



Open Access and repositories

Beyond green and gold



Jens Vigen

CERN

Geneva, Switzerland

Global Access to Science
Scientific Publishing for the Future

IATUL, Stockholm, 11th June 2007



“Standing on ye shoulders of Giants” eScience requires digital shoulders

Scientific progress

- New findings were always built on previous results
- Adequate access to information is as needed in eScience as it is in science, but quicker, deeper, and more accurate

Open Access revolutionized the access to information

- Preprints are the main vehicle of OA HEP information exchange
- HEP is now moving towards Open Access publishing

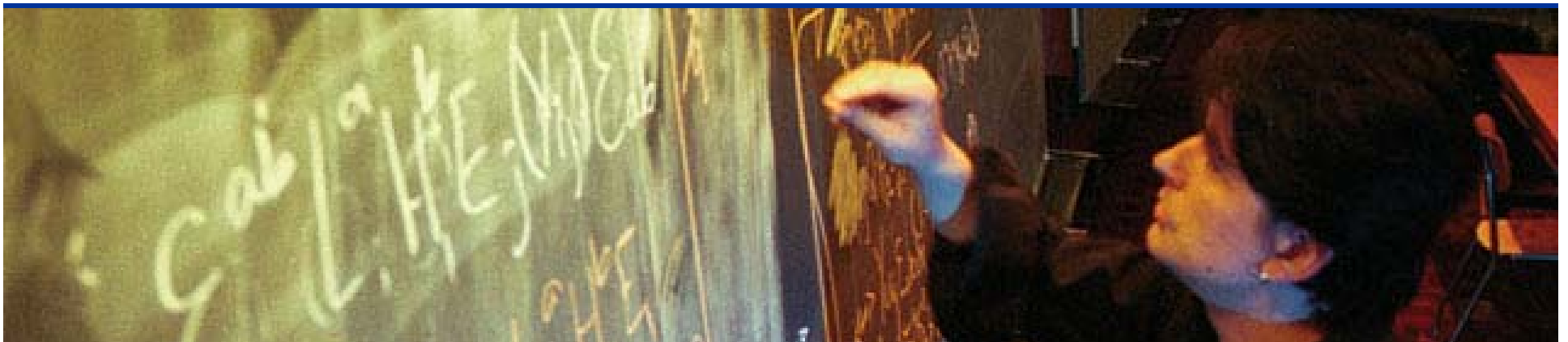
A long-awaited watershed

- More than 15 years after the invention of the Web scientific information remains an electronic clone of the paper era
- Specialized libraries can play a pivot role in preparing the route for their communities towards eScience



Scientific information provision in the era of eScience

- Full text and data-mining applications
- Detection of *relations* between articles
- Treatment of large datasets for statistical and citation analyses
- Identification of popular and influential articles and authors with complementary ranking criteria; alternative metrics to ISI
- Access to numerical information from figures and tables within published articles
- Offer integrated access to primary scientific data





The “Digital” shoulders of giants

HEP as an example

- Infrastructure for repository of scientific information
 - There is urgent need for an integrated repository for the HEP community employing state-of-the-art technology for storage, retrieval and information analysis
- Entire corpus of the HEP information in one place
 - *E.g.* the CERN Document Server hosts today 915 000 entries; half of which are catalogue records (just metadata) and the other half are objects freely available for download: full text articles but also slides, videos, photos, etc.
- Current priority
 - Empower the repository with new technology and content; enabling researchers to explore information matching the emerging expectations of the eScience era.



