AN OPEN ACCESS MANDATE IN ICELAND
Njörður Sigurjónsson

Recently Bifröst University became the first higher education institution in Iceland to adopt an Open Access mandate. The mandate, or policy depending on definitional preferences, was initiated by the faculty and is a declaration of the faculty member's preference to publish in Open Access journals and their obligation to store research articles in the university's open repository. The mandate, which is closely modelled on similar ones passed by for instance Harvard's Faculty of Arts and Sciences and by the Harvard Law School,[1] was then taken up by the Bifrost University Council, which gave it a status as a university wide policy. Below is a rough translation of the text with a brief explanation on its meaning and rational.

The Mandate
Bifrost University is committed to the objective of making the research output of its faculty available to as many as possible. For that purpose the academic staff of Bifröst University will seek to make their scientific articles available in open access, either by publishing in open access research journals or by depositing them in a research repository. Every member of the academic staff allows the university to make their published research articles available and to store them in an open repository, such as "Skemman". This holds for every research article published in a scientific journal authored by the researcher, alone or with others, during the time of his or her tenure at Bifrost University.

Exempt from this policy are books, teaching material, reports or other material that does not fall under the category of research articles published in scientific journals. Exempt are also research articles that are completed before the adoption of this policy and articles that were already underway and are bound by restrictions that are incompatible with this policy.

The Rector or the Rector's designate, will waive application of the policy for a particular article, or delay its appearance in the open repository, upon written request by a Faculty member explaining the need.

Each Faculty member will at no charge provide an electronic copy of the final version of the article no later than at its publication date, to the appropriate representative of the Rector's Office in an appropriate format (such as PDF) specified by the Rector's Office.

The Rector's Office may make the article available to the public in an open-access repository. The Rector will be responsible for interpreting this policy, resolving disputes concerning its interpretation and application, and recommending changes to the Faculty when appropriate. The policy will be reviewed after three years and a report presented to the Faculty.

Discussion
The next few months and years will be a trial period and no doubt there will be obstacles in implementing the policy. However, with the issue of Open Access becoming an ever more pressing issue for both academics and the general public, it is worthwhile to reflect on the process of introducing policy change in an institution. What is it that helps reaching consensus on a policy? Bifröst University is a fairly small institution, even by Icelandic standards; the process of reaching an agreement is perhaps not as long winding as in larger organizations. Nevertheless, it took some discussion to reach an agreement, and the focus that helped in the discussions at Biföst were: 1) keeping the message simple; 2) the use of exemplary institutions abroad as a reference; 3) the benefit of being early adopters in your area; 4) the idea of Open Access as a public good, and 5) emphasize the opt out available in exceptional cases.

Readers knowledgeable about Open Access discussions, policies and mandates, see from the start the resemblance with the Harvard policy mentioned

---

[1] http://www.earlham.edu/~peters/fos/2008/02/text-of-harvard-policy.html http://www.earlham.edu/~peters/fos/2009/03/mit-adopts-university-wide-oa-mandate.html

above. When discussing the mandate and gathering support this became important, since there were "respectable" research universities abroad that had already adopted a similar policy.

In a similar way as the Harvard model mandate, the first part of the Bifröst mandate describes the intent and what the faculty is committed to do. The purpose of the mandate is to make the scientific output at Bifröst accessible to everyone, everywhere, on the Internet. The way to do that is to either publish in open access journals or by depositing the articles in the university repository.

This point was mentioned during the debates about the rationale of the mandate, and the argument about more democratic and fairer distribution of knowledge was convincing. Other important lessons from the discussion process at this time was to keep the message simple and not let the discussion spin into a general debate about intellectual property, the scientific merit of particular journals or the general developments in publishing across the globe.

The next part of the mandate differs from the Harvard one. With the mention of Open Access journals in the Bifröst mandate, the emphasis is on the University's commitment to OA publishing. The University publishes its own open access journal Bifröst Journal of Social Science (bjss.bifrost.is) that uses international open access software (OJS) and is listed in the Directory of Open Access Journals (doaj.org).[2] This fact, and the relevance of the journal to the researchers at Bifröst, made the idea of publishing generally in OA journals more natural than perhaps in other places were the culture is different.

It came out in the discussion, that Bifröst had already adopted an Open Access culture and now it could be the first in Iceland to decide upon an OA mandate.

The reader will notice, that the wording at this point doesn't require the faculty to send their material directly to the repository, which might sound "weak". However, there is a clause that gives the University permission to store every article in the repository. Skemman is the one we use at Bifrost together with several other Icelandic institutions,[3] and makes the University at least partly responsible for gathering and making the material available. The later sentence: "Each Faculty member will at no charge provide an electronic copy of the final version of the article no later than at its publication date…" makes it however clear, that it is also the faculties' responsibility to make the material available. The University can demand that a given article is sent to the repository, and that the administration has the responsibility to gather the material.

The "exemptions" clause in the end was necessary to convince the skeptics, and is, with the rest of the mandate, similar to the Harvard one. Being able to opt out became an important point in the discussion, i.e. that in any unforeseeable, circumstances it is possible to be exempt from the policy by writing a request. How often this option will be used in the future we will have to see.

Njörður Sigurjónsson Ph.D., is an Assistant Professor Bifröst University, Borgarbyggð, Iceland and has taught Cultural Policy and Management at Bifröst University since 2004.

2
http://www.sciecom.org/ojs/index.php/sciecominfo/article/view/1765/1398

3
http://www.sciecom.org/ojs/index.php/sciecominfo/article/view/4762/4323

PATIENCE IS A VIRTUE? SOME REFLECTIONS ON MANDATE IMPLEMENTATION AND UPTAKE
Jonas Gilbert

What expectations concerning uptake and compliance are realistic when you are implementing an open access mandate?

There are several recent studies that can provide useful information and comparative examples, but I believe, that one also needs a practical apprehension of the local level to be able to plan what resources and strategies that are adequate.

You will have to interpret and present the progress to your stakeholders: are we doing good or not-yet-good-enough? When can we expect to reach a tipping point and to see that the policy uptake gains momentum?

Not least there is a need to prepare yourself and your team for an assiduous work, where you many times will be moving two steps forward and one step back. I remember one colleague with several years' experience advocating open access, saying in a discussion that "open access isn't a suitable work if you don't have the patience". I can see that this is true to a large extent, and it is also clear, that patience alone won't be enough to achieve what we want (and our stakeholders expect us) to achieve.

I sometimes compare the open access work with my previous experience from implementing a university-wide process to register all the bibliographic references for the published research output. Of course, this also required resources to provide information and instructions to authors and administrators (as well as participation in discussions on why this was necessary and how we should do it in a way that didn't do injustice to different publishing habits etc).

It was still a rather straightforward process, where we had a clear goal within reach when all the stakeholders got on the (same) train.

I find it useful to compare this process with the work to promote and implement an open access mandate policy. There are of course similarities, but also quite another kind of dynamics in the questions involved in open access. One of the reasons for this is, I believe, intrinsic to the way mandates/policies are being formulated. They can contain requirements as well as recommendations and also, I would say, even prayers: You shall deposit your manuscript in the repository; You are recommended to do your original publishing in an open access journal, and please, Dear Researcher, don't sign away your copyright.

Of course it is possible to promote and work with all these questions, but doing it at the same time and still being able to get a focused communication to the research community is definitely a challenge. However, the main reason for the dynamics in this process is that not all of the players in the field are playing the same game. While universities, researchers, libraries and funders are playing one game according to one set of rules, the publishers are playing another game.

