Notes on the Public Hearing on access to and preservation of scientific information

30th May 2011, European Commission, Luxembourg

Opening remarks, Mr. Richard Swetenham

Mr. Swetenham explained the context of the hearing. There was a Communication on scientific information adopted in February 2007, followed by Council Conclusions in November 2007. Currently a new Communication is in preparation, which will describe the progress made since 2007, and explain what the Commission intends to do in the coming years. In addition, a Commission Recommendation to the Member States will be prepared. The Communication and Recommendation will be preceded by consultations, of which this hearing is part, and which will also include a public on-line consultation. Several Directives are of importance in this context, such as the PSI Directive, which is currently under review, and where one of the questions is whether scientific information should be included in its scope (as it is now explicitly excluded). There are diverging views on the strengths and weaknesses of the present scholarly system, as well as on how to improve it. The hearing should help to create an overview of these views.

Opening remarks, Gilles Laroche

Mr. Laroche explained the structure and functioning of the hearing.

1. Research evaluation systems

Statement Mark Patterson (Public Library of Science)

Considerations:

- The prevailing methods of research evaluation are stifling research and innovation. Because researchers are judged largely on the basis of the journals in which their work is published, and specifically on the journal's impact factor. A journal's impact factor is a poor metric for predicting the impact of an individual article. The consequence is that researchers are forced to publish in journals that are optimized for print publishing, not online communication.

- Impact evaluation can be done at the article level, with PLoS's work as case in point. Every article gets an automatically updated collection of metrics data. Usage, citations, social bookmarking, and blogging activity are included. This has been received well by authors, who can use that data in grant applications, as well as met with interest with other publishers. The concept of impact needs to be expanded further to include effects on e.g. clinical practice or policy development. Vision is a 'dashboard' of info around each article.

- Encouraging use of impact data at article level will accelerate open access publishing. Open Access (OA) journals take time to build prestige which is influencing where authors submit work. Article level metrics will thus support OA journals, as most of these journals are new.

- New approaches to research communication are possible, focussing on evaluation at the article level. Using journals reputation as impact indicators is a system that has arisen in print. Most articles are submitted to multiple journals before eventually being published, which is slow and wasteful. PLoS ONE peer reviews all articles on scientific rigour, not on impact/relevance. Impact can be assessed after publication using article level metrics, other approaches that become possible on-line. This year more than 10.000 articles will be published in PLoS ONE. Other publishers are adopting this approach. This approach speeds up research communication (just one submission is needed), and it removes delays and costs, speeds up transition to OA, and thus speeds up research.

Recommendations:

- Key recommendation is to focus on article level for the assessment of published research outputs.
- Journal impact factors should have no place in research assessment
- Applications from researchers for funding, jobs etc, should include evidence on impact of their articles, and taken into account.
- Researchers should be asked how open their articles and underlying data are. EU funders could ask these questions , as part of the dissemination plans included in applications.
- Relevant initiatives should be supported (ORCID, PIRUS2)
- Policies need to be developed that require and facilitate OA publishing, as it is central to all of this. Such as ensuring funds for OA publication fees.
- Pragmatic steps can be taken now towards article level research assessment, which would improve research communication and the research process itself.

Statement Cameron Neylon (Open Knowledge Foundation)

Considerations:

- Our current system has historic roots that measures what could be measured, not what we wanted to measure. Today we can track and measure much more, including contextual factors. This brings opportunity, but we also need to work out what the right questions are.

- There are many reasons for research(er) evaluation, naming three.
  - Researchers evaluating themselves against their peers. Subjective and discipline specific, it is interesting but not a matter of policy.
  - Making objective decisions on distribution of research resources. A challenge is that formulaic approaches can be gamed, and that by nature metrics are backward looking. We need leading indicators.
  - Using evidence to design optimal architecture for the research enterprise. Evidence based policy making is lacking. There is opportunity to build strong, transparant, useful evidence base, but only if we include and understand the social context of that evidence (e.g. How does measurement change researchers' behaviour? Gaming of measurements? Differences in nations, disciplines, age groups?)

- Open approaches will lead to faster, efficient and effective research. Concerns around business models, quality assurance and sustainability exist. We can collect evidence, debate what important measure are, and base decisions on real data and understanding.

Recommendations:

- We need data, and be able to share it.
  - Research organizations must be encouraged to keep records of downstream usage of their published artifacts.
  - The EC and national funders should articulate that the provision of usage data is a key service for publishers to provide. Where direct payment for publication provision is made, such data should be included, and be technically and legally reusable.
  - The EC and national funders should support work towards standardizing vocabularies and formats for this data as well critiquing it's quality and usefulness.

- We need to integrate our systems of recognition and attribution into the way the web works through identifying research objects and linking them together in standard ways.
  - The commission should support efforts to open up scholarly bibliography to the mechanics of the web through policy and technical actions. The recent Hargreaves report explicitly notes limitations on text mining and information retrieval as an area where the EU should act to modernize copyright law.
  - The commission should act to support efforts to develop and gain wide community support for unique identifiers for research outputs, and for researchers. Again these efforts are diverse and it will be community adoption which determines their usefulness but coordination and communication actions will be useful here. Where there is critical mass, such as may be the case for ORCID and DataCite, this crucial cultural infrastructure should merit direct support.
  - Similarly the commission should support actions to develop standardized expressions of links, through developing citation and linking standards for scholarly material. Again the work of DataCite, CoData, Dryad and other initiatives as well as technical standards development is crucial here.

