand renew commitment to boost cooperation under CSEP - Channel News Asia

Source: fmnmkswrnkx327.blogspot.com

Yesterday at 1:07pm · Comment · Like · Share

Polar Humenn

In this photo: Simeon Warner

Greg and Kyla Wedding Weekend.

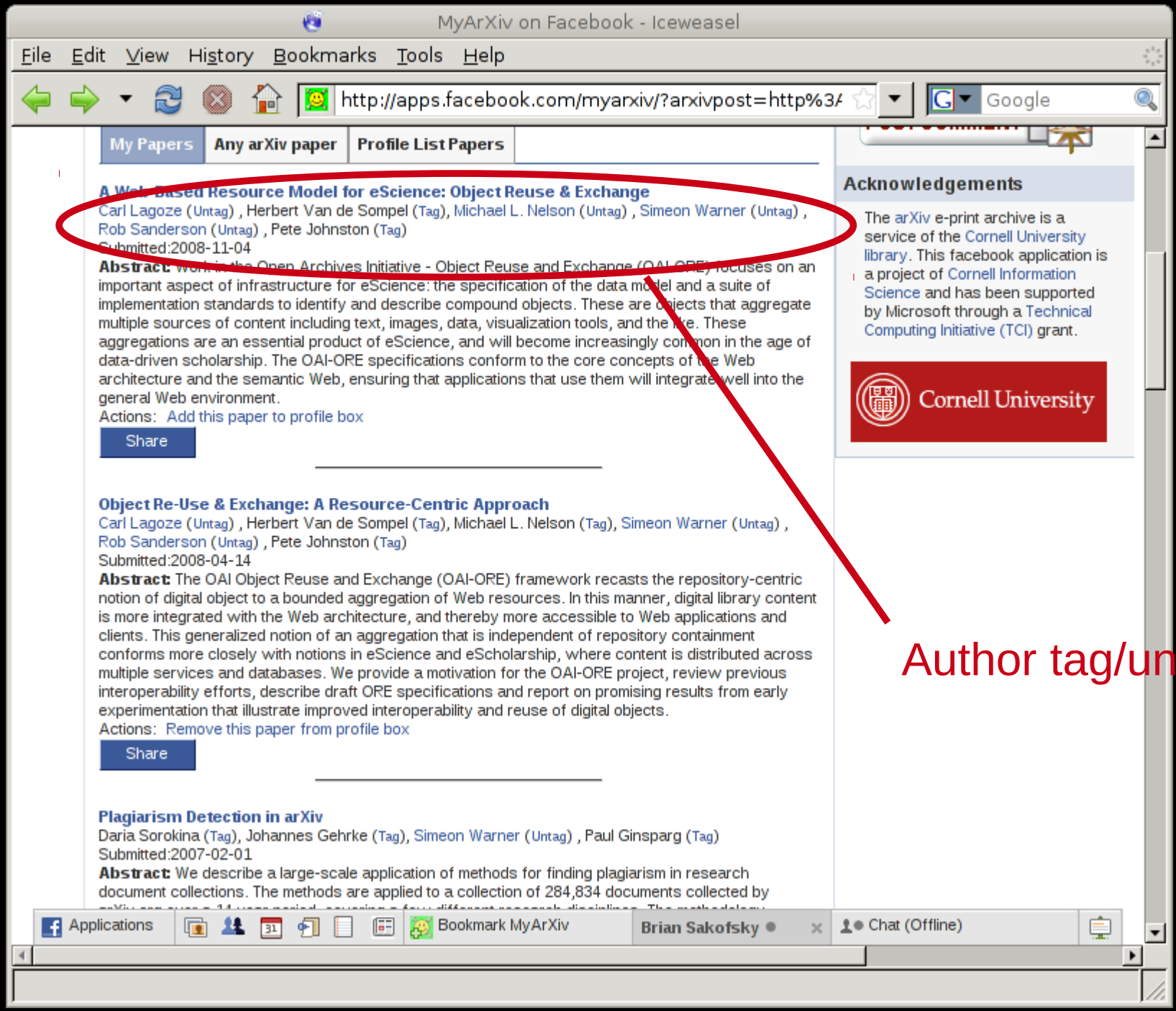