I think it would have been easy (or at least more comprehensible from a pedagogical point of view) if we were just opposite teams, but that is clearly not the case. Instead it can now sometimes look like we are aiming at the same goal, but in reality it is different games being played out on the one and same field. And, as can be observed from any schoolyard during a break, this leads to a wide array of dynamic interactions.

Jonas Gilbert works at the library at Chalmers University of Technology where an open access-policy was adopted in 2010. He leads a section in the library for publishing services and bibliometrics. jonas.gilbert@chalmers.se

THE – HOPELESS? – QUEST FOR GOLD…

Jan Erik Frantsvåg

Here, at the University of Tromsø, we are trying to populate our repository Munin to the best of our ability. This entails both e-mailing authors and locating gold OA articles with licenses that allow us to post them to Munin without asking the author(s).

This we do because we feel it is important to try to show the world around us what comes out of the funding we are given, and also because it seems that more copies could result in a better chance of being read (thus used, thus cited …). And we believe that it is easier to persuade an author to self-archive in a repository where there are already a number of articles in his field, than asking him to be the first author in an empty and barren repository.

Combining data from the national CRIS Cristin – giving us full information on everything published by anyone associated with our university – and data from DOAJ and from Sherpa has given us lists of OA articles ready to be harvested, and articles in journals allowing self-archiving of post-prints. These last lists have been e-mailed to a total of 725 local authors, asking them to find their post-prints and self-archiving them. Needless to say, we are not overrun with articles, but an occasional self-archived article suddenly turned into a dozen a day for some days. And we believe this also raises awareness of self-archiving itself, lowering the barrier to spontaneous self-archiving at some later point. We have, after all, seen a small but consistent growth of such author-instigated self-archiving over the years.

But: What about the unknown number of articles in hybrid journals, where our authors have paid dearly to make their articles Open Access? Both DOAJ and Sherpa list journals, not articles. Surely, the publishers must have some tool for us to mine the gold? But, alas, this is not how it works. Even Springer, whom I knew to have such a service, had made it impossible to locate this service from their start web page. (At least, it was a task I was unable to do.) Only a bad habit of hoarding useful URLs[1] saved me, and made it possible to identify and download a list of all OA articles in Springer journals. This did not differentiate between Open Choice articles and articles in Springer Open journals, but as the latter ones seem to be few and far between at this time, it did not pose a problem. My collection held no links to similar services for other publishers, and after spending some hours going through Sherpa's list of publishers offering OA options trying to locate such services I conclude that either there are none, or they are actively hidden.

I know for certain that we must have paid for some kind of OA option with other publishers, but as things stand we are unable to show these articles to the world through our repository.

We should note, and make publishers note, it is **our** gold they hoard in **their** vaults!

Jan Erik Frantsvåg Open Access adviser, dept. of digital resources and services, University Library of Tromsø, Norway.

---

[1] http://www.authormapper.com

BIBLIOMETRIC RESEARCHERS – NEW SERVICE AT VILNIUS UNIVERSITY LIBRARY

Žibutė Petrauskienė

Bibliometrics is the application of mathematical and statistical methods to study and analyse the flow of documents and their bibliographic characteristics. A statistical analysis (citaton) method for bibliographic references is used to evaluate the the efficiency of scientific and scholarly publications, researchers and research group performance; to certificate researchers according to their citations and impact to scientific progress; to identify scientific schools and major branches of scientific and scholarly publications; to optimize the supply of information to the scientists; to improve the organisation and management of science, etc.

The idea of using the *Impact Factor* (IF) to evaluate scientific journals was first suggested by Eugene Garfield as far back as 1955[1]. The journal *Impact Factor* is the average number of times articles from the journal published in the past two years have been cited in the JCR[2] year.

The IF shows journal articles citaton frequency, the popularity, value and use of the journal. and hasa huge impact on the evaluation of scientific output, although scientists and specialists make contradictory assessments of this impact . The reason is, that the usage of IF for other purposes (i.e., for journal subscription selection processes in the libraries, for estimation of individual scientist's or research group's productivity in science management) faces a dilemma if scientific assessment on a macro level (science, country, world) applies the same criteria on the micro (individual) level. Citaton is the indicator of the influence of an article, author or journal, but is not necessarily a *quality indicator* of the individual publication or the researcher's activity. The language issue is a relevant problem. For example, scientists who publish their articles in Lithuanian riskbeing cited or referred to only by those who know of this language,

which means that the visibility of such article worldwide will be poor or equal to zero. To minminimize this problem, abstracts in major world languages are provided. In addition, different areas of science (such as humanities – history, philosophy, philology), especially if the research is conducted over a long period, do not publish their results in journals. In this case, books and monographs are one of the main forms for dissemination of research results.

Results of the survey on the *Need for bibliometric research*

The idea of providing bibliometric research services at the Vilnius University Library came up a few years ago, but it was important to know how it would be regarded by people at the University t: would they see it as necessary, useful, timely.

For these reasons,the Vilnius University Library (in January 2010) conducted a survey on the *Need for bibliometric research* ed.

The aim of the survey: to find out if researchers think that bibliometric research is useful and necessary and, if the answer is yes, what kind of research should be performed.

Main goals:

- To question VU scientists;
- To analyze the results of the questionnaire;
- To provide findings concerning the research priorities.

Organization of the survey and a sample: A questionnaire was prepared and placed online. A letter inviting participation in the survey was disseminated via the VU email list to University lecturers, researchers and doctoral students. 141 answer was received.

Key indicators of research results

In total 141 respondents replied to the questionnaire: 80 lecturers, 34 doctoral students, 19 researchers, and

---

[1] Garfield E. The History and the Meening of the Journal Impact Factor. JAMA. 2006, No 295 (1), p. 90 – 93. Available at: http://jama.ama-assn.org/content/295/1/90.full.pdf+html. Accessed February 22, 2012.

[2] JCR – Journal Citation Report
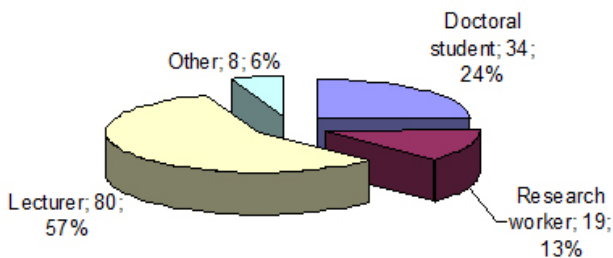
8 other respondents (see Figure 1).

Figure 1: Respondents by occupation.

The first question of the survey *Bibliometric research is necessary* was aimed to clarify whether the scientists think that such research should be carried out. 80% of the respondents agreed that such analyses are needed., 16% said they had no opinion, 3% did not know what it meant, and only 1% (1 respondent) said that such research is unnecessary (see Figure 2).
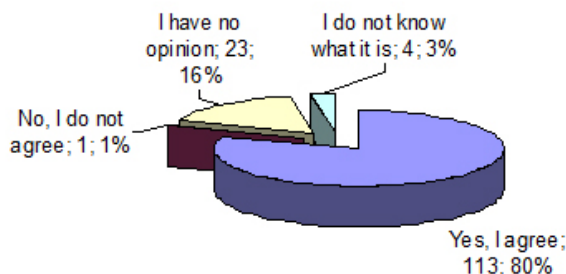

Figure 2: Answers to the proposition: "Bibliometric research is necessary".