Transforming our library web sites

Radical trust

Lara Cohen, Dec.15 2006

- I wish I could show you examples of exemplary academic library Web sites, but I can't. There aren't any. Yet.
- Get ourselves moving in order to stay relevant with today's users
- The service is about our users, not about us
- Library web masters will be replaced by blogs, wikis and RSS

Library 2.0

An academic's perspective →





Current perceptions and new requirements

- Survey user perception of present HEP information systems
- Assess user requirements and preferences
- Learn nitty-gritty details for short-term (easy and feasible) improvements of current systems
- Look for the killer application(s) of the next years

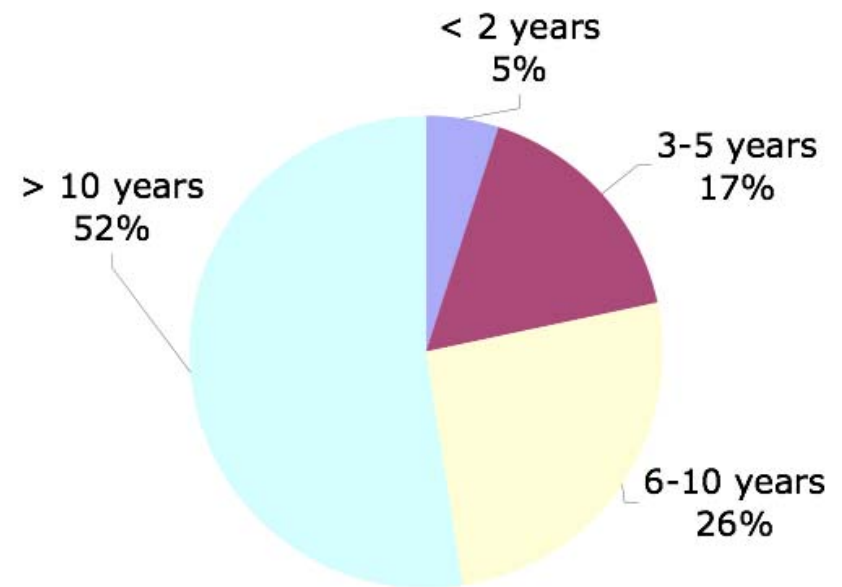
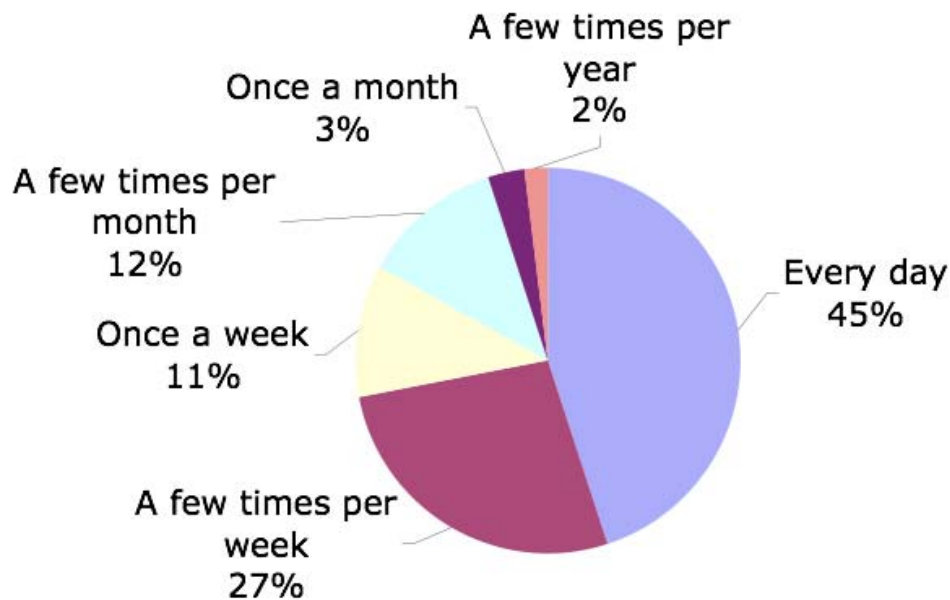




Highly-qualified answers ...adding up to more than 2000 replies!

Frequency of use of HEP Information Services

Experience with HEP Information Services



Users are highly concerned:

43% wish to receive the results via e-mail

89% answered to two or more of the nine “free-text” questions



Which system do you use the most?

3% Commercial services

- ~ 0% pay databases
- 3% publisher portals

11% Internet search engines

- 11% Google



86 % Community services

- 28 % Subject repositories
- 58 % Specialized libraries





What's on a user's mind today?