I wonder what she sees in me

Applications

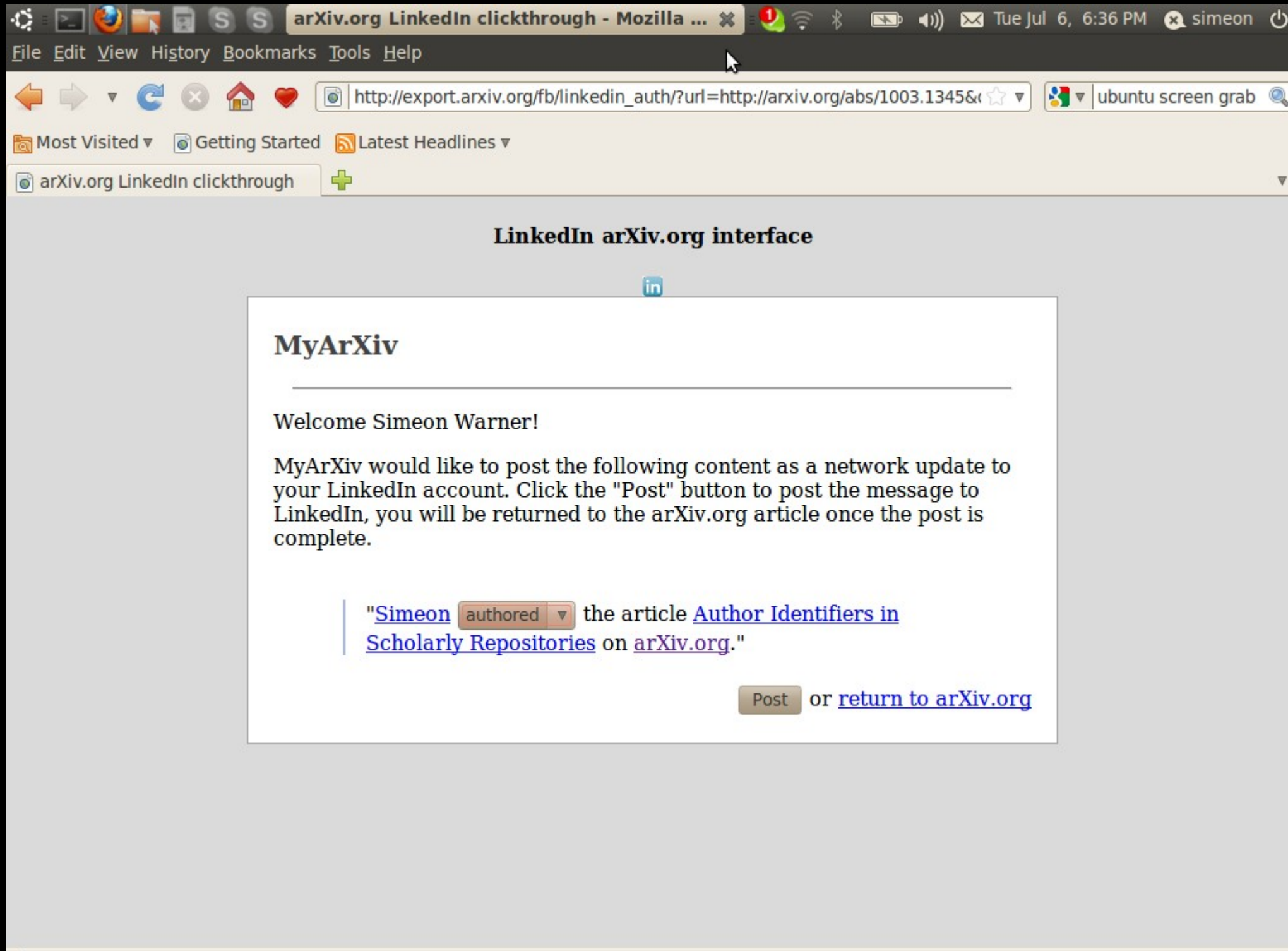
Brian Sakofsky

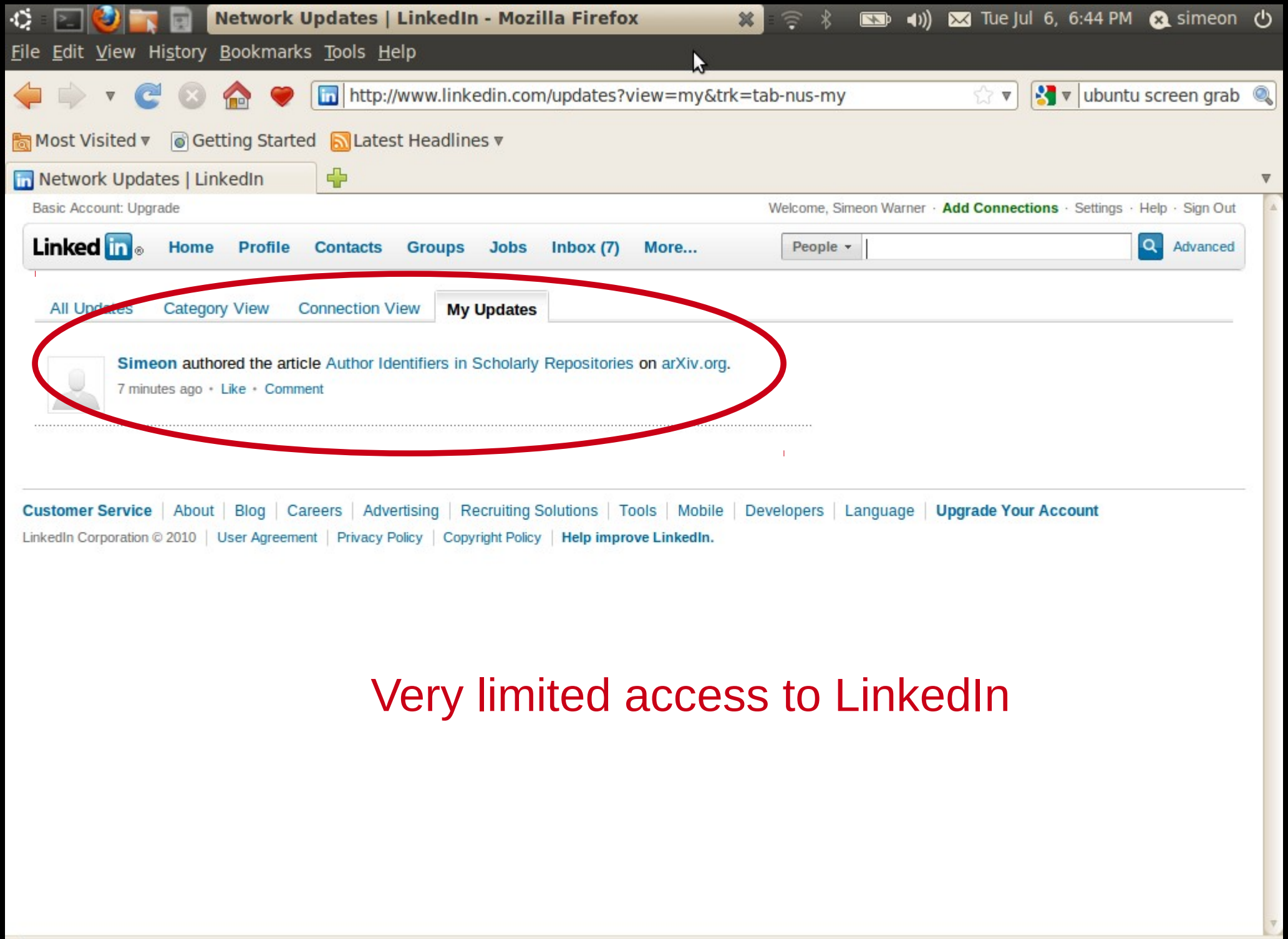
Chat (Offline)