- We must study the context in which data collection and indicator assessment develops. Social systems cannot be measured without perturbing them.
  - We have a generational opportunity to make our research infrastructure better through effective evaluation and evidence based policy making and architecture development. But it takes understanding ourselves and the system we work within.

- The commission should act to ensure that current nascent efforts work efficiently towards delivering the technical, cultural, and legal infrastructure that will support an informed debate through a combination of communication, coordination, and policy actions.

Statement: Suzan Hezlet (London Mathematical Society)

Considerations:

- Slow science can be good science. In mathematics old(er) texts are still often cited, having impact over longer periods of time. Current evaluation system is focussed on quick impact.

- As a matter of principle evaluation should be impartial, and therefore be disconnected from people involved in funding decisions.

- As a matter of principle there should be no barriers to entry for authors.

- Having no barrier to entry is very important. E-submission systems and subsidized subscriptions have helped. BRIC countries are now more part of the system than before. The system has internationalized, though still US dominated.

- Mathematics research evaluation has unique traits. Some 500 peer reviewed journals, which is almost the largest group. This leads to articles being offered to multiple journals before publishing.

- Review is meant as validation of the work, and to establish relevance. But communication between the referee and author is also a process which improves a paper, for which we have evidence. This is going against an evaluation system that focusses on quick impact.

- Any evaluation system will have a large degree of error.

Recommendations:

- The policies on refereeing of articles need to be opened up and transparent.

- Attention is needed for the role of the communication between referee and author in improving the work.

Discussion on research evaluation systems

- Focus of remarks was on the research data and process achieved. The big push for OA is to do with speed and how material gets out there for criticism and review. This is clear communications between peers. However there's also the evaluation if it can be accepted for practical purposes or not. Present system tries to do both, although social and economic impact often discussed in other journals, not the peer reviewed. It's important for knowledge transfer to society and economy. Can we count this application literature towards citing etc.? Ultimately it's where you want to see impact, but it is outside classic evaluation system.

- There is a normalization problem around establishing impact. The normalization has to do with subject area, its size, and the number of journals in it, and then is used to say something about individuals.

- To establish impact there are many more things we can use as data. Look at it as a research problem. Collect the data, see how they apply and with what error margins. Take into account wider uses downstream.

- Get beyond the notion it is about articles. That's only a small part of scientific output.

- In social sciences/humanities monographs are more common than articles, and we haven't mentioned them yet at all. They could be measured as are articles, but you may not want to go there. Downstream measurements look interesting. In humanities as in mathematics the time scale in which works are used is different (more long-term) than others.

- Metadata of usage: should we address how to access that metadata? Not just open access to the article, but also to metadata (which can be social, contextual) If we want to use this in a scientific way we need access to it.

- Real breakthroughs have been often disbelieved, suppressed and not cited. But we want to have breakthroughs. Can we measure how controversial a certain piece is with the majority?

- Focus on awareness raising among researchers, involve them. Often they only think about the impact factor, but can we make them aware of our access concerns?

- Is there experience in establishing the accuracy and validity of impact factors and citations? There have been efforts to reproduce impact factors, but most relevant data, citation data is not available, which is a big problem. Also the way citations are used is not taken into account, all citations regarded as equal. Any transparent evaluation system needs to give access to this data.

- Many attempts at post-publishing peer review have failed. PloS ONE seems to work, how to clone and transfer that to other areas?
  - (Patterson:) We don't think PloS ONE is doing brilliantly yet, as far as post-publication commenting is concerned. Thinking broadly about what happens post-publication and what you can aggregate around that content. Other interesting things are happening like Mendeley. Experiments: PLoS hubs, create journals after article publication, aggregation of content around certain topics, doing human expert filtering for a specific community after publication. What are the incentives for those people to do this? Potential benefit is decoupling some things that are now all wrapped up in the one process of publishing a journal. To reduce cost and delay, get the content out there, and do the other things, such as impact assessment and content organization afterwards.

2. Preservation of scientific outputs

__Statement David Giaretta (Alliance for Permanent Access)__

Considerations:

- The general context in which we should view preservation is that "European citizens should be able to sensibly use data from anywhere in order to improve their health, wealth and happiness." This is a global issue but Europe can play a leading position.

- Challenges of how to get scientific data deposited:
  - For data producers: surveys show people are unwilling to share data. Concerns around possible misuse, legal issues, lack of perceived benefits to justify the effort. How do we overcome these? Some experiments with funders imposing requirements for deposit and the creation of data management plans. How to build confidence in this, and where to deposit the data?
  - For funders: Costs models are currently too crude. How can costs be controlled/shared/reduced, and justified? It is suggested to initially buy time only (3-5 years after project?) before long term commitment. How can we make sensible selections?

- Challenges for justification of the costs involved: How can we benefit from preservation? Preservation is for future generations, but in difficult times that perspective tends to be forgotten. Preservation is needed because hardware, software and general concepts over time become unfamiliar. The same is true for what prevents use of data from one area in another: unfamiliarity. So actions taken for preservation can help with wider usage right now, and open up possibility for greater benefits. Here data is not just evidence with which to redo previous processes, but a resource for doing new things now (new processes, combination of data) in new ways.

- Cautionary note: the techniques of "emulate or migrate" which are adequate for libraries and their documents are not adequate for data

- Cautionary note: the term "metadata" is too coarse grained and causes misunderstanding and confusion. We need the finer granularity of terminology such as that provided by the Open Archival Information System Reference Model.