Considering the scientists' opinion and proposed bibliometric research topics, the Scientific Information Data Center at Vilnius University library accomplished and presented the following analysis to the VU Senate:

- The Impact Factor and other parameters of those journals in which VU scientists published their articles in 2009.
- Comparative analysis of Vilnius University, Vilnius Gediminas Technical University and Kaunas

Technological University publications registered in Web of Science (WOS) in 2009.
- Analysis of VU publications registered in WoS in 2010.
- Analysis of VU publications in the humanities and social science registered in WoS .
- Analysis of VU publications in humanities registered in SCOPUS.
- Comparative analysis of publications in the humanities from Lithuania, Estonia, Finland, and Germany.

Representation of humanities publications in the Web Of Science database
At the end of 2011, the research project *Reflection of humanities in the database Web Of Science* was accomplished. The aim of the research was to determine how the humanities are represented in the database *Web of Science*, and whether this source is advantageous and appropriate for the registration and evaluation of articles in the humanities..
Chronological range: 1990–2011.
Source: Web of Science (WOS) database.
Subject: articles in humanities from Lithuania, Estonia, Finland, Germany in WoS.

The results show, that there are 1,395 humanities journals registered in WoS, whereas the number of social science journals is 4,500 and of science journals is 8m500 (see Figure 3).
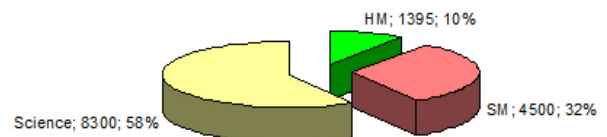

Figure 3: Journals' distribution in DB WoS according to the branches of knowledge.

Although the number of journals in humanities in the WoS database presents a fair amount, it is only 10% of all journals registered in this database.
The study shows, that the amount of humanities publications in WoS from the surveyed countries is slightly increasing (see Figure 4, 5, 6, 7).
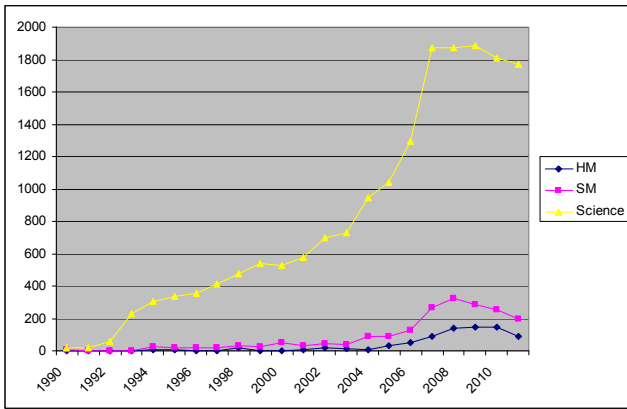
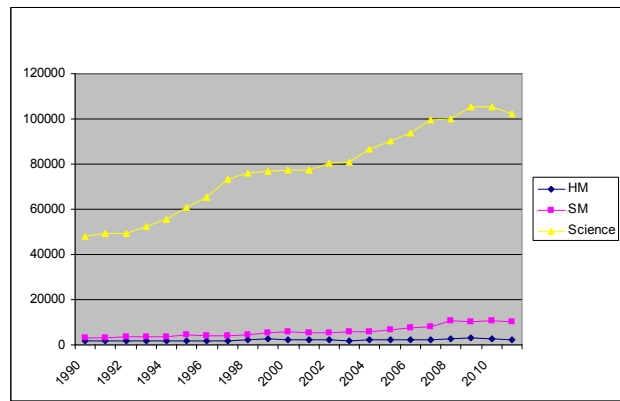*Figure 4: Dynamic of scientific publications, Lithuania, WoS, 1990 – 2011.*

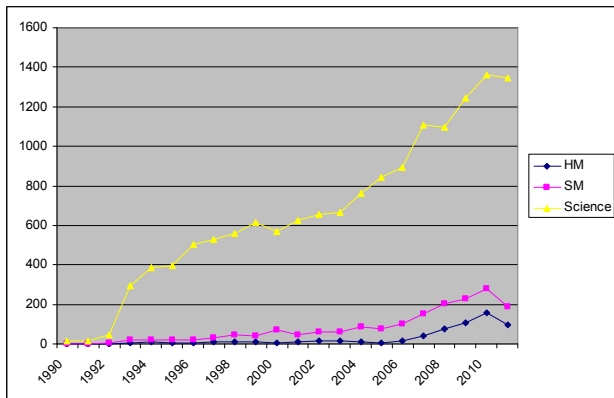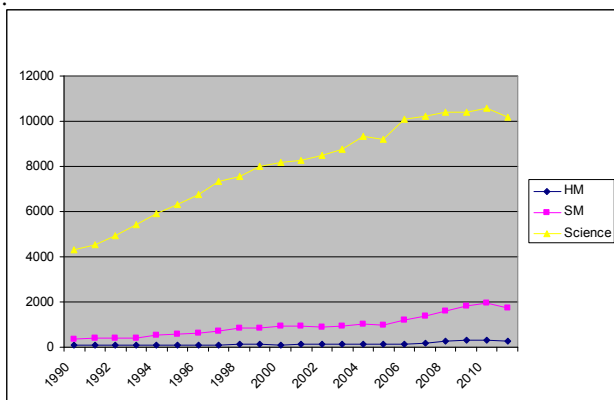*Figure 7: Dynamic of scientific publications, Germany, WoS, 1990 – 2011..*

However, it must be acknowledged that in comparison with publications in other subject areas humanities cover a very small part of the database(see Figure 8).
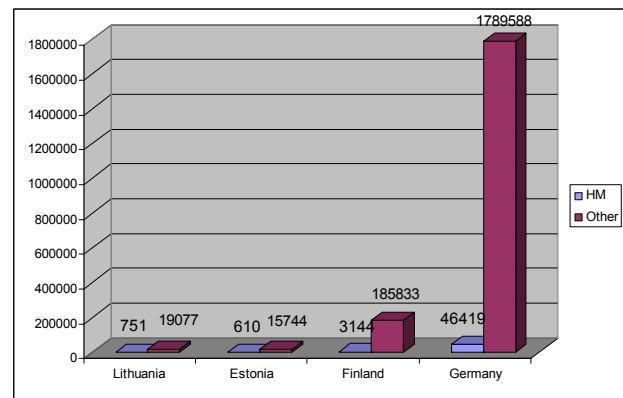

Figure 5: Dynamic of scientific publications, Estonia, WoS, 1990 – 2011.

.

Figure 8: Comparison of humanities and other publications, WoS, 1990 – 2011.

Figure 6: Dynamic of scientific publications, Finland, WoS, 1990 – 2011.

Summary:

A general trend can be observed: humanities publications in WoS registered journals represent a very small proportion compared to other sciences: from Lithuania and Estonia – less than 0,04%, from Finland – less than 0,02%, from Germany – less than 0,03%. In addition, journals in the humanities consist only 10% of the total amount of journals registered in the WoS database. Therefore, it can be concluded, that the Web of Science database offers an insufficient representation of research in the humanities.