Done


MyArXiv on Facebook - Iceweasel
File Edit View History Bookmarks Tools Help
http://apps.facebook.com/myarxiv/?arxivpost=http%3A
Google
My Papers
Any arXiv paper
Profile List Papers
A Web-Based Resource Model for eScience: Object Reuse & Exchange
Carl Lagoze (Untag) , Herbert Van de Sompel (Tag), Michael L. Nelson (Untag) , Simeon Warner (Untag) , Rob Sanderson (Untag) , Pete Johnston (Tag)
Submitted:2008-11-04
Abstract: Work in the Open Archives Initiative - Object Reuse and Exchange (OAI-ORE) focuses on an important aspect of infrastructure for eScience: the specification of the data model and a suite of implementation standards to identify and describe compound objects. These are objects that aggregate multiple sources of content including text, images, data, visualization tools, and the like. These aggregations are an essential product of eScience, and will become increasingly common in the age of data-driven scholarship. The OAI-ORE specifications conform to the core concepts of the Web architecture and the semantic Web, ensuring that applications that use them will integrate well into the general Web environment.
Actions: Add this paper to profile box
Share
Object Re-Use & Exchange: A Resource-Centric Approach
Carl Lagoze (Untag) , Herbert Van de Sompel (Tag), Michael L. Nelson (Tag), Simeon Warner (Untag) , Rob Sanderson (Untag) , Pete Johnston (Tag)
Submitted:2008-04-14
Abstract: The OAI Object Reuse and Exchange (OAI-ORE) framework recasts the repository-centric notion of digital object to a bounded aggregation of Web resources. In this manner, digital library content is more integrated with the Web architecture, and thereby more accessible to Web applications and clients. This generalized notion of an aggregation that is independent of repository containment conforms more closely with notions in eScience and eScholarship, where content is distributed across multiple services and databases. We provide a motivation for the OAI-ORE project, review previous interoperability efforts, describe draft ORE specifications and report on promising results from early experimentation that illustrate improved interoperability and reuse of digital objects.
Actions: Remove this paper from profile box
Share
Plagiarism Detection in arXiv
Daria Sorokina (Tag), Johannes Gehrke (Tag), Simeon Warner (Untag) , Paul Ginsparg (Tag)
Submitted:2007-02-01
Abstract: We describe a large-scale application of methods for finding plagiarism in research document collections. The methods are applied to a collection of 284,834 documents collected by
Acknowledgements
The arXiv e-print archive is a service of the Cornell University library. This facebook application is a project of Cornell Information Science and has been supported by Microsoft through a Technical Computing Initiative (TCI) grant.
Cornell University
Author tag/untag
Applications
Bookmark MyArXiv
Brian Sakofsky
Chat (Offline)

LinkedIn – hoped this might suit our user community better. Request for app never answered, put up minimal “post an update” when that API became open


arXiv.org LinkedIn clickthrough - Mozilla ...
Tue Jul 6, 6:36 PM
simeon
File Edit View History Bookmarks Tools Help
http://export.arxiv.org/fb/linkedin_auth/?url=http://arxiv.org/abs/1003.1345&
ubuntu screen grab
Most Visited
Getting Started
Latest Headlines
arXiv.org LinkedIn clickthrough
LinkedIn arXiv.org interface
MyArXiv
Welcome Simeon Warner!
MyArXiv would like to post the following content as a network update to your LinkedIn account. Click the "Post" button to post the message to LinkedIn, you will be returned to the arXiv.org article once the post is complete.
"Simeon authored the article Author Identifiers in Scholarly Repositories on arXiv.org."
Post
or return to arXiv.org
Done


Network Updates | LinkedIn - Mozilla Firefox
Tue Jul 6, 6:44 PM
simeon
File Edit View History Bookmarks Tools Help
http://www.linkedin.com/updates?view=my&trk=tab-nus-my
ubuntu screen grab
Most Visited
Getting Started
Latest Headlines
Network Updates | LinkedIn
Basic Account: Upgrade
Welcome, Simeon Warner · Add Connections · Settings · Help · Sign Out
Home
Profile
Contacts
Groups
Jobs
Inbox (7)
More...
People
Advanced
All Updates
Category View
Connection View
My Updates
Simeon authored the article Author Identifiers in Scholarly Repositories on arXiv.org.
7 minutes ago • Like • Comment
Customer Service | About | Blog | Careers | Advertising | Recruiting Solutions | Tools | Mobile | Developers | Language | Upgrade Your Account
LinkedIn Corporation © 2010 | User Agreement | Privacy Policy | Copyright Policy | Help improve LinkedIn.
Very limited access to LinkedIn
Done

Lessons

- Some users value service, positive email, better replacement for link to arXiv author search
- Either Facebook is not the right venue for our users ***or*** we have not hit upon the right way to connect
- Working with Facebook requires monitoring/fixing as there are “breaking” API changes about twice per year.

