

School of Information and Library Science
Summer 2008 e-publishing project
London Journal
student
Cedomir Kovacev



**monday
june 16th**

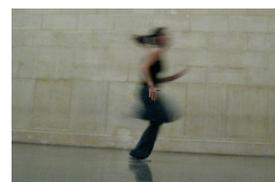
“The Digital Transition and All That” was the title of the presentation by David Nicholas from the CIBER Institute. The informal character of the title indicated the perceived need for a new approach suitable for the phenomenon we are all accustomed to referring to as “the digital age”.

Nicholas concentrated on the consequences of digital transition and changes to the information environment as it has moved to the virtual space. In his view the information environment evolved on many different levels:

- mediated information environment is increasingly becoming non-mediated. This contention is arguable, since the nature of mediation has changed, but mediation itself still exist, through hardware, software, content providers etc. There is only an illusion of non-mediated environment, since there is a sense of so much power placed into our hands. Katherine Hayles in her seminal work “Writing Machines” explores extensively this issue.
- virtual, interactive and social networks replace bibliographic systems,
- everybody performs a search,
- little choice is replaced by massive, overwhelming choice,
- the amount of available information is increasing at a stunning rate,
- stability of the old environment is replaced by ever increasing volatility,
- the new environment is dominated by the unknown,
- user becomes consumer. This raised a red flag in my mind. Are we all going to become consumers? I find the idea of treating the entire information environment in

*After all Web 2.0 is all
about making money,
isn't it?*

Peter Murray-Rust



*Martin Creed, Work No. 850,
Tate Britain, London 2008*

such a way deeply troubling.

The focus, Nicholas argues, has to be on the digital consumer. The content is not a king anymore, but consumer is now a king. Consequently, researchers should concentrate on user studies.

This sounds to me as unnecessary glorification of supposedly omnipotent market approach to everything. I do share an excitement about the new digital age but I do not believe in placing a market value on everything. Usability studies are certainly necessary and that just proves the point on a new kind of mediation that is silently taking place before our eyes.

Emergence of E-books will fast-forward digital transition process, Nicholas argues. That could lead to disconnection with the user base, since users are more remote, anonymous. User's behavior has dramatically changed and that has not been acknowledged.

There was an emphasis on problems of access, but not enough studies on the impact the access had. What is good or bad information seeking and how do we know uninterrupted access is helping us?

Libraries do not have money to invest into user research like Tesco supermarket, so they should work with publishers on that.

Libraries face potential nightmare scenario – they could experience disconnection from users, publishers and faculty. In order to avoid that they should recognize that users are consumers who have choice, they should move from monitoring activity to monitoring users and understanding their information-seeking behavior. They should also partner with publishers and that way get user data and branding support.

The implicit idea here is that libraries should compete with commercial environment by, among other things devising new branding strategies. I don't think libraries are supposed to enter into such a competition at



*South Bank Graffiti, Waterloo, London
2008*

all, since their function is not selling of information but provision of an efficient access to it.

The second part of the presentation dealt with CIBER's Google generation study that got a lot of media attention, at least in the UK. The study focused on activity metrics (pages viewed, time spent etc.), information-seeking characteristics (type of content viewed, searching style etc.) and user characteristics.

The results were very interesting and revealing on many levels. Half of the visitors to scholarly sites are robots, which mimic human behavior to get entry. There are national, age and gender differences. Germans turned out to be the best searchers. Elderly users have most problems with search with two-thirds of zero results, while women are more likely to return to a site.

As far as scholars are concerned no conceptual framework that would guide them was found, there was uncertainty about quality of searches and little time was spent on evaluation of results. A surprising finding was that even scholars have problems usually identified with more general users. Scholarly behavior can be described as active, bouncing, navigating, promiscuous, diverse and volatile.

→ a visit to the British Library

Nothing could have been more appropriate for the first day then a visit to somewhat architecturally eclectic but nevertheless in many ways impressive British Library. After a brief tour of the building there were several presentations focusing mainly on the ways British Library is coping with rapid changes in information environment.

Emergence of Open Access is slowly changing still predominant subscription model. Different Open Access models are emerging and research councils as well as some prominent institutions (Harvard University for example) are increasingly supporting Open Access idea.