Dr. Žibutė Petrauskienė Head, Scientific Information Data Centre, Vilnius University Library, Universiteto str. 3, LT-01122, Vilnius, Lithuania. http://www.mb.vu.lt
zibute.petrauskiene@mb.vu.lt

BIBLIOMETRIC RESEARCH OUTPUT INDICATORS AND UNIVERSITY FUNDING IN THE NORDIC COUNTRIES

Hampus Rabow

Funding of research and the role of governments

Traditionally, decisions about research funding have mainly been informed by three basic criteria. One has been the potential practical utility of the research, another, the quality of the research as perceived by peers, and finally the conformity with the preconceptions or imagination of the funders. The last criterion is generally perceived as illegitimate and will not be further considered here. The first criterion has with the advancement of science and technology comes to be more and more dependent on the second. In other words, it has become increasingly more difficult for a non-peer to determine the probability that a certain research project will provide some beneficial applications.

There are of course usually easily discernible differences between research primarily devoted to some intra-scientific problem, and that with some extra-scientific application in mind. Between these two extremes, there are however a whole spectrum of possibilities. It is also frequently the case that some more theoretical research is discovered to have practical applications, whereas many applied research projects turn out to be practically useless. Another important aspect is that practical applications of research frequently appear in other contexts than those originally intended. This is not the place to discuss the complex relationship between theoretical research and possible applications, but it is important to note that the need to protect basic research from the demand for immediate results has been of fundamental importance for the development of the modern research system.

When, towards the end of WWII, Vannevar Bush wrote his famous report, which stressed the need for government funding of science and education, and outlined, what would eventually become, the *National Science Foundation*, he also formulated five essential principles which may be summarized (somewhat simplified) as follows: (1) The funding must remain (more or less) stable over several years, (2) decision makers should be non-partisan professionals with a good understanding of science and education, (3) the funding agency may only distribute funds to outside institutions such as colleges, universities, and research institutes, (4) as long as the provisions of the funding application are followed, the funding agency can not exercise any influence over the research, and (5) the funding agency should be assured "complete independence and freedom for the nature, scope, and methodology of research" while at the same time being responsible to the executive and legislative branches of government.[1] These five principles were derived from the necessity to sustain "basic research". From a global perspective, basic research was required to make progress in applied research possible, and from a national perspective, basic research was required to maintain a competitive edge. As Bush put it: "A nation which depends upon others for its new basic scientific knowledge will be slow in its industrial progress and weak in its competitive position in world trade, regardless of its mechanical skill. "[2]

The road towards bibliometrics

The system proposed by Vannevar Bush turned out to be quite stable and successful, and was, with some minor modifications, duplicated in many countries around the world. Yet, the increased demand for government funding of science also created a demand for increased accountability, which would seem to require a more efficient and transparent (to the taxpayers) process for allocating resources. Apart from the lack of transparency to outsiders, the peer review process for grant applications tended to be either too unreliable or too costly. A reliable review required many days of work both by the applicants and the reviewers. In the early 1960s, political criticism of the *NIH* led to a large evaluation carried out by experienced research administrators. The final controversial report suggested that NIH reviewers lacked competence in certain areas and proposed that Bush's third principles should be implemented. It also recommended an increase use of "administrative devices" in the decision process.[3]

At the same time, the complexity and size of the research divisions of some large corporations had grown so much, that the research managers needed

---

[1] Bush (1945) esp. pp 32-33
[2] Ibid p 19
[3] For a summary see Greenberg (1965). Criticism in Cooper (1965)

some systematic and science-based approach to monitoring and quality control. In 1958, the main US organization for research directors began publishing an academic journal called *Research Management* adopting approaches from operations research and management science to the problems of research administration. During this period, new successful methods for the quality control of production had been developed in the Japanese manufacturing industry, and many research managers believed that similar methods could benefit research & development.[4] This was also the period of the Cold War, and there was a widespread fear that the Soviet ability to steer resources to prioritized areas of military research would give them an advantage in the arms race. During the early 1960s both the US navy and *NASA* sponsored large projects to increase the efficiency of research management.[5]

In 1963 the physicist and historian of science Derek de Solla Price proposed a new research programme in order to solve many of the above-mentioned problems. This research programme was first called the "science of science" and later "scientometrics", and it laid the foundation for modern applications of *bibliometrics* in the areas of science and technological innovation.[6] Price wanted a unified approach where knowledge from the history, sociology and psychology of research would inform the statistical models used in research administration and thus provide support for decision makers in the realms of research management and science policy.[7] Although the overall programme had very limited success, some of Price's core ideas gradually became more and more influential. He proposed that the new bibliographic database called *Science Citation Index* could be used as the main data source for a statistical analysis of scientific research. It would be possible to examine and measure the spread of scientific information, model the structure of science (in terms of research fields and their relatedness) and study patterns of cooperation between individuals and institutions.

One of the major areas of application for the new bibliometric methods was, just as Price had envisioned, for research evaluation and funding decisions. At the same time, it was initially difficult to find the right place for this kind of studies as an aid in funding decisions. The bibliometric indicators were often seen as crude and unreliable in comparison with well established methods of peer review. The main use of bibliometrics in this context has been to calculate performance indicators that help ascertain the

quantity, quality and impact of previous research. A number of studies have shown that simple bibliometric indicators correlate well with the results of large complex evaluations based on peer review. Problems only occurred when too few of the most important publications were included in the citation database.[8]

What you measure is what you get

Administrative decision makers in many types of organizations often use performance indicators to create negative and positive feedback loops that counteract or promote certain types of behaviour. The indicators may also be used simply to see how the input in terms of resources and funding corresponds to the output in terms of some useful products. In the scientific research system, bibliometric research indicators may thus be used to steer the research activity in a certain direction or simply to see how much funding is required to produce a certain quantity of scientific output or impact.

In their classic paper on "the balanced scorecard" Kaplan and Norton stressed the importance of the positive feedback loop with the use of performance indicators. As they succinctly put it: "What you measure is what you get".[9] Outcomes not reflected in the indicators may be ignored in funding decisions, and individuals or units tend to give priority to the activities that are counted. Rather than trying to find extremely sophisticated indicators that are sensitive to all important activities, Kaplan and Norton's solution was to use a complex scorecard, where different perspectives were balanced against each other. Thus most forms of "Larsen effects", where a particular activity is given too much priority, may be avoided. On the other hand, a more complex system may be seen as less transparent, and the direct steering effects may be smaller. Ronald L. Straight has stressed the need to clearly define the goals, metrics, and weights, in any performance metric system.[10] Applying his principles to bibliometric performance indicators, the goals implied in the metric must be generally accepted among the scientists, the metric must match these goals as closely as possible, and the weights should be explicitly adjusted to match research priorities. Looking at the bibliometric indicators most frequently used in research evaluation from the 1960s and onwards, they fulfil Straight's first criterion reasonably well in most cases. Goals like "scholarly impact", "submission to qualified peer review", and "productivity" are widely accepted in the scientific and scholarly communities. For the second and third criteria, however, the situation has been more ambiguous. This may be one of the reasons why the

---

[4]See Montgomery (2007) for an introduction to modern quality control. It includes a brief historical overview.
[5]For more examples see e.g. Dedijer (1966)
[6]Price (1963). See also Garfield (1987) for an appraisal.
[7]Price (1965)

[8]E.g. Norris & Oppenheim (2010) , Oppenheim (2007)
[9]Kaplan & Norton (1992)
[10]Straight (2000)

general use of bibliometric indicators in funding decisions has been slow to gain acceptance in the Nordic countries

Bibliometrics and government funding in Nordic countries

In the early 90s several international bodies promoted a greater international conformity of the evaluation of higher education and research. The most important initiative for the Nordic countries was probably a pilot project launched by the European Commission in 1994. The aim was to develop a unified approach to academic evaluation in all European countries. All Nordic countries took part in the the project despite some not being EU members.[11] These evaluations did not generally make much use of bibliometrics, but they were an important step towards standardized performance indicators (even if only in the form of qualitative reviewer scores). At this time there were also many *ad hoc* research evaluations performed for various purposes and such evaluations, especially in the fields of technology, medicine and natural science frequently made use of bibliometrics as a complement to more qualitative peer review. This increased focus on academic evaluations also led to attempts to follow up on such evaluations with strategic funding of institutions with a high performance or potential. Finland created special *Centres of Excellence*, and Sweden transferred the money from the employee funds to a new foundation for "strategic research", in order to support research "of the highest international standard".[12] On the other hand, the core research funding for colleges and universities were largely left intact, except in so far as that funding depended upon external grants.