Recommendations:

- EU can be catalyst to co-ordinate national and commercial funding. Organisations such as the APA can play a role in bringing these efforts together across nations, beyond Europe, in science, academia and commerce.

- Support the creation of realistic data management plans by encouraging/mandating the certification of repositories so people know where they can deposit their data in confidence – Europe has a lead here in Framework for certification of repositories

- There is a need to harmonise licencing, copyright and other legal issues etc, and the promotion of data as a first class publication

- Promote the training of researchers (and other data producers) with a coherent view of preservation and re-use applicable to all types of digitally encoded information

- At the technical level we need a Support infrastructure to sit on top on the vital GEANT and EGI to support preservation by sharing the costs and effort and enabling usability – not just file access
  - Identifier system which is really persistent –the current ons have issues with sustainability - the EU can throw its weight and long term financial backing, perhaps via an endowment, behind one system or another
  - Provide active brokerage so repositories which close can find another to hand on in a chain of preservation
  - Share the effort in maintaining understandability and usability into the future using the intellectual "cognitive surplus" – and support usability right now.

Statement Peter Burnhill (University of Edinburgh)

Considerations:

- Two key concepts for preservation of scientific data are making data public and gaining recognition.

- Two key challenges therefore are how to make data available into the future and for the future, and how to provide the reward of recognition

- Data need not be digital, all that is digital is not always data, but might become so

- New and existing organizations are stepping forward to take stewardship of e-journal and other digital content (LOCKSS, CLOCKSS, Portico, national libraries such as BL (UK) and KB (NL))

- peprs.org is a project to act as monitor of who is looking after what e-journal, how, and with what terms of access. Help is welcome on how to govern this.

- Research literature is an international concern, and requires international action. Relying on legal deposit legislation is not enough. The EU and Commission have an important part in ensuring that Europe has a lead role here.

- Data archiving needs a demand-focus, assisting researchers to discover and obtain access to data produced by others.

- University of Edinburgh has institutional policy to guide staff in management of digital research data. Elements, setting standards for themselves, of that are: a) Research data of future interest will be offered to repositories, or stored in own repository, open

for re-use, b) no exclusive re-use rights are handed over to commercial publishers, c) keeping track of which repositories have the data.

Recommendations:

- To preserve the record of science for the future, we should include research literature (next to data) as an important part of the record of science. Both have evidential value for research, and the relationship between the two is also important. (separate note on that relationship submitted in writing) (taking into account three types of data: a) source and reference databases, b) datasets upon which conclusions in published literature are based, c) supplementary data files that accompany e-publication. Responsibilities for different types differ)

- Future-proofing requires we make data available as though for researchers beyond our immediate peer group and for the machine-as-user. To ensure that future researchers can use data for their own research.

- This means opening up knowledge now locked in document formats like PDF, so that scientific literature becomes scientific data

- When the Commission revisits the grand societal challenges to which research can and should address it should regard 'assured and continuing access to digital content' itself as a grand societal challenge. One in which Europe can make a significant and lead contribution.

- We should not have a narrow view of science and scientific data

Statement Robert Campbell (Wiley-Blackwell)

Considerations:

- The traditional role of journals i registration, certification, dissemination and preservation

- As a publisher we try to ensure preservation and making investments towards it

- We are taking part in Portico and CLOCKSS (which has a dozen hosters, not more, to avoid confusion)

- If for some reason Wiley-Blackwell should stop providing material on-line, Portico is triggered to take over
- Publishers are ensuring long term preservation of articles, but this is only a part of the scientific output. Original data may be lost, and only the article left, and that is only the rhetorical presentation of arguments to convince others

- Digital research libraries are hindered by costs, format obsolescence and scanning technology limitations etc.

- OA tries to avoid the added value of publishers. Given this attitude how can we expect publishers to invest?

Statement Rachel Bruce (Joint Information Systems Committee)

Considerations:

- Research output is also software, not just research texts and research data. Software was not mentioned yet.

- We are only starting to understand how to describe software as an artifact.

- Preservation is about different areas: content, policy, infrastructure, sustainable business models

Recommendations:

- There needs to be greater harmonization of IP directives across EU MS. In terms of research texts people say preservation is dealt with, but not so for these policy issues.

- OA is not just about economics but also a policy for preservation. OA can also mean open formats for research texts.

- Move forward on data management plans. Also to learn, as we don't know yet how these might be implemented. A step forward would be to include in data management plans where data should be placed, what data would be produced etc.

- Acknowledge software as artefacts and important output of research. This is urgent to research. There is a need for legacy systems to run longer, skills of IT professionals to be maintained to be able to understand the software record of science.
- Issues of sustainability of infrastructure need to be addressed. More collaboration across MS and governance models. This is where the EU can support.

Discussion on preservation of scientific outputs

- Publishers don't claim copyright to data

- Who's going to pay for this? EC has put some money up for data repository projects. OpenAIRE is dealing with publications, harvesting them and linking them to FP7 project data, and adding usage statistics for assessment. So the EC as funder can know about the output of research they fund. Preservation is not within the scope of the project. OpenAir relies on a network of independent repositories, and is aligning their work with them, and CERN powers a repository for researchers without access to one in their own institution, assuring long term preservations. EC should set policies on re-use statistics and have a repository site.