British Library, London 2008

Some governments from countries with strong publishing industry do not support Open Access, because of potential loss of tax revenues. That must be the reason why European Commission is still ambivalent about this issue.

Publishers are quick in accepting new technologies and are becoming more efficient. Usage of standardized metadata was mentioned as an example. Also publishing companies are merging and that way they are becoming more powerful. In blunt words this means that publishing companies are finding ways to monopolize the market to a certain extent.

In this context the issue of data mining was mentioned. Publishers take control not just over the articles, but also over supplemental data and software used for research. It is clear that such practice cripples scientific research and exchange of ideas on a global scale but the speakers at the British Library were careful not to use that kind of language since in the present situation everybody is kind of supposed to dance around real issues in order to not upset the big money makers that hold the keys to the treasure.

How is emerging semantic web going to influence research and all of these issues around it? This is a matter of speculation at this point. I am getting my first ideas about the paper I am going to write.

Lee-Ann Coleman, head of the STM division at the British Library, served us with some useful and stunning facts. 99% of STM journals are online. 2000 publishers are fighting for a share in a \$5 billion market, and just five of them emerge with 50% of revenues! Open Access is only accounting for less than 5% of published articles.

British Library is involved in the development of UK Pub Med Central, a digital archive of biomedical and life sciences journal literature. It provides free access and is driven by deposit mandates by funders.



British Library, London 2008

The Library is involved in development of data mining and data sets frameworks. It sound like the library is taking bold steps into preparing for the future but it is severely limited by publishers' hold on journal publications.

Benjamin White talked about mass digitization, currently under way in the British Library. It is the project undertaken with Microsoft Corporation and it is in limbo right now since Microsoft gave up on the whole idea a while ago.

There was a short visit to the digitization room. Mr. White explained that images are saved in JPEG 2000 format, the reason for that being available server space as well as enormous amount of scanned images. The loss in quality is an issue, he admitted. TIFF format, commonly used in printing would be preferable, but it results in several times larger file sizes.

**tuesday
june 17th**

Our energetic and charmingly passionate host, professor Anthony Watkinson, introduced Sally Rumsey, ORA (Oxford University Research Archive) Service and Development Manager, as having a “very broad view” on digital repositories. It turned out that this is just another way of saying that Miss Rumsey argues for something that should be fairly obvious – research articles as well as supplemental data should be freely available to their creators as well as to the wider scientific community.

Scholars at Oxford University are primarily interested in preservation and not in achieving visibility, hence a different role of repositories at Oxford.

Repository uses Fedora, a management tool for digital preservation. They are using a lot of semantic web technologies, such as RDF for establishing multiple relations between documents.



*Martin Creed, Work No. 850,
Tate Britain, London 2008*

Given the idea behind a repository concept it is not surprising that Miss Rumsey prefers Open Access. The benefits of having a repository are many: Universal Unique Identifier applied to articles, persistent URL's, possibility of a cross-search with other repositories, effective research over several disciplines etc.

Research data is most likely not going to be held at one place – repository will just reference data held somewhere else. ORA repository uses a program that harvests metadata from different databases. Automatic metadata generation does pose a problem though, since finding a balance between quantity and quality is not always easy.

Eventually, Miss Rumsey argues, there are going to be many repositories representing different collections brought together by some kind of framework. This is essentially semantic web idea and it is being realized as we speak. Emergence of semantic web is going to make the issue of compliance between different software components of the system obsolete because underlying semantics are going to override today's common problems. This is now definitely a possible idea for a final project.

→ **a visit to Elsevier**

Niels Weertman, Director of Scopus and Scirus at Elsevier talked about the company that provides access to more than 2000 science publications. 250 000 new articles are made available each year, while a number of accessed articles currently reaches 20 million.



Elsevier, London 2008

It is interesting that Duch Royal Library has a copy of Elsevier database; obviously a consequence of an unfortunate fact that a privately held company has a lock on a vital source of scientific research.

Scopus was presented as collaborative project in building a search tool. The company follows what it calls a

“customer-driven strategy” based on market and user research. It was established that interface, ease of use and search capabilities are most valued features by researchers. Collaborative tools are still in the state of development.