The Norwegian model

Not until around the year 2000 was the next step taken towards a more direct feedback from university research performance to government funding. In the Norwegian government report *Frihet med ansvar* ("Freedom with responsibility"), the committee proposed that the core funding to the institutions should be made partly dependent on "the results of the institutions' activities".[13] This proposal was later followed by a decision to create a national standardized publication database and a bibliometric output indicator for Norwegian research. In 2003 the university of Oslo introduced a very simple output indicator called the "publication score". This publication score was calculated by giving different weights to publications depending on whether the *publication channel* (journal , book or series) had an international scope or not. The publication channel

[11]Thune & Staropoli (1997)
[12]SOU 1996:29 p 186-187
[13]NOU 2000:14 p 46

also had to have a generally acknowledged scholarly or scientific importance for the publication to be counted att all.   The amount A_f given to each faculty f was calculated from the publication score P and a "cost factor" C according to the following formula: $A_f = A * (Pt_f / \sum Pt)$, where $Pt = P * C_f$ A was the total amount of funding to distribute, and the cost factor C (which was optional) was supposed to reflect different research costs at different faculties.

The Oslo model was clearly too simplistic to be useful, but it were to provide the basis for the new national output indicator.[14] Only scholarly and scientific publication channels with at least a national reach were accepted. Out of these, around 20% with an especially high international importance were placed in "level 2" and assigned a higher weight. The selection of publication channels was done by specially selected committees of researchers for around 40 different subject areas. Three different publication types were used: (1) articles in a journal or series, (2) monographs, and (3) book chapters in anthologies. Thus no distinction was made between e.g. different types of articles or between diffferent kinds of anthologies or between conference proceedings and yearly reviews. Once the type of a publication and the level of its channel had been determined, a publication score was assigned as follows:

Publication type	Level 1	Level 2
Monograph	5	8
Article in journal or book series	1	3
Article in anthology	0.7	1

The Swedish model

In Sweden, a new model for resource allocation was proposed first in general in 2005, and then in details in 2007. Rather than using the simple Norwegian output indicator, based upon the quantity of output and the general importance of the publication channels, a citation-based model was proposed. The citation model had been developed by Ulf and Erik Sandström and was based upon standard bibliometric indicators such as the Leiden "Crown Indicator", but with some modifications[15]. The Sandström model was approximately as follows (the actual model has been a little simplified here): A citation score was calculated based on all publications from each college or university in the *Web of Science* during a 4-year period. The number of citations to these articles was divided by the world average for the same research area

[14]See Sivertsen (2003) and Sivertsen (2007)
[15]Sandström & Sandström (2008) and SOU 2007:81

and the same type of article. When many universities had contributed to the same publication, the citation score was divided equally among them. The total citation score would thus indicate the total impact of a university or college in the corresponding research area. In order to adjust for the difference in productivity between different research areas, the total citation impact was adjusted by a "Waring factor" corresponding to the estimated average productivity for each of 23 different research areas.

There were several reasons why Sweden preferred the Sandström model to the Norwegian model. The division into publication types and publication channels was seen as too crude and overly simplistic. Another problem was the disregard for differences in productivity between different research areas. A study by the Swedish Research Council had shown how an average doctoral dissertation in Medicine only gave 28% of the score of a Humanities dissertation. A third problem was that the Norwegian model ignored the actual scientific impact of the research. Seglen and others have shown that there is very little correlation between the impact of individual papers and the overall impact of their respective research channels.[16] Finally, it was felt that the Norwegian model was too conservative and failed to send any clear signals to researchers.

The Sandström model was not completely uncontroversial in Sweden. The Research Council argued that it had several flaws, and especially that the adjustments made for differences in productivity between different subject areas were based on an erroneous model. Another difficulty was that some research areas, especially in the Humanities, only had few of their publications in the form of journal articles, and only some of these journals were indexed in the *Web of Science*. Thus, it was decided that actual citation impact would be ignored for the humanities.

Other Nordic countries

Denmark began to implement their model slightly after Sweden and thus had the option to base their model on any of the previous two. In the end, they decided to use a version of the Norwegian model, though it is not entirely clear why. When reading the Danish discussion of the Swedish model, it is difficult to avoid the conclusion that it is partly based on misunderstandings. The Danish model contained some improvements, notably some clarification of ambiguities and a more complete list of publication types.

Recently Finland has also decided to use a version of the Norwegian model with the full implementation beginning in 2015. Finland has added an additional

level for publication channels, meant to indicate a channel of the highest international importance.

The situation in Iceland is a bit special with only one large university and six smaller. Only a small amount of the total public R&D expenditures could be labelled "competitive funding", mostly through council grants and centres of excellence.

Discussion

When discussing bibliometric output indicators in relation to funding, it is important to distinguish clearly between two different functions. The bibliometric scores may, on the one hand, be used to indicate the *overall volume* of various research activities, and on the other to indicate the fulfilment of some *normative criteria* for these research activities. The main purpose of using bibliometric in the first case is simply to relate output to input. In other words, the funders have a legitimate interest to know what they get for their money. If an institution produces comparative little in relation to the resources given, there are legitimate questions to be asked. The indicators are used mainly as a gauge of efficiency. In the second case, funding is used to steer research activity according to certain policies or epistemic norms. Here the indicators are used mainly to regulate behaviour.

As was mentioned above, bibliometric performance indicators have often been used with a limited adherence to Ronald Straight's second and third principles. In the case of the Norwegian model, there is also considerable ambiguity concerning the first principle. Is the indicator meant to have some kind of positive effect on the *kind* or *quality* of research produced, or is it primarily intended to relate the supplied resources to the achieved outcomes? With the Swedish model, the authors have been very clear that it has a dual function. It is not only meant to indicate to what extent the objectives have been fulfilled, but also to reward research institutions with a significant international impact, and to punish institutions whose research fails to impress the wider research community.[17] As concerns the second principle, the choice of adequate metrics, it is of course impossible to evaluate the metrics without a clear notion of the purpose. But if the only purpose is to have an indicator of efficiency and help estimate the amount of resources required for a certain volume of research, then it should be imperative that the research indicator actually correlates well with the actual amount of research activity, with adjustments for differences in efficiency. In the development of the Swedish model, this problem has been at the forefront, even if the solution is far from perfect. In the development of the

[16]Seglen 1997

[17]Sandström & Sandström (2008) p 8. Compare also Vetenskapsrådet (2012)

Norwegian model, the problem has been largely ignored. Straight's third principle relates to priorities. Here it is essential that the funders and recipients agree what the priorities are. This obviously does not mean that all researchers should agree with the government's priorities, but they should be aware what they are and be able to see how they are reflected in the metrics.