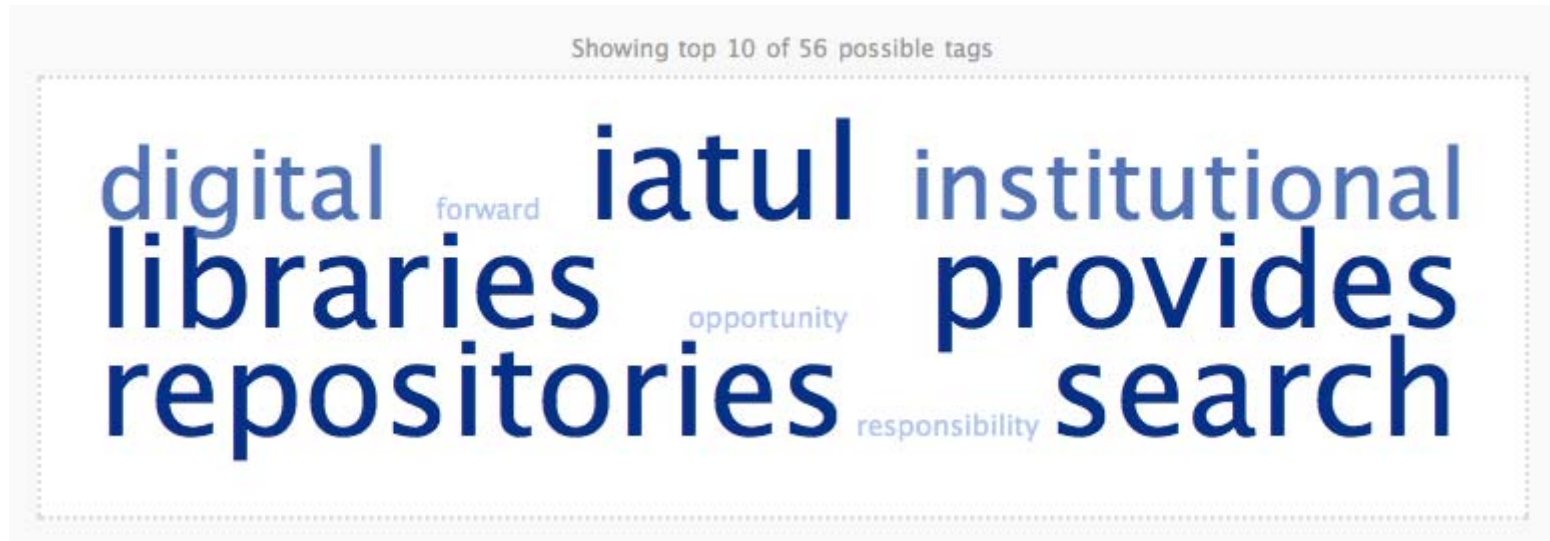


Which changes do you expect?





IATUL is head on!

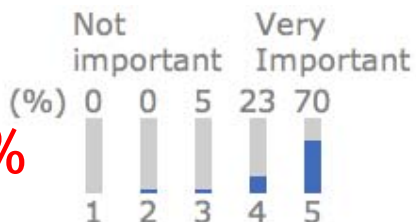




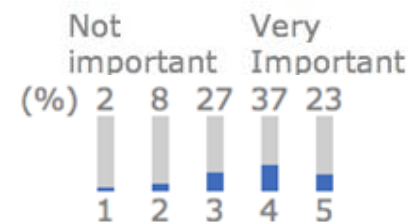
How important are these features of an information system?