- Who's role and responsibility is it for doing and funding this? This is not clarified yet in the community. Researchers don't know the difference between preservation and access, so there is a serious debate to be had about roles and responsibilities. Some universities may well want to do this, but not everyone. We need a roadmap from the present confused position to a place with more clarity. EC has a role in stimulating clarification in the FP7 and FP8 frameworks.

- There is data clearly not worth preserving, and data that clearly is. The largest volume is in the middle where judgments are to be made, and that's the problematic bit. What seems a lot of information now, is only a tiny bit of the future volume and storage. Volume is not a key issue. The really difficult thing is keeping the data understandable and usable.

- CERN now holds 30 PB of data from the LHC accelerator. Keeping it is not the issue. The 'tragedy' is in making it reusable and understandable, even for the researchers that created the data themselves. We need to act now to make it accessible now. Preserving data is not 'sexy' enough, and there is systemic lack of political understanding of the importance. There's a need for a clear political statement that preservation is hugely

important as well as investing in opening it up.

- The problem of preservation for a singular researcher is one of time, more than money. Which incentives are there for a researcher to put that time in? Such as the time needed for reformatting data, or changing my collection process. So we need to turn the attention of researchers to it. Don't mandate the publication of data as prerequisite for funding (you will just get singular spreadsheets attached), but measure the downstream re-use of data as grounds for them to get future funding.

- Legislation is needed to prevent certain data from being deleted. E.g. Clinical trial data in the UK is deleted after 15 years, but there are cases where after 20 years medical research on pharmaceuticals is called into question, and then you cannot go back to the underlying clinical trial data.

- Research in humanities often makes use of other people's data, reanalysis of existing data has always been key. Other sciences, such as climate science, are doing this as well. Can we leverage long living institutions for this? Maybe have something like an insurance fee for this, or build up an endowment? This creates means independent of current budgets/revenues and policies. The idea is to separate long term costs of current needs to raise revenu. This would need political action, and may be hard to reconcile with the EC's normal approach to funding.

- OA cuts out the costs of restricting access

- What is the scale of the problem?

3. Access to scientific information (including Open Access)

Statement Laurent Sesink (Netherlands Organisation for Scientific Research NWO)

Considerations:

- NWO has included OA in granting conditions, has in incentive fund for OA publications, does awareness raising actions around OA

- NWO is involved in OApen project, bringing OA to monographs

- OA for data is another issue, and NWO is doing the following:
  - Setting up guidelines for good OA data provision
  - OA where possible, and restricted access to data where necessary
  - As co-financer NWO stays co-owner of the research data is financed
  - Giving incentives to publishers to go OA, calling for new OA journals in certain domains with NWO funding

- Aiming to increase impact factor for OA journals, a.o by prompting researchers to use OA journals

Recommendations:

- An EU level challenge is getting to common operational procedures. There is a need for common EU goals on common access to articles, monographs and data.

- Cost transparency of publishers services is needed

- Datamanagement and perhaps a one time fee for data deposists need to be budgeted in grants

- We need to reassure researchers, to get them to share data, that data will maintained in trusted repositories that use certain guidelines

- Discussion needed on which data these repositories should cover

- We need to stimulate scientific credits for good quality data sets being published, and community review of data sets.

Statement Michael Mabe (International Association of Scientific, Technical & Medical Publishers)

Considerations:

- STM publishers are committed to the wide dissemination of, and unrestricted access to, their content

- We support any and all sustainable access models that ensure the integrity and permanence of the scholarly record

- We do not support unfunded mandates that constrain scholarly authors or affect the sustainability of the publishing enterprise services that publishers provide must be paid for in some way

- While 93% of university researchers surveyed are satisfied with journal access, lowest satisfaction is with access to data sets.
- Access to journals is sub-optimal for SME's and 'other', read to mean the general public.

- STM is working with other stakeholders to understand and bridge those gaps

- STM initiatives to expand access include development of new business models (gold open access, author pays models, DeepDyve, pay per view models, and pilots aimed at access for SME's) and information philanthropy (e.g. patientinform and Research4Life)

- Higher VAT rates on electronic products over print versions are acting as barriers to a fully digital environment.

Recommendations:

- We believe that a level playing field with light touch regulation is the quickest and best way to realise the goals of widening access we ALL share, and recommend this course to the Commission.

- An important action that needs to be taken in Europe is to resolve the significant problems that exist with data.

- STM's position is that the lowest possible (or zero) VAT rate should apply to all kinds of academic and professional publications (journals and books) irrespective of the format in which they are produced, commercialized or delivered , and we encourage the Commission to take this initiative up.

Statement Saskia de Vries (Amsterdam University Press)

Considerations:

- There is a clear need for Open Access publishing and funding models for monographs to bridge the gap with OA articles, but also because the traditional business model for books is losing its sustainability.

- The OAPEN project found a financial model which is appropriate to scholarly humanities and social sciences monographs, it created a publishing platform which is beneficial to all users and finally, OAPEN created a network of publishing partners across Europe and the rest of the world.

- Funders of research should take care of the costs for the Open Access edition as the dissemination of research should be seen as part of the research, but also because they *are* already taking care of most of the costs of scientific publications.

- The structure of costs of traditional publishing and the OAPEN model can be found in the slides

Recommendations:

- Special attention is recommended for Open Access for scholarly monographs, in particular in relation to the Humanities and Social Sciences.

- Academics should be made more aware of how much money is spent by libraries for them to obtain access to the results of research.