Plans

- Working to add OAI-ORE resource map RDF/XML in addition to Atom as machine readable format
  - Spin-off from work with Data Conservancy project
  - A step along the road to linked data for arXiv
  - (Also working on other UI pages, e.g. /abs/)
- Investigating how to share authorship information with INSPIRE project and get data back from their matching and claiming services
  - Perhaps also provide authentication site for INSPIRE
- Will keep Facebook and LinkedIn apps going, interested in new ideas...
- Working with ORCID and will add facilities to store and expose alternative identifiers in arXiv profile

2) ORCID and repositories

See http://orcid.org/ for details of ORCID and of participating organizations

See “Author Identitity: The Shape of the Problem” http://www.orcid.org/node/19, a presentation by Geoffrey Bilder (Crossref) at STM Innovations 2009 for an excellent overview of the issues. Have taken slides marked [Builder] from his talk

While I have attempted to accurately represent ORCID, the slides are mine and may have errors!

ORCID Open Researcher & Contributor ID

Working together to align the global network

An independent, community effort to standardize researcher identification

The Initiative

Name ambiguity and attribution are persistent, critical problems imbedded in the scholarly research ecosystem. The ORCID Initiative represents a community effort to establish an open, independent registry that is adopted and embraced as the industry's de facto standard. Our goal is to resolve the systemic name ambiguity, by means of assigning unique identifiers linkable to an individual's research output, to enhance the scientific discovery process and improve the efficiency of funding and collaboration. MORE

Mission Statement

ORCID aims to solve the author/contributor name ambiguity problem in scholarly communications by creating a central registry of unique identifiers for individual researchers and an open and transparent linking mechanism between ORCID and other current author ID schemes. These identifiers, and the relationships among them, can be linked to the researcher's output to enhance the scientific discovery process and to improve the efficiency of research funding and collaboration within the research community.

News

Subscribe

Announcement: New Participants ->

2010-06-25

ORCID welcomes new participant organizations:

Library of Congress; KaraM, Center for the Black Sea Studies; CABI; Royal Society of Chemistry; American Geophysical Union; Proceedings of the National Academy of Sciences (USA) (PNAS)

Announcement: New Participants ->

2010-05-28

ORCID welcomes new participant organizations:

Human Frontier Science Program (HFSB); Symplectic Ltd.; Taylor and Francis; American Society for Biochemistry and Molecular Biology (ASBMB); Anianet; Environmental

Mission Statement

ORCID aims to solve the author/contributor name ambiguity problem in scholarly communications by creating a central registry of unique identifiers for individual researchers and an open and transparent linking mechanism between ORCID and other current author ID schemes. These identifiers, and the relationships among them, can be linked to the researcher's output to enhance the scientific discovery process and to improve the efficiency of research funding and collaboration within the research community [http://orcid.org]

To rephrase:

- Aim to facilitate inclusion of data unambiguously identifying authors/contributors in the scholarly record in machine actionable form
- In order to build better information services that will make scholarship more efficient
- And it might reasonably hoped that this will not only create a better end result but also save effort/money in the long run

Stakeholders

Researchers

Libraries

Institutions

Funding agencies

Publishers

Services and agencies working with scholarly information

Repository designers and maintainers

Note that contributor identity is qualitatively different from work(article) identity so more stakeholders than for DOI scheme for example

Contributor not Author

- Many possible roles significant, often badly represented in current systems:
  - Author
  - Editor
  - Programmer
  - Advisor
  - Collaboration Name? (e.g. in high energy physics)
  - Reviewer
  - …
- Similarly, *work* includes articles, data, programs, video...

For this work,
who was responsible for it
and what was the nature of their responsibility?

[Builder]

For this person,
what works have they contributed to,
and what was the nature of their responsibility?