Depth of coverage

93%

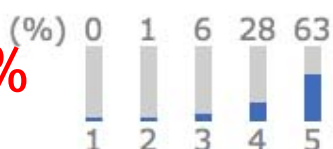


Submission interface

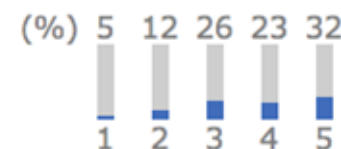


Quality of content

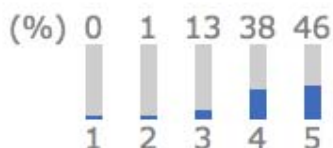
91%



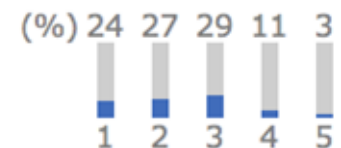
Citation analysis



User friendliness



Multimedia content

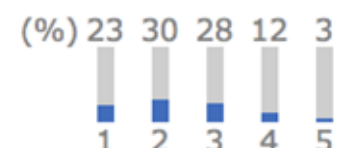


Access to full text

94%



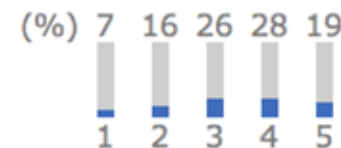
Personalisation



Speed to find what you want



Keywords and classification

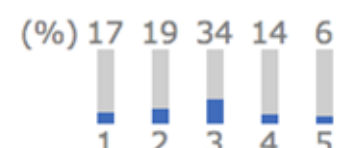


Search accuracy

93%



Collaborative tools





Which changes do you expect?

Summary of recurrent and inspiring answers

- Seamless (open) access to older articles via a unique portal
- Improved full-text search
- Indexing of conference .ppt slides (interlinked with the corresponding article)
- “Publication” of “ancillary” material:
 - Data in tables, figures
 - Correlation matrices
 - Data (high-level objects)
- (A new kind of) Peer-reviewing overlaid on subject repositories
- “Smarter” search tools



Any features you would dream of? (I)

Cover the **contents of all famous journals** going back to their starting dates and a database which contains something like flow diagrams showing how certain articles initiated further research. (Papers citing the original ones etc)

A more **clever system of searching for a paper that is "connected"** by title, citations, references, ... to a given paper

Follow the 'paper trail'. **Citations/References linked** to be a single click away inside of the article and open access from anywhere and include peer-reviewed journals...



Any features you would dream of? (II)

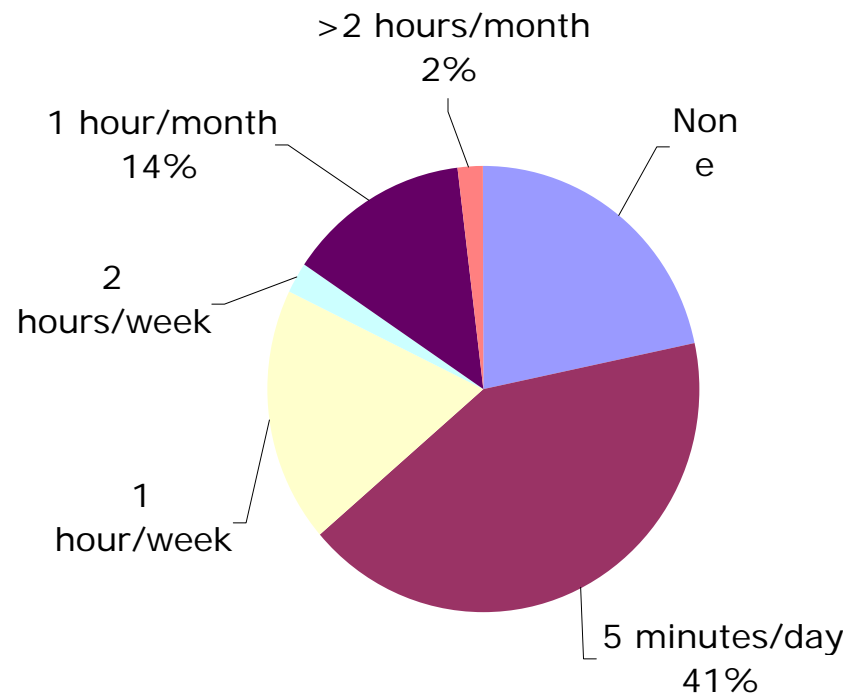
A weekly alert of all the new preprints and publications in certain predefined (but rather general) fields, sent by email, in a compact form (html), with the title, first few authors, with a hidden link to a page with the abstract and a link to the PDF file, similar arXiv today

Access to code fragments which could accompany a plot or equation. This would make it easier to re-use results without making many errors.