Mr. Weertman acknowledged that the nature of connections between information has changed. Static centralized networks are being replaced by ever changing relations between multiple networks. Here is another allusion to the emerging semantic web. It looks like everybody is aware of where the things are going. Decentralization poses a problem for companies like Elsevier, since their mode of survival relies on guarding against “unauthorized access” to the information they consider proprietary.



South Bank Graffiti, Waterloo, London
2008

Elsevier is focused on information content and it is not clear whether it is going to branch off into software development.

The issue of Open Access was obviously not a pleasant subject for Mr. Weertman, but predictably he said that there is always going to be a need for this kind of service, offering more features etc. Fine point, but what about taking research papers funded by someone else and selling them back to the scientific community?

**wednesday
june 18th**

Dr. Mark Patterson, Director of Publishing at Public Library of Science has a background in human genetics and is a forceful Open Access advocate. Free access in his words equals open access. Open Access means free, immediate access, deposition of materials in a digital public archive as well as unrestricted reuse under Creative Commons Attribution Licence. Materials are free for use but the author has to be cited, while no permission is needed for copying. Open Access would create a network of not just individual documents, but of

different databases as well.

In order to do this text-mining should be enabled. Text-mining refers to analyzing a number of texts using computation and finding sometimes unexpected links and relations between phenomena. By analyzing words and phrases within the text connections are established and further explored. Currently text-mining is done through abstracts, but the idea is to be able to perform it using full texts.

Mr. Patterson then talked about PLoS ONE, an international, peer-reviewed, open access online journal that takes advantage of web 2.0 technologies. It is focused on objectively measuring technical quality (it is peer-reviewed), streamlined production (ideally publication process lasts no more than three weeks) and encouraging of scientific debate through community comment and annotation.

In order to provide open access PLoS charges authors or research sponsors a publication fee. It also offers a complete or partial waiver for authors who do not have funds to cover publication expenses. This is one possible model, and the one that seems to be working, although there were some comments after Mr. Patterson's presentation questioning some of his claims. I can see no reason why publication fees wouldn't be included in a research grant and why such publishing model wouldn't be sustainable. On the other hand, it is obvious why some interested parties would rather see the whole project fail.

I was impressed by Mr. Patterson's energy and his ability to convey the strength of his convictions.

The second presentation was by Panayiota Polydoratou of UCL Library Services and was related to Repository Interface for Overlaid Journal Archives (RIOJA). The project involves academics from astrophysics and cosmology who were unhappy with the existing journal publishing scheme. Overlaid Journal Model makes several assumptions: content is deposited and stored



*Martin Creed, Work No. 850,
Tate Britain, London 2008*

in one or more repositories; quality is assured; open access; financially sustainable.

The project included a survey of more than 4000 researchers in order to find out what kind of a model would be most successful.

RIOJA seeks to develop a robust system of quality assurance within the process of submitting papers to an electronic repository, and that way bypass traditional publishing model with all of its pitfalls and restrictions. There is an evidence that a significant number of researchers are frustrated with the current model, which generates errors, different versions of a paper and numerous restrictions on published materials.

The project should result in a toolkit that would improve quality assurance process in electronic repositories.



South Bank Graffiti, Waterloo, London
2008

→ a visit to **Joint Information Systems Committee (JISC)**

The series of five dense presentations given by JISC staff seemed to leave everybody thoroughly confused. It was the consequence of a language heavily infused with managerial lingo that left most of us listeners wondering what this organization is actually about. We heard that JISC is into “network and access management, info environment, e-resources, e-research, e-administration, business and community engagement”.

Once the official part of the whole affair was over and the question time came up the things became clearer. JISC brokers licensing agreements with more than 80 publishers in the name of universities, libraries, schools etc. JISC makes sure institutions get “a real value”, whatever that might mean, supports accessibility standards, open URL as well as institutional repositories, deals with copyright issues. It is a publicly funded committee with financing coming from institutional members of the committee.

JISC is also involved in frequent, some of them high-profile, studies examining current trends. As an example, we heard about an ongoing study focusing on a way e-textbooks are accessed and used.

JISC activities go beyond scholarly communication, but as far as this aspect is concerned the wish for finding a middle ground is admirable, but that does not change the basic fact that the currently used system is broken. In a way JISC is helping publishers continue with their questionable practices.