The current Swedish model has recently come under review, and Anders Flodström's report has suggested that Sweden move towards something more akin to the Norwegian model.[18] Unfortunately Flodström's discussion of bibliometrics is often quite difficult to follow. One example: "Either the quality is measured through how many times the publication has been cited in other publications or an impact factor for the the journal where the publication occurs is used."[19] Perhaps this is simply a very careless statement, which should not be given too much weight, but bibliometrics is based on *statistics* and *indicators*. Bibliometrics can never "measure" the "quality" of a scientific publication. Another example is when Flodström discusses the adjustments of the citation score with regard to the the different levels of productivity (as measured by publications) in different research areas. As was mentioned above, the solution provided by Ulf and Erik Sandström was far from perfect, but Flodström has apparently not given it much thought at all. He writes as follows: "The current distribution system makes use of the Waring model, which takes account of different research areas through the field adjustments of publications and field normalization." Perhaps this is simply a grammatical mistake, but taken together with other statements[20], and hardly any discussion of the issues involved, it is difficult to avoid the above conclusion.

A fundamental problem with all performance indicators is that adequacy frequently must be sacrificed for simplicity and transparency. A more advanced indicator may conform better to the desired outcomes ("what you measure is what you get"), but it may be difficult to fully understand by administrators and practitioners. Also it may make comparisons over time, or with other similar organizations more difficult. There are today many advanced bibliometric indicators that reflect intuitive concepts like "scientific impact" or "productivity" fairly well. On the other hand, a funding system based on such indicators would probably be rejected, since the considerations involved would be insufficiently understood outside the world of bibliometric research. Thus it may be preferable to use simpler indicators, such as those in the Sandström model, although with some improvements. One of the major problems with citation based measures has been the lack of coverage for certain academic fields. Recently, however, the *Web of Science*, has been complemented with a *Book Citation Index*. Thus this problem will now only be a major issue for a few research areas, and for these areas some improved version of the Norwegian model is probably to be preferred. A major weakness with the Sandström model is that it ignores the status of the publication channels altogether. This has the effect that there are several years of delay in the feedback loop, and that world class research may be ignored, if bad luck with the timing prevents it from being cited. For these reasons, an improved version of the Sandström model with the addition of a channel based score would probably be the best solution for the Nordic countries.

Hampus Rabow Malmö Högskola, Sweden, hampus.rabow@mah.se

[18]Flodström (2011)
[19]Flodström (2011) p 34
[20] e.g. "With the use of the Waring method the citations are field normalized" and "the productivity per researcher may be seen directly" (p 35)

References

Bush, V., 1945. Science, the endless frontier. AYER Co. Pub.

Cooper, J.D., 1965. Onward the Management of Science: The Wooldridge Report. Science 148, 1433 –1439.

Dedijer, S., 1966. The science of science: A programme and a plea. Minerva 4, 489–504.

Flodström, A.. 2011. Prestationsbaserad resurstilldelning för universitet

och högskolor. Stockholm.

Garfield, E., Essays of an Information Scientist, Vol: 10, p. 72, 1987 Current Contents,# 11, p. 3, March 16, 1987.

Greenberg, D.S., 1965. NIH Study: Wooldridge Committee Praises Past Efforts, But Urges Major Organizational Revisions.

Science, New Series 147, 1556–1559.

Karlsson, S & Jacobsson C., 2007. Vad är en svensk doktorsavhandling värd enligt det norska systemet attvärdera publikationer?

Bilaga 7 SOU 2007:81 Resurser för kvalitet.

Kaplan, R.S., Norton, D.P., 1992. The balanced scorecard–measures that drive performance. Harvard business review 70, 71–79.

Montgomery, D.C., 2007. Introduction to statistical quality control. Wiley-India.

Norris, M., Oppenheim, C., 2010. Peer review and the h-index: Two studies. Journal of Informetrics 4, 221–232.

NOU 2000:14 Frihet med ansvar : om høgre utdanning og forskning i Norge : utredning fra et utvalg oppnevnt ved kongelig

resolusjon 30. april 1998.   Statens forvaltningstjeneste, Informasjonsforvaltning :, Oslo.

Oppenheim, C., 1997. The correlation between citation counts and the 1992 research assessment exercise ratings for British

research in genetics, anatomy and archaeology. Journal of documentation 53, 477–487.

Price, D. de S., 1963. Little Science, Big Science. New York: Columbia University Press.

Price, D. de S. J., 1965. The Science of Science. Bulletin of the Atomic Scientists 2–8.

Sandström U. & Sandström E., 2008. Resurser för citeringar Rapport 2008:18 R Högskoleverket.

Seglen, PO 1997. Why the impact factor of journals should not be used for evaluating research -- Seglen 314 (7079): 497 --

bmj.com 314.

Sivertsen, G., 2003. Bibliografiske datakilder til dokumentasjon av vitenskapelige publikasjoner: En utredning for Utdannings-og

forskningsdepartementet. Oslo: NIFU.

Sivertsen, G., 2007. "Den norske model": Et bidrag til diskusjonen i Sverige.

SOU 1996:29 Forskningsfinansieringsutredningen. *Forskning och pengar: slutbetänkande*. Stockholm: Fritze

SOU 2007:81 Resursutredningen. *Resurser för kvalitet: slutbetänkande.* Stockholm.

Thune, C., Staropoli, A., 1997. «The European Pilot Project for Evaluating Quality in Higher Education». Standards and quality

in higher education 198–204.

Vetenskapsrådet (2012) Remissvar till rapporten Prestationsbaserad resurstilldelning för universitet och högskolor.

http://www.vr.se/download/18.13384c8f135aad61b5516f6/GD-beslut+28_120313_Remissvar+resurstilldelning.pdf

KUBIS DATAVERSE NETWORK
A repository for primary data at the The Royal Library / Copenhagen University Library and Information Service

Thea M. Drachen, Asger V. Larsen, Bertil F. Dorch

The KUBIS Dataverse Network (https://data.kb.dk/dvn) is a digital archive for scientific primary data for use by researchers at The University of Copenhagen. KUBIS Dataverse Network is available to researchers, departments and institutes at the University and research groups with an affiliation to the University of Copenhagen.

Some of our researchers expressed a demand for this kind of service

The KUBIS Dataverse Network was established on the basis of the results of a "Report of a qualitative study of social sciences collection, use and sharing of primary data" (in Danish) (http://www.hprints.org/hprints-00451000). This report concluded, among other things, that there was an urgent need among a considerable part of the surveyed researchers for backup procedures and safe archiving of research data. Furthermore that there was a need to be able to archive data in many different formats: audio, images, text, numbers, etc. In addition there was a desire amongst scientists to share data with others and that the researcher him or herself could control who had access to these data.

What is the Dataverse Network?

The KUBIS Dataverse Network serves as an additional backup system, which archives data in a structured form for both short and medium term preservation. It can also serve as a way of sharing data. Each researcher/group can create his/her own Dataverse in the KUBIS Dataverse Network and can store and process the data, and if he/she chooses to share his/her data. The open source application was developed by The Institute for Quantitative Social Science (IQSS) at Harvard University (http://thedata.org/, link to the IQSS Dataverse Network: http://dvn.iq.harvard.edu/dvn/). The KUBIS Dataverse Network is not hosted in the cloud, but on secure servers at the Royal Library / Copenhagen University Library and Information Service, using 256-bit encryption.