- The Guiding principles for scholarly communication should be:
  - Research and the dissemination of results belong together
  - Funders' OA policies should include OA publishing
  - Publishers should develop OA Publishing as a service to the scholarly community

- The 'authors pay system' used in STM for the publication of journals and articles should also be introduced in the HSS, where the monograph is the most common way of disseminating the results of the research.

- Funders of research should take care of the costs for the Open Access edition as the *dissemination of research* should be seen as part of the research, but also because they are already taking care of most of the costs of scientific publications through the library budgets.

Discussion

- With monographs the reimbursement of costs (and scientific journals) should extend beyond a project as they cost more time to make. However with up front funding that is not an issue and revenues are not a factor.

- In Germany an OA pilot is supported by all states and universities through research grants. There also rights for secondary publications are asked.

- Taking the golden OA path is a concern for unfunded researchers. If we go that route we need to think deeply about its impact on unfashionable and unfunded research. That needs to be investigated before it becomes the regular route.

- If we are to bring the costs of publication down, money needs to be in the hands of researchers.

- In mathematics they have looked at the percentage of unfunded researchers. Emergent nations have this problem a lot, and it impacts the quality of the field.

- VAT is barrier, as it makes e-versions more expensive than paper versions. MS say its

internal market problem and to talk to the EU, EU says it is MS problem and go talk at national level.

- The peer review system is very biased through language and the cultural background of the reviewer. We need to prevent the payment model from becoming another bias, raising thresholds to entry for developing countries

- Libraries can shift their money from regular publishers to other forms of publishing, so they are important as stakeholder.

- When talking about OA we should also talk about re-use

- Many OA publishers waive the fees for those who cannot pay, trying to keep trhesholds low

- If we invoke the developing world it is not just a question of how they get access to the world research literature, but also a question of how they can contribute (which could be releasing their work in OA)

Statement Peter Gøtzsche (Nordic Cochrane Centre)

Considerations:

- Medical research results are being published selectively. Furthermore, harms of drugs, even potentially lethal harms, are often omitted or downplayed in published reports. Selective reporting and biased data analyses can have disastrous consequences.

- Clinical study reports submitted to drug agencies for marketing approval are much more detailed and truthful than published reports. It is therefore of vital importance for EU citizens to have easy access to such reports

- The EMA consistently denied access, arguing it needed to protect the drug industry's commercial interests. The European ombudsman intervened, setting an important precedent. However the same problem now pops up in the UK.

- Calls for data sharing have been put forward by many leading institutions, including the OECD, the WHO, the US National Institutes of Heath, the US Congress, the European Commission, the European Ombudsman, journal editors, The Cochrane Collaboration, and several funders, for example the UK Medical Research Council, the Wellcome Trust, the Bill and Melinda Gates Foundation and the Hewlett Foundation.

- If commercial success depends on withholding data that are important for rational decision making by doctors and patients, then there is something fundamentally wrong with our priorities in healthcare.

Recommendations:

- We must ensure that all EU drug agencies adopt the same new openness policy that the EMA introduced in 2010.

- We must also ensure easy access to ALL drug trials, also those performed after the drugs have been marketed, and also those performed by academics, whether or not they are published.

- We must also ensure easy access to trials that are not about drugs.

- We must get access to the raw data. Data sharing would lead to tremendous benefits

for our societies, as a lot of research could be done on existing data, making it unnecessary to collect new data. Furthermore, the incentive for bias and outright fraud would be reduced when other researchers can check the data.

- If legislation is needed to make this happen, then let us have this legislation. Our patients deserve nothing less. Without their altruistic willingness to run personal risks, we could not carry out clinical trials.

- Data sharing is a moral imperative considering that the lack of data sharing is harmful to patients.

Statement Jean-Pierre de Cuyper (Royal Observatory of Belgium)

Considerations:

- Digitalization of historic photographic materials bring with it a whole range of easily overlooked difficulties. Digitalization can easily result in loss of information that way. See slides for a more detailed listing and examples of technical aspects that can lead to information loss. Commercial scanners are unfit for this purpose, custom made scanners are deployed.

- When done right digitalization results in huge databases and metadata sets.

- Digitalization may not be the best way for preservation for these materials, but it is crucial for access, analysis and re-use. The raw digitized data needs to be archived as well.

- Digitalization of photographic materials create data that is in the original photo inaccessible to the human eye. Special software is needed to visualize the huge image files and to make the greyscale information accessible in full detail.

Recommendations:

- Making all the scans available as OA is crucial. This and the scale of the databases involved require additional funding

- The way various MS tackle this problem is not always obvious, and clarification is needed

Statement David Carr (The Wellcome Trust)

Considerations:

- As a funder key activities towards open access are
  - Policy: all papers funded must be placed in PMC or UKPMC archives within 6 months of publication.
  - Funding: cost of open access publishing is legitimate research expense, dedicated funds provided to institutions
  - Infrastructure: established the UK PubMed Central repository (currently with 18 funder partners)
  - Advocacy (working with publisher towards compliance, raise researcher compliance, raise OA awareness)

Recommendations:

- Use golden OA

- Establish a clear policy on open access for EU-funded research, building on ERC policy and FP7 pilot.