"2001: A Space Odyssey", The Stanley Kubrick Archive, University of the Arts London, London 2008

**thursday
june 19th**

Steve Hall, formerly of Blackwell Publishing gave a thorough and unapologetic defence of publishers' perspective.

Publishers use licensing because existing copyright laws were written for print and are therefore insufficient for digital editions. That way publishers protect their electronic content against unauthorized use. Cornell University's holdings of multiple copies of textbooks as well as British Library's attempt at direct digital document delivery were cited as examples of such practice. Piracy and fraud in developing countries are also emerging problems for publishers in this regard.

The issue of Big Deal is actually pretty straightforward and Mr. Hall didn't mince words while talking about it. Publishers get stable and predictable income while a customer commits to a subscription to a package and that way supposedly gets lower price. Potential benefits for a customer are wider reach and readership, greater use of non-core journals and the possibility of improved impact-factor.

Libraries in Mr. Hall's opinion sign on Big Deals because of predictable expenditure, lower price increase, access to more journals, support for interdisciplinary research,



South Bank Graffiti, Waterloo, London 2008

meeting demand from faculty and students etc.

Mr. Hall had a different view on the issue of rising subscription prices. He claims that percentage of expenditure for acquisitions in libraries haven't changed in the last 20 years. On the other hand he acknowledges that libraries are tied to subscriptions from a few large publishers, meaning that a talk about supposedly existing choice is largely removed from reality.

The last issue discussed was emergence of e-books. Publishers don't exactly know what to do, how to sell e-books. They are different than journals and are not sold by subscription. Reference works are mostly going to go digital and the same stands for monographs.

Mr. Hall ended his talk with a telling statement that librarians are the worst negotiators for the best deal. The question is should they enter the game and wage legal and market battles they are certainly bound to lose or should the journal publishing system allow for unrestricted access to information that should be in a public domain in the first place?

→ a visit to the Nature Publishing Group (NPG)

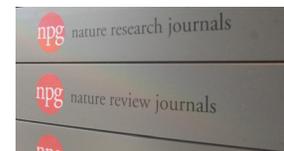
NPG publishes and maintains high impact journals, gateways and databases. It is well known for their innovative approach and introduction of cutting-edge technologies.

NPG provides for both individual (print or online) and institutional access. Several newest features were mentioned. Among them, Corotea is online reference management for scientists while Nature Network is an online networking tool for scientists. They also recently introduced podcasts, blogs, web feeds etc.

They digitized the entire Nature catalog back to 1869 and launched regional and national web sites.



*Martin Creed, Work No. 850,
Tate Britain, London 2008*



*Nature Publishing Group,
London 2008*

friday
june 20th

→ a visit to Oxford University Press

Claire Dowbekin introduced Oxford Scholarship Online and OUP e-book projects.

There are print and online editions of Oxford English Dictionary. There are several subscription options for online edition. The Dictionary is online since 2000. New search capabilities are constantly being introduced. It is read widely, across professions and is the most used Oxford Dictionary online.

Dictionary of National Biography (DNB) involves 10000 contributors. It uses XML as standard and is highly searchable. It has direct link to other resources (museums etc.).

OUP offers four year contract to British public libraries for print and online access. It adapts the price of the product to the budget of a library and it allows for selective or customized subscriptions instead of forcing a purchase of a bundle.

Fiona Bennett talked about Oxford Journals and Open Access. OUP publishes around 220 journals. Two thirds of them are published in partnership with learned societies. Oxford Open is based on Gold Open Access model (free online access, free reuse for non-commercial purposes). There is an optional full or hybrid Open Access model. Nucleic Acidic Research Journal was cited as a success story. It features unlimited reuse for research and educational purposes. Authors retain copyright and self-archiving is not restricted. Author charges are becoming dominant part of the revenue.

Michael Popham of the Bodleian Library described the Oxford-Google digitization project. Google was chosen because of an emphasis on access, not conservation. They identified extensive holdings of out of copyright material and began with 19th century material. The books are selected in Oxford while Google deals with



*Martin Creed, Work No. 850,
Tate Britain, London 2008*

metadata checks, digitization, quality control, OCR and indexing. The output is in the form of large raw color images in JPEG 2000 format and uncorrected OCR.