The Dataverse Network is hierarchically organized with a number of dataverses (each created by e.g. individual researchers, university institutes, or a research project groups). Each dataverse creator has the option to create one or more "studies" in that particular dataverse (as a top level organization of the contents). Files can be public or restricted as default, but each file can be set to public, restricted to all, or restricted to all but the persons or groups granted access by the uploader.

Each study is given a persistent identifier (Handle) at upload and version history is saved, enabling the author to give permanent links/handles to the uploaded data. When creating a study and uploading a file, the uploader will be asked to fill in metadata for that file, which greatly enhances the possibility of reusing data at a later stage by the same researcher or others he/she might give permission. These metadata are typically filled in at the study-level, so the amount of metadata one has to fill in for individual files is limited.

The amount of metadata that can be attached is vast, but just a few fields are mandatory (such as name and creator). There are no restrictions to the file format in a Dataverse, what you put in is what you get out (only SPSS file formats SAW/POR will change to TAB).

... and what is it not

The Dataverse is not a virtual research environment or an analysis tool. Apart from some statistical analyses (using R), you need to download the file(s) you want to analyse.

You cannot bulk upload to the Dataverse Network. Each file needs to be uploaded separately. For files you want to share in a collective group, you might want to zip the files to cut down the number of times you need to upload.

Accessing the Dataverse Network the first time can be cumbersome. Once you get the hang of the very hierarchical organization of the Dataverse Network it is fairly easy to navigate in it, but it is not very web 2.0.

Derived services

We got a request from a research group at the Department of Cross-Cultural and Regional Studies at the University of Copenhagen. They had data they wanted to share and research they wanted to promote, they asked us for help and we ended up giving them a

combination of a webpage, finding and uploading to repositories the publications for which we could get copyright clearance, and setting up a dataverse for the Department for their data. Read more here (http://shkaratmsaied.tors.ku.dk/).

This was such a success, that we started up Research@KUBIS (http://libguides.culis.kb.dk/researchatkubis), which uses the Dataverse Network as one of its three components. It is a help towards 1) building a project site, 2) promoting data and 3) promoting publications.

The project site platform (LibGuides) is especially interesting for researchers in cross-institutional collaborations to use, since project collaborators are not required to be affiliated with a certain University or Institute to be given editor-rights to a project page. And the system is WYSIWYG (What You See Is What You Get) and very easy to work in.

We have only just started promoting the KUBIS Dataverse Network and the Research@KUBIS services, but an increasing number of researchers and research groups have shown an interest in using the systems.

Thea M. Drachen, Research Support Service, The Royal Library / Copenhagen University Library and Information Service, forskerservice@kb.dk

Asger V. Larsen Research Support Service, The Royal Library / Copenhagen University Library and Information Service, forskerservice@kb.dk

Bertil F. Dorch, Research Support Service, The Royal Library / Copenhagen University Library and Information Service, forskerservice@kb.dk

PERSISTENT IDENTIFIERS IN DENMARK ANNO 2012
Adrian Price

At long last the initial steps have been taken to introduce a Persistent Identifier (PID) infrastructure in Denmark. A project group, which has received support from Denmark's Electronic Research Library (DEFF)[1], has over the last year or so been working on establishing the basic building blocks of this infrastructure. The members of the project group come from the Royal Library in Copenhagen (project lead), the State and University Library in Århus, the Library of the Faculty of Life Sciences, University of Copenhagen, and the Technical Information Center, Technical University of Denmark. After considering the requirements for a PID system, which would fit well into the Danish setup and the available resources, the very basic building blocks have now been established.

I will in this article describe these "basic building blocks", as well as give a brief overview of Phase two of the project, which has also received support from DEFF and will be starting shortly in continuation of the first part of the project. There is no ambition here though to cover in any way the variety of "PID systems", standards and software available for exploiting the use of PIDs for research activity - for example DOI, Ark, Handle, the Datacite organisation, URN etc. There is ample information available, with excellent coverage of all the alternatives, so Yet Another Identifier Overview will not be attempted!

In the first stages of the project a short-list of basic requirements for a PID system was drawn up, and parallel to this the available software alternatives were investigated. In light of the available resources and other activities which were emerging in related areas of the research landscape, it was decided to use the Handle[2] software as the basis for the administration and resolving of PIDs. There were several reasons for choosing Handle: it is well-established, has a global resolving capability, has a large international user-base, and although at the moment it is managed by the

Corporation for National Research Initiatives (CNRI) in the US, there are plans to find an international organisation which can take over its administration. This last point is considered important, as a PID infrastructure by nature requires a stable and permanent organisational basis. But another important factor also influenced the decision: using the Handle software would not require using a large amount of resources on software development, both in the initial establishment and in continual future development.

The actors in the PID landscape

The actors in the Danish PID landscape are for example universities, research libraries, cultural heritage and government institutions – which both publish and preserve research publications, reports and other digital objects. This digital material requires a means of ensuring permanent access as well as preservation.

Organisations and their information management systems are unstable things. When one institution gets taken over by another, are links to digital objects maintained, are even the electronic objects themselves retained? It is becoming more and more imperative that an infrastructure be established, with rules and regulations regarding the preservation of electronic material, as well as, and not least, the preservation of active links to this electronic material.

The research publications from Danish universities are a very important part of the "content" of the infrastructure. The universities are in a rather special situation: all Danish universities use the same repository software, PURE, developed by a Danish software company, Atira.[3] In a version released towards the end of 2011, it has been possible to install and configure a Handle server (in a manner similar to DSpace), which enables the issuing and resolving of Handles for local objects. Unfortunately, the details of how this "integration" between PURE and the Handle

---

[1] http://www.DEFF.dk/
[2] http://www.handle.net/

[3] http://www.atira.dk/en/pure/

server was implemented, was not available in phase one of this project, so exactly how the Danish universities will be brought into the infrastructure is unclear at the moment. This will be a major task to be addressed in phase two, which will begin immediately after phase one has been completed.

In the PURE-Handle implementation the administration of Handles is taken over by PURE, so there has to be established a mechanism whereby "PURE Handles" can be integrated into the central Handle resolver. Another reason that this is an important issue is the fact that this situation with parallel and independent Handle servers, will most certainly crop up again in the future: in fact it already exists, in that two Danish universities already have Handle servers which function in a similar manner. The Copenhagen Business School and Roskilde University both have DSpace, and thereby Handle servers, connected to their PURE installations.

Now that in Denmark the foundation has been laid, future phases of the project will seek to include the important actors in the landscape as a way of stabilising access and preservation in the research information environment.

Organising PIDs: danPID

As has been emphasised over and over in the persistent identifier literature, the challenges are not only technical but also administrative. Any system which relies on the implementation and maintenance of standards relies on the existence of and respect for these standards, which in turn requires an administrative setup between the partners. Another important factor of shared systems is the question of resources and financing: who will pay for maintenance and further development of services. PID systems are certainly no exception to this.

In phase two the ambition is to create an organisation which will administer the future development of the Danish Handle services. This organisation has been given the nickname "danPID". The detailed tasks of danPID are not clear at the moment, but a cooperation of member institutions, who actually contribute to content, is envisaged. The main tasks will be connected with further development of services and standards, supervision of quality requirements and keeping a watchful eye on PID developments internationally. The question of financing will be

attacked from several directions, but as with all infrastructural development in the meta-research environment, resources are limited and under pressure.