- Provide dedicated funding to meet open access publication costs

- Work with others to engage:
  - researchers in receipt of EU-funding – to advocate benefits of open access and track compliance
  - European research institutions – to establish policies and processes to support researchers
  - publishers – to enhance workflows and accelerate transition to author-pays model

- Provide dedicated EU funding to develop and sustain key infrastructures needed to ensure access to research literature and data
  - vision for Europe PMC, linked into the ELIXIR initiative

Discussion

- Re-use rights are something to consider. Any OA policy must both incorporate access and re-use. Particularly in biomedical area. A survey (SOAP FP7 project) on the status of OA journals and their conditions showed bigger (STM) publishers tend to add re-use licenses (such as Creative Commons), covering up to ¾ of articles, but most other publishers don't take this into account yet. Most OA journals do not use CC. To fully achieve the OA benefits of re-use of scientific information standardisation is desirable.

- It is desirable to have more transparency in business models, so researchers can see how costs and funds are calculated, to build trust they get a fair deal.

- We need to look at blogs, social media and new kinds of writing and incorporate them in our evaluation models

- There are a lot of practical issues around access. Librarians want transparency in costing structures. It takes a constructive dialogue between the various stakeholders.

- The Dutch NWO forces transparency of cost structures in their rules.

- It is important that the law is enabling rather than hindering, so that it permits the granting of use and data mining, without infringing database law or copyright. It needs to be easy to see those rights have been granted, not just in academia but also in other sectors. Data mining will become a bigger activity driving innovation.

- SMEs have the least good access, as a dividing line in licensing is the 'non-commercial' clause, which excludes SMEs entirely.

- We talked a lot about preservation, but key is dissemination of scientific information and the speed and timeliness of dissemination. So we need to reduce the cost of access to scientific information. Can we do something similar as in the telecommunications industry where the EU drives the harmonization of telephone and mobile data fees?

- Re-use rights are a barrier to transition when moving from subscription to OA models. 70% of income for academic publishing comes from re-use. We need to make sure they can continue to make a contribution.

- Most calls for access center on publicly funded research. When are we also going to look at industries? E.g. drug companies keep data under lock that was created by

patients. That should be made illegal.

Statement Alicia Wise (Elsevier)

Considerations:

- Elsevier's vision revolves around universal access, quality and sustainability

- Elsevier sees 6 drivers for access initiatives:
  - digital preservation (all titles archived in CLOCKSS, Dutch Royal Library and Portico, pilots on institutional preservation)
  - easy flow of scientific content (content mining, applications, discovery, integration of data+content+services, semantic linking)
  - journal pricing concerns (currently 2 author pays journals, 800 with sponsorship option, 30 with delayed access model, green manuscript posting policies)
  - public access to information (DeepDyve pilot with 50 Elsevier journals, now expanding to 250 titles)
  - showcasing scientific output in a specific context (bringing content together in a single context, such as all Hungarian researchers' publications)
  - understanding research strengths and trends (better data for decision making in academia)

Recommendations:

- Looking to work with the EU on shared standards and interoperability, to support easy flow of scientific content

- If transition in scholarly communications to OA is wanted then Elsevier is willing to work with EU on this

- EC could sponsor free access to information in their funding decisions

- A 50% OA article goal in FP8 may be too far form practical reality, given that the 20% FP7 pilot goal is only now starting to be reached because of delay between funding research and publication of results.

- Adopt a test-and-learn approach so policies and practices evolve in evidence based ways

- We'd like to work closely with other stakeholders as collaboration is essential for appealing workable solutions.

Statement Paul Ayris (Association of European Research Libraries LIBER)

Considerations:

- It is not a question of if Open Access is an option – it is rather a questions of how Open Access solutions can make a difference

- The world will not become Open Access overnight. There are steps which can be taken, however, to change the current way the world works – for education and research to enjoy the results which more openness brings.

- LERU, the League of European Research Universities is about to launch its Open Access Roadmap in Brussels in mid June. This is a Roadmap that ALL Universities can follow if they wish to adopt Open Access practices in both the Green and Gold routes.

Recommendations:

- The EU should make changes to the current EU copyright framework.

- There should be a fair dealing exemption for text and data mining for non-commercial use. This would strengthen research across the EU

- The EU should identify ways for extended collective licensing to take place for out of commerce and orphan works, so that these materials can be digitised for the benefit of research and education.

- The EU should ensure that contracts and licences cannot overturn Fair Dealing exemptions, as currently happens.

- Fair dealing exemptions should also cover format shifting to allow libraries and memory institutions to digitally preserve for the long term the digital content that European researchers use and need.

- Were this progress in the EU's copyright/IPR frameworks to be made, Member States could follow suit. This is exactly the conclusion reached by the UK's Hargreaves review of copyright/IPR frameworks. Greater openness would result if these changes were made.

- The LERU Roadmap identifies a number of opportunities for studies and infrastructure development to support European researchers. It is important that these initiatives and EU funding programmes are aligned, and so ongoing dialogue between relevant parties is necessary.

- There is a wide variety of Open Access platforms and Discovery To Delivery tools in Europe. How do, for example, Europeana, the OpenAire pilot project for the deposit of Open Access materials, Europeana Libraries (a LIBER EU project), the Bielefeld aggregation of Open Access repositories, the Directory of Open Access Journals (and the content they contain) all fit together into the European Information landscape? There needs to be dialogue between all these stakeholders and the EU can help with project funding to help identify how the Open Access infrastructure in Europe can be developed and supported.

- European research needs sustainable digital preservation infrastructures, to complement current digital preservation research. The EU should work with relevant stakeholders across Europe to identify the needs and the funding to meet those needs.