Instead of a paper centric system it would be nice to have an idea/topic centric information system. Following the ideas of wikipedia/myspace, each idea in HEP would have its own page, with, for example: a brief outline of the idea, a list of the most recent review papers, a list of the latest cutting edge research in that area, a summary of experimental results, a list of people working in that area with links to their personal pages.

Web2.0 applications

If a simple web interface would show you an article and offer a set of categories to which it could belong, how much time would you spend in this tagging system to give a service to the community?





Vision (I)

Build a complete HEP information platform

- In collaboration with all stakeholders;
- Integrating the content of present HEP repositories and databases to host the entire body of HEP metadata and the full-text of all HEP OA publications, past and future;
- Developing and deploying new approaches to automatic subject detection, key-wording and classification of articles





Vision (II)

Enabling new full-text and data-mining applications on all publications

- Detecting **relations** between documents carrying similar information, or produced by an overlapping set of authors, or referencing a similar set of papers;
- Creating datasets to exercise new hybrid metrics to measure the impact of articles and authors and evaluating the scientific production of research groups;
- Extracting numerical information from figures and tables within published articles

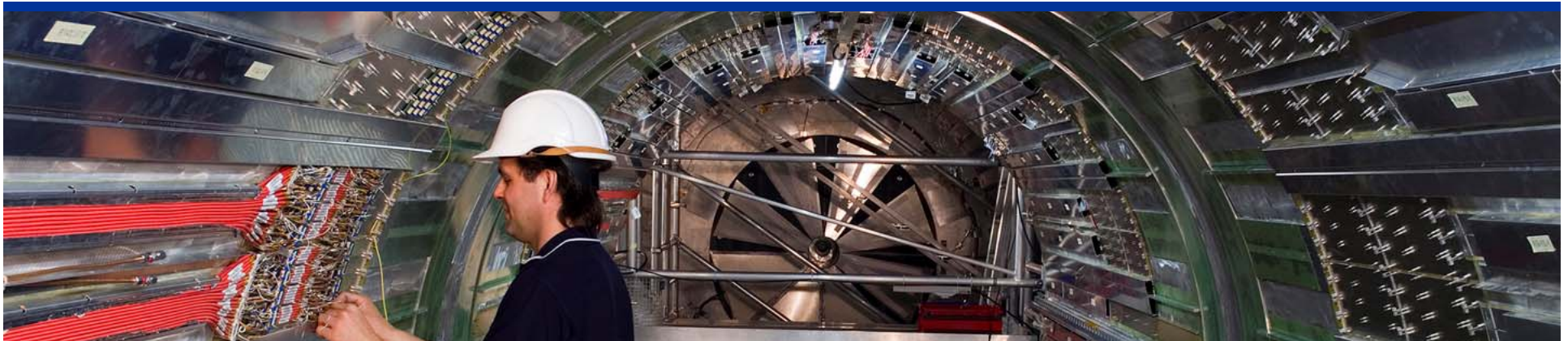


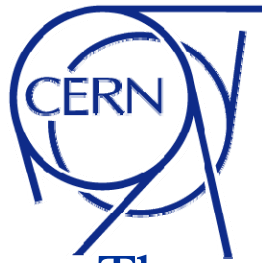


Vision (III)

Demonstrate and deploy Web2.0 applications in the domain of sciences

- Involving readers/authors in subject tagging, altering automatically assigned keywords/classification codes;
- Enabling the possibility to review and comment on articles, adding links to additional documents or other digital objects;
- Providing collaborative tools for effective management of co-authorship within distributed collaborations;
- Introducing community-based alternatives to the established peer-review system





Conclusions

- The era of eScience is still ahead of us
- eScience requires Open Access
- Web 2.0 and Library 2.0 are buzzwords, but a change will come. Users want to be involved
- Libraries must regain control over the literature
- Librarians have the opportunity to play a key role in the era of eScience ... provided we listen to our users!
- HEP is an ideal test bed for new ideas and concepts – express interest if you would like to join