**monday
june 23rd**

Chris Beckett of Atypon gave a presentation on online hosting of e-journals and e-books.

The quality level of publishers' online presence is related to their willingness to invest into new technologies. There are two types of hosting providers. With the Aggregators (Metapress, Ingenta, Connect, Atypon Link) publishers share the same platform and the interface looks pretty much the same regardless of the content. The Customized Builders (Highwire, Scitation, Atypon, Metalink) are more flexible and scalable and typically employ several servers.

Hosting provider's offer is becoming more granular, a consequence of a slow shift from subscriptions towards accessing components of an article (video or audio content, supplemental data etc.).

Hosting providers offer a number of services: standardization of content, link generation, metadata distribution, content indexing, browse and search user interface, web design, branding, access control and licence management.

Content management is either done by the provider or by a customer through real time control of access, content, management etc. Providers pay special attention to content online visibility by introducing search, alerts, recommendations and syndication.

Providers are responsible for allowing access depending on a subscription model.

There are three types of aggregators: Hosts (the content owner's or their provider's web site), Gateways



"2001: A Space Odyssey", The Stanley Kubrick Archive, University of the Arts London, London 2008

(web sites that index and link to hosts) and Full Text Aggregators. There is a tension between two models: fixed article an author is mainly interested in because of attribution, and the article with added features such as community links etc.

Roger E. Schonfeld, the manager of research for Ithaka organization talked about humanities and social sciences in the context of the transition from print to electronic format. Ithaka organization is mainly focused on employing collaborative technologies for communication between universities, libraries and publishers.

Attitude towards electronic only environment are locally defined and based on space and budget. There is a broad consensus across disciplines about the importance of preservation of electronic journals. Electronic archiving is currently a priority for librarians. Technology for digitizing is available, but there is no collaborative effort for preservation, since it is not clear who should be responsible for that.

There are signs that mergers typical for publishers of science journals is going to become common in humanities as well.

Mr. Schonfeld's presentation was very dry and relied heavily on surveys and raw data. Since the focus was on humanities it is surprising that discussion of a broader ethical and social context was missing.

→ a visit to Thomson Reuters

Bob Stembridge introduced the Derwent World Patent Index.

The value is added to otherwise freely available information through the provision of indexing, classification and improved search.



Thomson Reuters, London 2008

Protection of rights is central to the idea of patent. A patent expires after 20 years. A complete disclosure of invention is required in order to get a patent. 1.5 million patents are issued every year. 38000 are processed by Derwent every week. They receive patent information in a variety of media and process them by correcting data, classification application, abstract creation, manual coding and indexing.



Thomson Reuters, London 2008

Patents are cross-classified across all relevant patent aspects.

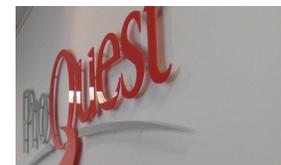
As far as academic research goes Thomson Reuters provides background research through Web of Science, Biosis, Derwent etc.. Managing of research results and collaboration are supported by End Note and End Note Web software solutions. Article submission and management are available online.

There was a short demonstration of a manuscript submission process as well as manuscripts management features.

→ **a visit to ProQuest, Cambridge**

**tuesday
june 24th**

ProQuest is largely known for aggregation of periodical content. It licences content from a large number of publishers and authors. The company selects the content and creates tools for access to relevant information. Search function is tailored to specific disciplines.



ProQuest, Cambridge 2008

Duncan Cambell talked about ProQuest's Literature Online project. It is a fully searchable database of English literature, covering poetry, American literature etc. For each entry there are cross-reference links to reference works, biography, bibliography, index etc.

Dan Burnstone presented ProQuest's digitization

process. Using the example of a newspaper page he explained the process and benefits of article zoning. Sections of a page are indexed separately which allows for more precise searching as well as a potential article-level blocking in case of copyright issues.

In case of the Medieval Collection the text is re keyed and then coded and prepared for the web.

John Johnson Collection (the Archive of Printed Ephemera) is another digitization project. Every item was scanned at a very high resolution and is richly indexed. The project was funded by JISC and Bodleian Library.