[Builder]

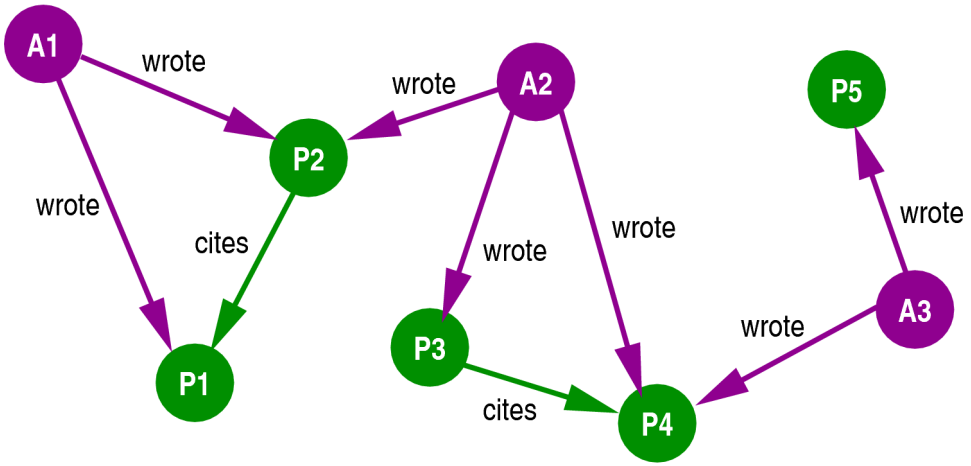
Add to author(contributor)-article(work) graph

Graph is augmented with citation data (work-work relationships). Many metrics, sub-graphs and forms of “related” are then available, e.g.

- co-authorship
- co-citation
- citation ranking
- ....

Even more with usage

No join issues if authors are uniquely identified



Simeon Warner
Simeon Mark Warner
S M Warner
S Warner
Warner, S
Warner, S M
Simon Warner

Other sources of ambiguity

- John Smith
- Much more so with Wang, Zhang, Chen, Lee
- Roman ↔ Chinese mappings not 1:1 or reversible
- Name changes
- Fraud?

How to resolve ambiguities?

- You can tell that I'm *this* Simeon Warner because I'm 192cm tall, was an author of arXiv:1003.1345 and arXiv:0906.2135, have social security number 123-45-6789 (no, not really), and work at Cornell
- Store *some* information in a profile associated with id that can be used to help match in new situations
  - Privacy concerns (no SSN!)
  - Legal restrictions (differ in different jurisdictions)
  - Should not contain more information than necessary for the purpose of disambiguation in the scholarly domain
- Who creates initial profile? Who contributes data? Who has access to (how much of) the profile data?

Roles for repositories?

- Information source (item X has contributors A, B, C with roles i, j, k)
- Information consumer (which John Smith is this? “also by...” etc.)
- Identity registrant on behalf of institution?

(Use term “repository” broadly, could be some other university researcher information system.)

Design of contributor identifier

- Unambiguous
- Persistent
- Universal (discipline, geographic, institutional)
- Contributors, not just authors
- Support use in broad range of contexts: formal, informal, code, data...
- “Open” while respecting privacy
- Some degree of control by contributor (safety)

After [Builder]

Use at arXiv – a disciplinary repository

- Associate internal identity with ORCID (author choice, admin effort, automated lookup, feed from external service?)
- Expose author–article associations (including DOI in many cases)
- Query to resolve ambiguities within arXiv
- Query and connect to external services based on ORCID (in particular within HEP and Astrophysics with Inspire and ADS)

Use in eCommons/VIVO – institutional services

- Associate internal identity with ORCID
- Serve as registration, correction service for Cornell researchers (follow best interests of researcher and insitution). When/how does someone first get an ORCID?
- Expose contributor–work associations
- Query and connect to external services based on ORCID

Why is ORCID exciting?

- Last year I predicted that there would be many separate author identifier schemes. Key commercial companies (TR and Elsevier) were pushing their own (closed) systems
- ORCID has the promise of creating a single scheme and system that will be used by publishers, aggregators and academic insitutions. Even with limited services the shared identifiers will be powerful
- ORCID is being built by a mix of commercial and academic representatives so is well placed to address the needs of all stakeholders (Please let me know what you think!)
- (I am happy for my prediction to turn out wrong!)

Questions?