Statement Astrid van Wesenbeeck (Scholarly Publishing & Academic Resources Coalition Europe)

Considerations:

- SPARC support OA to scholarly literature (articles and books) and data

- SPARC Europe upholds that Open Access via Institutional Repositories and Open Access via publishing are complementary strategies. We welcomed the FP7 Open Access pilot which secures permanent access via repositories and supports Open Access publishing by making publication costs eligible for reimbursement. The EC Open Access policies can set an example for governments and funding agencies on a national level.

- SPARC Europe upholds that Open Access in Institutional Open Access Policies should refer to freedom of access + freedom of re-use, in other words "Libre" Access.

- SPARC Europe upholds that Open Access means immediate access, thus access without any delaying mechanisms.

- The current research evaluation system that is aimed mainly on indirect measures based on average journal performance (e.g. the journal impact factor) is disadvantageous to Open Access. Besides the fact that Open Access journals face difficulties in receiving an impact factor we are convinced that digital research environments offer much more quality and performance indicators than citations alone.

- To be allowed to reuse publications and data ensures efficient and innovative science and prevents the recollection of already available data. A lot of work in Open Research Data, Data Management Plans and the development of international standards for access to data and the use of applications is currently conducted at different levels.

Recommendations:

- Further implementation of the EC/ERC Open Access policies conducted in the FP7 Open Access Pilot with an outreach to member state level
  - Extend the Open Access policies to all research areas covered and funded under the forthcoming FP8 and build upon the network of open repositories that has been established through the OpenAIRE project.
  - Ensure researchers compliance with the Open Access policies by focusing on means of communication with researchers and increasing awareness amongst researchers about the benefits of Institutional Repositories and the equal quality of peer-reviewed Open Access publications to toll-access publications;
  - Strive for the shortest embargo periods possible where Open Access is realized via open repositories. We understand that subscription publishers need the embargo periods for sustainability issues but embargo periods in the Open Access Pilot lead to publications not being accessible for the first 6 or 12 months, depending on the research area. We emphasize that Open Access Publishing offers a solution for the embargo dilemma.
  - Support projects and initiatives that aim to provide an infrastructure enabling publisher content to be harvested and deposited into institutional repositories or subject based repositories.
  - Support and encourage projects that aim to develop, implement and monitor Open Access policies by national governments and funding agencies.

- Libre Open Access refers to Open Access Publishing: Boost Open Access publishing
  - Make funding (EC and national level funders) available to cover the costs of Open Access publishing by including the dissemination of research as part of the grant requirements and by making publication costs eligible for reimbursement

  - Recommend Open Access publishing by pursuing efficient and focused communication with grant recipients; communicate that the quality of Open Access peer- reviewed journals is equal to the quality of subscription peer-reviewed journals.

  - Explore in a collaborative approach with the research community how recognition of peer-reviewed Open Access journals can be improved in research evaluation.

  - SPARC Europe calls for subscription–based publishers to start the transition of subscription journals towards Open Access.

  - And to allow authors and institutions to deposit metadata and publications into Open Access repositories and support Creative Commons licensing of these materials.

- Support a reformation of the Research Evaluation System
  - We call for EU funding streams to foster a shift from a research evaluation system that is based on journal level to an evaluation that is based on article level, and to explore together with the research communities which indicators that are emerging from the digital environment could be added to this system.
  - Including Open Access availability of the publication in the criteria for research evaluation.
  - Supporting projects that develop alternative bibliometric tools and services or aim to implement them.

- Start a pilot on Open Access to Data as part of further efforts to build up and extend European Infrastructures. We recommend the EC to encourage and fund the openness of research data in order to take the next step towards concrete implementation by starting a pilot on Open Research Data.

Statement Thomas Parisot (Cairn)

Considerations:

- Cairn has been publishing 130.000 articles since 2005, from 80 publishing structures of various nature.

- Publishers choose between two models (all articles on Cairn available for free without embargo, or using a moving wall where more recent articles are pay per view or licensed)

- Only 4% of journals available at Cairn use the completely free access model.

- Embargo periods are a concern. The average moving wall at Cairn is at 3.7 years, in conflict with a 6-12 months embargo which seems far too short.

- Interest 'half-life' in articles differs greatly between various disciplines

- Publishing structures in the humanities and in minority academic languages as French prove to be fragile. This make the 'green' OA route very difficult.

- Going the 'golden' OA route has been estimated to cost 44.8M Euro for the humanities and social sciences in French journals. The current budget in French universities for subscriptions is 5.3M

Recommendations:

- Take into account differences regarding interest 'half-life' and define embargo periods discipline by discipline.

- Take into account the specific aspects of disciplines like the humanities and social sciences as well as minority academic languages

- If the golden OA path is chosen, it must be conceived as a one-way street without return, and with financial conditions of public funding fulfilled. So that all actors know what they are up to.

- For the humanities and social sciences, who have a wider public reach in French general interest journals, these other publishers need to be taken into account as well, as it may impact them negatively, or as in the case of publishing essays, threaten them.

Discussion

- LIBER's mention of 'fair dealing' is not about adopting the US legal notion of 'fair use' but seeks to build on the Hargreaves report which suggest fair dealing for non-commercial use, as it could be an enabler.