Proquest provides an option of “permanent archive”, where a customer can get the copy of the content in case they want it with an assumption that ProQuest would not be able to provide it sometime in the future. Another, more common model is subscription.

Roger Tritton talked about ProQuest’s position within the changing aggregation business. ProQuest has large full-text periodical collection and what it calls an “intuitive platform”. It owns ABI/Inform, the biggest and the oldest business database, covering 4000 journals.

Aggregations in general are growing and expanding beyond periodicals, they are becoming more specialized.

After getting the material from publishers ProQuest is doing indexing, abstracts, citations etc. The company introduced so called “deep indexing” where for example illustrations or photographs are supplied with a full caption with article citation, link to the abstract, category, subject, geographical and statistical descriptors.

→ a visit to professor Peter Murray-Rust

The absolute highlight of the Cambridge trip and possibly of the whole London e-publishing course, the brief lecture by professor Murray-Rust was an exercise in precision as well as in unbounded idealism. It was a



South Bank Graffiti, Waterloo, London
2008

much needed breath of fresh air after a market-driven list of Powerpoint selling points delivered earlier in the day in ProQuest offices.

Professor Murry-Rust says that management of digital identifiers is going to be central to our profession.

The answer is not clear to the question who owns copyright to articles - author or publishers?

He demonstrated Open Source software named OSCAR, which is designed to capture and analyze chemical data within articles and visualize detected changes. The alternative service is "Chemical Abstracts" for which one has to pay. Is it possible to go straight from chemist to a reader? Publishers are main barrier to that; "Chemical Abstracts" has to make millions by selling abstracts.

Authors have to start thinking about changing their relationship to publishers.

Professor Murry-Rust also presented the CrystalEye project which aggregates crystallography from various web resources and makes it available for easy browsing and searching.

Publishing copyright terms are murky. Academics think most of the things they publish are free. Publishers on the other hand routinely hold copyright over both articles and supplemental data which is essential for scientific research.

Publishers' business model is the problem. There is also a culture that considers papers not published by reputable publishers as less valuable.

Within less than two hours professor Murry-Rust managed to hit the main sore points in the current crisis in scientific publishing and openly challenge ethical issues related to it. He also demonstrated the power of Open Source technology and indicated efficient ways in which scientific data can be freed from business-model con-



*Professor Peter Murray-Rust,
Cambridge 2008*

straints. I found it remarkable that a scientist who deals with something that is I assume obscure to most of us has an ability to demonstrate the issues in such a clear and convincing way.

**wednesday
june 25th**

Ruth Jones, Director at Ingram Digital, talked about E-books. Ingram supplies digital content management, hosting and distribution for publishers, libraries, retailers and higher education. It has 130000 downloadable e-books in its catalog.

One of the first problems she identified was the issue of unique identifiers or ISBNs. Some e-publishers refuse to use unique identifiers. There is also an ongoing battle between different platforms.

There are several types of e-books: reference via web browser, basic e-book download, e-book with added value and audio book.

Miss Jones also touched on using widgets (small self-contained pieces of software that behave like applications) as a marketing tool for previewing content. Is there a place for widgets in a library setting? Why not, in my opinion, although there could be some copyright issues related to that.

Aggregated platform, such as Ingram, has obvious advantages for a customer, says Miss Jones: cross publishing search, ability to buy cross-publisher subject collections, consistent metadata etc.

Big Deal concept is not a prevalent model in e-book industry (yet).

Development of software is making possible an e-book as a personalized experience - making highlights, search across multiple books, export to other applications, change page layout etc.



South Bank Graffiti, Waterloo, London
2008

→ a visit to the Office of Public Sector Information

Mr. Alan Pawsey gave a presentation on e-Government. The role of the Office of Public Sector Information is to encourage persons and private sector to use government information.

Mr. Powsey showed us the copies of London Gazette, a government publication, published from 1665. All the issues of the paper are available online.

There are six legal deposit libraries in the UK where government records are kept.

Government offices and institutions are widely present on the Web - UK Parliament, Office of Public Sector Information, Scottish Parliament and the Scottish Government, DirectGov web site, 10 Downing Street web site.

Private providers, such as Lexus-Nexus, pay for XML feeds provided by the Office. All text is already encoded. Not all papers are automatically released. Royal family papers are some of them.



Office of Public Sector Information, Whitehall, London 2008

→ the Virtual Scholar Conference, day one

thursday
june 26th

Geoffrey Bilder started the session with an entertaining talk on the state of publishing industry.

In the digital age, Mr. Bilder claims, publishers' responsibilities are shrinking - they are mainly responsible for trustworthiness of the materials. Researchers, on the other hand are faced with both a deluge of articles and a convoluted access procedures on web sites that consistently fail usability tests.



Second Bloomsbury E-Publishing Conference, UCL, London 2008

While Web 2.0 enabled effective sharing of knowledge, but also resulted in an uncontrollable mass of information, We 3.0 should help us extract relevant information, through data mining for example.

Web is chronically lacking in ability to establish trustworthiness of the source. This is the consequence of the fact that identity is a fluid category on the web - it is difficult to establish its origins and multiple identities are possible. OpenID was mentioned as an example of a new distributed identity system.

Related to this issue is a nonexistence of a mechanism to identify peer-review articles. Mr. Bilder argues that such a practice is necessary and could follow the Honor Code for medical and health web sites.

This presentation mainly touched on semantic web ideas while focusing in the second part on persistent identity issues that have been undermining the web since its very inception.

Ian Rowlands presented the Google Generation project by the CIBER Institute. The project examined user behavior related to accessing electronic content.

One of the most important findings was that demographics related to user behavior are not just about age, as is usually assumed, but also about gender, subject etc. Just around 20% of youngsters aged 16-25 are pushing boundaries, while most others are passive consumers.

The second focus of the study was at how people use digital libraries. The study found no difference in user behavior between undergraduates and professionals. Digital libraries are replacing physical libraries, but their confusing interfaces make navigation extremely difficult.

Mr. Rowlands concluded with an emphasis on the importance of understanding information seeking behavior,



*Martin Creed, Work No. 850,
Tate Britain, London 2008*

something commercial companies have recognized almost 20 years ago.

There is a question, though, related to employing marketing research methods to examination of user behavior in library or scholarly communication. Commercial research is aimed at luring people into buying things.

David Nicholas, also from the CIBER research group, talked about digital transition, CIBER research methods and the findings on the way students and scholars are accessing and using E-journals.

The first half was mostly covered in the lecture Mr. Nicholas delivered on the first day of the e-publishing course. It centered on understanding of the phenomenon of digital transition, tracking user behavior through digital log analysis, demographic studies and information-seeking characteristics.

The findings on E-journal access and usage included: there is a huge demand for e-journals; home audience is relatively small compared to the foreign audience, while the students are the biggest users; power-browsing is the prevalent mode (viewing versus reading); half of the time is spent navigating etc.

Touching on the issue of information availability Mr. Nicholas wondered why the people expect free access to information while they readily pay for a toothpaste in a supermarket. Following this concept people who seek information are reduced to shoppers and information becomes just another commodity. In his excitement for effective user research, proved to be so successful in commercial milieu Mr. Nicholas makes a crucial mistake of applying similar logic to scholarly communication and its products.



South Bank Graffiti, Waterloo, London
2008

Peter Burnhill, Director of EDINA and Head of the Edinburgh University Data Library talked about research related to what scholars require and what they actually do as a part of their scholarly activity. In his words, scholars need access to licensed material as well as access to Open Access repositories.

Virtual scholar tasks include several steps: discover, locate, request, access and use.

Research into scholarly activity is not sufficient, since the target audience is not precisely defined - the sample often amounts to 5% of what direct-mail typically reaches.

Carol Tenopir presented findings of a survey on faculty reading habits in the context of the transition from print to digital. The survey originally examined reading habits of more than 40000 scientists.

Main conclusions were that faculty members read more within the same amount of time, use more ways to search for and locate sources, use both electronic and printed sources, read higher percentage of older articles.

Scholars use several ways to locate information: browsing, searching, following citation, recommendation from other user, alerts. Personal subscriptions took a huge hit while percentage of library provided material increased significantly.

Interdisciplinary readings are increasing with increasing number of sources and tend to be tied to electronic sources.