Statement Elena Giglia (University of Turin)

Recommendations:

- The author's right to self-archive the post-print in their institutional repository should always be retained without embargo (this helps meet public access, long term preservation needs, and evaluation purposes.
- Concerning copyright, scientific publications („give away literature") must have their own rules and regulations, different from revenue generating and fiction works. (scholarship is a great conversation, and exists only if output is shared and circulated freely, with only attribution as condition)

- Open Access should be recognized as a national and European strategy at political and legislative level. (OA as main strategy in ERA, OA Pilot in FP8 enlarged, Clause 39 to be mandated, promote synergie between stakeholders, need for a normative umbrella, grants to build evidence of OA benefits, such as cost/benefits ratios and ROI, and innovation of evaluation systems)

- For economic sustainability, the transition from subscription based models to an article processing charges scenario should be supported in institutions and library systems (library budget cuts and research funding impact both subscription publications and golden path OA negatively, EU as policy making, funding and infrastructure making body could help find new ways)

Statement Salvatore Mele (CERN)

Considerations:
- OA works, as CERN and its discipline high-energy physics has experienced since 50 years ago

- 97% of articles in journals in their field are online as OA months before journal publication as pre-prints, submitted to subject-specific OA repositories by the authors.

- This is universal green OA. Without mandates, but using opportunity.

- OA preprints collect five times more citations than journal only articles. OA preprints get cited from 6 months before publication.

- This shaves months off the research process, and is a real opportunity to accelerate progress.

- Scientists do not communicate through journals, they do that through OA infrastructure build by the community for the community.

- Other fields, not just high energy physics are also seeing this happening (astronomy, astrophysics, economy, condensed-matter physics, social sciences, and now in mathematics which is fastest growing section of arXiv)

- OA accelerates science and that's established fact.

- The SOAP FP7 project survey of scholars around the world and across disciplines shows 90% of scientists think OA is beneficial, yet only 8-10% of articles is published as OA. This gap stems from 2 barriers: funding and quality.

Recommendations:

- Embargo is not a solution, but makes things worse. It creates an unfair gap between those who can read the article, and those who have to wait and cannot timely contribute to progress.

- One size does not fit all, some disciplines where patents are at stake cannot or reluctant to share preliminary results on repositories.

- OA journals are the only opportunity to accelerate science and innovation

- We need to tear down the funding barriers and quality barriers to OA

- We need to convert high-quality journals to OA. (This is what the SCOAP3 initiative is doing for the field of high-energy physics, which is bringing together stakeholders and redirecting subscription funds to convert journals to OA model)

- Access to scientific information is a global issue which needs brave, creative and global solutions built on consensus We hope the European Commission will be brave, forward looking, and committed to accelerate science. For its part, CERN is committed to share its Open Access experience and support the European Commission and all stakeholders in this process.

Statement Caroline Sutton (Open Access Scholarly Publishers Association)

Considerations:

- The publishing landscape is now different from 2007. Then OA publishers could be counted on one hand, now there are many. This is an opportunity to experiment with the potential that OA can afford.
- OA publishers are useful partners for future projects for Open Science.

- A lot of potential is still latent and theoretical, we can do and experiment more.

- Putting our articles in repositories is unproblematic as they are already Creative Commons licensed.

- EC funding of OA reimbursement does encourage OA publishing, as it makes it easier for researchers to choose OA as an outlet. It recognizes publishing as part of the research process.

Recommendations:

- Can we link OA to other areas of Open Science, OASPA is willing to experiment.

- We can look at improving our connections to repositories

- OA publishers should be in dialogue with those involved in scientific data, to see what can happen if OA publishing and open data get connected.

- The EC can play a role in coordinating and facilitating the discussion between stakeholders (such as those of the morning sessions)

- The EC is a funder of OA reimbursement. We support expanding this to all areas of research in FP8 and extension beyond STM fields.

- We urge the EC to define OA as access and re-use of all research material, and to please use a Creative Commons Attribution license (CC BY).

- It is needed to pay attention also to the skills, knowledge and expertise needed to make the new systems work.

Final Discussion

- An issue to flag is that we did not mention research materials yet. We talked a lot about journals, little about data, and hardly about software. A standard materials transfer agreement is to be recommended. Also to expand the notion of research ethics of making components of the research process more available.

- Linked data and semantic web is a popular topic outside this discussion, which is about the notion of publishing for the machine. We currently hide the complexity of data structures for human usability. If we go to machine readable, relations in data need to be part of it. Then publication of data schema and code books is important. This gives data scientists the option of getting credit for publishing those, and thus get rewarded in terms of reputation.

- As a publisher I was even surprised we talked so little about data, and even less about software. How are we going to make progress towards taking on the data challenge. Publishers made progress on texts, but other areas are dramatically uncovered.

- There are great differences in accessibility of libraries, seen from the viewpoint of the consumer. It should be always possible to access scientific information that relates to yourself. (e.g. information on an illness you have)

- We make a distinction between OA in articles and data, but even if not so much visible today, in the Netherlands data is very much on the agenda. Several projects focus on data and tools for exploring data.

- Future hearings should focus on access to raw data, as it would be very useful. Results can be easily manipulated, not raw data.

- We did not address financing models for data, but we did address data itself, and how to select what to preserve. Active appraisal and disposal is something to look into more deeply. Most of an archiver's job is to make important decisions on what to throw away.

- It is not trivial to see what data is useful for the future. Not all data is digital, making it digital brings its own issues with it.

Concluding Remarks by Gilles Laroche (DG RTD)

Thank you for being here, and for all of your input. Minutes will be written and circulated for your feedback.

The next step will be a public on-line consultation by means of a survey. This is also a key input towards the planned Communication on scientific information. You are all invited to participate.