*Martin Creed, Work No. 850,
Tate Britain, London 2008*

→ **the Virtual Scholar Conference, day two**

Andrew Walkingshaw from the Department of Chemistry, University of Cambridge, opened the second day session with a blast. It was a powerful, yet down-to-earth demonstration of technology that promises to break down existing information barriers. Mr. Walkingshaw is a disciple of professor Peter Murray-Rust and his presentation touched on many issues already covered in professor Murray-Rust's lecture in Cambridge.

Stating that semantic web adds information about the links between bits of data, Mr. Walkingshaw explained the role RDF (Resource Description Framework) plays in expressing links between URIs (Universal Resource Identifiers). RDF adds semantic value to information.

Mr. Walkingshaw then demonstrated on a number of Open Source software applications (CrystalEye, Processing, Oscar) how data from separate sources can be harvested, interpreted, integrated and presented.

Richard Kidd of Royal Society of Chemistry talked about Project Prospect that uses new technologies to enhance journal experience.

He described several steps taken. They used HTML to link papers with same entities. RSS feeds were included and text-mining practiced using Oscar application. The final result was enhanced HTML, database and enhanced RSS.

Vince Smith from Natural History Museum London described challenges and opportunities taxonomist faces in a new digital environment.

Currently 80% of species are still not described. Besides that people and data are highly fragmented and distributed. Publications are not connected and use different models of citation. But all the data is linked by



*South Bank Graffiti, Waterloo, London
2008*

taxonomic names.

Several projects that are trying to deal with these problems were mentioned: Plazi, Encyclopedia of Life, Scratchpad, Biodiversity Heritage Library.

Plazi uses data mining, while Scratchpad is a web site and a publishing platform for taxonomic communities.

Vanessa Murtough from Institute of Physics Publishing described the research in developing a new journal platform.

Key trends the research indicated were isolated use of web 2.0 technologies, personalization, filtering and recommended content. On that basis the changes were introduced and included social book marking, tagging content RSS feeds, last 10 views, last 10 searches. For filtering the site uses Physics Astronomy Certification System.

The site also introduced several features in order to enhance personal experience - tagged articles, saved search, alerts, downloads and personalized saved articles.

Comments and ratings features present a challenge since it is difficult to know if they are competent.

Professor Roger Griffin from Oxford Brookes University talked about the web in the context of humanities. Web has emerged as indispensable research tool and reference library. It also functions as a tool for communication and trans disciplinary collaboration.

Joy van Baren talked about user-centered design at Elsevier. The process of design includes field studies, diary studies, personas and scenarios. Development of Scopus was used as an example. Steps were identified in the search process. RSS feeds, alerts and personal records were introduced.



*Martin Creed, Work No. 850,
Tate Britain, London 2008*

The discussion after morning session amounted to a clash between publishers and scholars. The question of access to primary data on which papers are based was raised. Scholars are not happy with being denied access to primary data by publishers. There is a concern about publishers undermining interdisciplinary exchange.

Mayur Amin from Elsevier started the afternoon session and presented his company's research into online behavior.

In libraries, Mr. Amin said, print is still not rapidly declining, but there is a clear trend towards e-only. There is also a preference for browsing e-journals at home, rather than at work.

Final published article is the most important for scholars while there is a high demand for older articles. Search is becoming more efficient and consequently productivity is increasing.

Michael Jubb from Research Information Network presented research evidence related to sharing research data.

He posed a number of important questions arising from the current publishing practices. Who decides on how the records are going to be managed and how to deal with various article versions? There are also important usability issues: "protected PDF", when should data be freely available, controlled or free access.

One of the major points is the question of who is responsible for funding. Related to that is of course the question of who owns the data. There is a total lack of clarity on this point. What is clear though is that public good is being used for private financial gain.



South Bank Graffiti, Waterloo, London
2008

James Pringle from Thomson Scientific predictably offered a different perspective. He concentrated on a problem of identity in scholarly communication. Percentage of single author papers is consistently declining and a number of scientists who are collaborating with others on a project is increasing.

The existence of multiple identities which has been the norm on the web since its inception is complicating the situation and often blocking effective scholarly communication. Mr. Pringle demonstrated Researcherid, a Thomson service which enables researchers to create an ID.