The basic building blocks

In the laying of the initial building blocks, several issues have been addressed. PID terminology is an important element in communication, and is not very standardised within the PID community. Therefore a basic list containing PID and Handle terms has been produced for use in this project and will be refined as the need arises. (It is not included in this article for reasons of space, and because a number of the terms have been defined for internal reasons and are not an attempt to standardise outside the bounds of the project.)

An important part of the terminology was related to how a "PID URI" based on the Handle system should be formed. This is illustrated in Figure 1, where a persistent identifier URL and a Handle in a citation format are shown. The separate elements of this correctly formed Handle PID illustrated in Fig. 1 will be treated in what follows as they apply in this project:
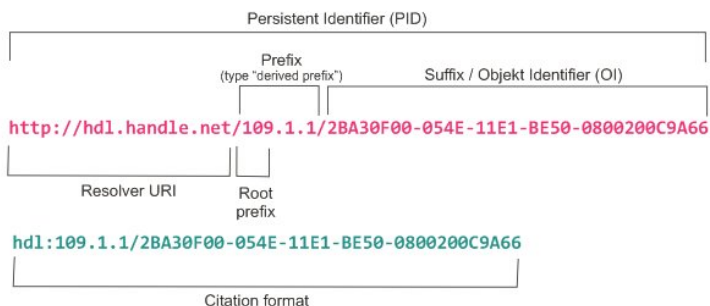

Fig. 1: Components of a Handle PID

Details regarding the individual components and rules as to how they should be formed can be seen in the Handle documentation. The relevant elements here are: the Handle prefix "109", which is the registered Handle prefix for Denmark. The project has decided that the two following positions will be used to indicate, first, the responsible institution and second, the system in question, which is owned by the institution in the previous position.

It is danPID which is responsible (or will be responsible), for issuing the first two positions ("109" and "1", the institutional position) and it is the institution which owns the system that wishes to issue

identifiers and be a part of the infrastructure, which issues the third position. This can be illustrated, as shown in Fig. 2:

Fig. 2: The derived Handle prefix

Using the example in Fig. 1 and 2:

109.1.1:
109: Denmark
1: The Royal Library, Copenhagen
1: A Dataverse system hosted by The Royal Library

Concerning the suffix, in the example used in Fig. 1 and 2, the suffix illustrated is a Universally Unique Identifier, a UUID. A UUID is a much-used type of identifier. On the initial short-list of PID requirements were a number of issues directly related to how the suffix part of the PID (the object, or unique identifier) should be formed. There are 7 conditions which apply concerning the suffix:

1. It should be possible to issue an infinite number of unique identifiers.

2. These unique identifiers must be able to be machine-generated.

3. There must be no limits as to how unique identifiers should be formed.

4. It must be possible to be able to generate unique identifiers locally (where by "locally" means the individual institutional actors in the danPID environment).

5. It must be possible to locally decide which system or method will be used to generate unique identifiers.

6. That each identifier is unique is a local responsibility.

7. It is recommended that unique identifiers are without semantics, i.e. there should not be used codes for collections, abbreviations for organisations, people's names etc. in identifiers.

A UUID has the advantage that it can be generated independently of the issuing system or organisation, which has responsibility for complying with the 7 conditions explained above. They can also be generated online or by the system where the object resides.[4] A disadvantage is that they are long (32 characters) and therefore not especially suitable for use in, for example, lists of references. Therefore it would be advisable if shorter identifiers could be created, which comply with the 7 conditions mentioned above. Systems do exist where this can be done.

As the first phase of this project come to an end, four Handle prefixes have been registered and are in operation at The Royal Library in Copenhagen and The State Library in Århus.

Conclusion
The first phase of the project is now drawing to a close, and the final report will be available from the DEFF website shortly. As already mentioned, DEFF has decided to support a phase 2, where the project group will continue with the further development of the infrastructure, expanding to include other important actors, especially universities and other public institutions.

Another important issue for phase two will be danPID: what form short danPID take, who will participate and how, as well as the vital questions regarding financing and which model will it be possible to agree on.

Adrian Price Information specialist, Library, Faculty of Life Sciences, University of Copenhagen

---

[4] But are there enough UUID's available? There are 340,282,366,920,938,463,463,374,607,431,768,211,456 available. " This means that 1 trillion UUIDs have to be created every nanosecond for 10 billion years to exhaust the number of UUIDs."

A BRIEF REPORT FROM THE 6TH MUNIN CONFERENCE ON SCIENTIFIC PUBLISHING 22-23 NOVEMBER, TROMSØ, NORWAY

Leif Longva

The University of Tromsø Library has since 2006 hosted an annual conference on scientific publishing. The 2011 conference had a special focus on new and additional services made available alongside and connected to the published articles. "Enhancing publications" was the subtitle of the conference.

Cameron Neylon, CTFC, UK: *"I need to publish more and read less! How new platforms will enable you to publish more effectively while reducing information overload"*

The first keynote speaker was Cameron Neylon from Science and Technology Facilities Council (STFC) in UK. Cameron focused on the increasing pressure on researchers to produce publications. The competition for increasingly scarce resources is bringing this pressure to critical levels, claimed Neylon. At the same time, the web creates new opportunities to communicate, both to other researchers through the sharing of things beyond the traditional journal publication, and to the wider community through lower costs of distribution. This is both a benefit, bringing opportunities for new forms of publication and reducing the costs of sharing research outputs like data, software, and research records, but it is also a burden, as it brings new expectations of the availability of information. The result of these pressures is an explosion of information which challenges our ability to cope.

Neylon discussed how we may shape the way that we communicate, and the way that we discover information. Building filters to help scholars avoid unuseful publications, is not the way to go, said Neylon. This will unavoidably also filter out useful publications. The solution lies in efficient discovery systems, through networks. Networks will help scholars to become aware of important news and new scientific findings.

Martin Rasmussen, Copernicus Publications, Germany: *"Ensuring Availability and Quality of*

Research Data through Open Access and Public Peer-Review"

Another invited keynote speaker was Martin Rasmussen, who is director of Copernicus Publications, located in Göttingen, Germany. Rasmussen talked about the scientific and economic value of research data which is enormous. In order to realize a successful subsequent usage, the scientific community needs efficient access to data. This access has to be reliable and persistent. In addition, the quality of the data has to be proved following standardized work flows. One solution to these preconditions is to apply the techniques of today's scientific publishing to research data. Besides the access to the data sets through data repositories together with some metadata, the data should undergo a transparent peer-review using a publication platform.

Rasmussen discussed two approaches: On the one hand, the data can be the basis for a research article and undergo a review parallel to the review of the manuscript. Data may preferably be stored in external dedicated data repositories, enabled to long time preservation of big data sets. These data may then be reviewed, and made available as a supplement to the scientific paper. On the other hand, the data itself can be the subject of a publication whose quality is then assured by peers. This is what Copernicus Publications do in their journal Earth System Science Data (ESSD), using an interactive public peer-review process.

Eelco Ferwerda, OAPEN Foundation, the Netherlands: *"Open Access for books - OAPEN pilot projects"*

OAPEN has developed from an initial project, and is now an independent foundation, based in the Netherlands, forming an international network of publishers and research institutions publishing monographs Open Access. OAPEN thus forms a library of freely available books in humanities and social sciences in various European languages. Per date, the OAPEN library includes more than 30 publishers,

and totaling more than 1000 books. A "Directory of Open Access Books" will soon be launched.

The OAPEN model is a hybrid model, meaning that the books may be available both online and in print, or print on demand. And publishers may operate commercially, by selling print copies and charging publication fees from the authors (or their funders). Calculations show, that publishing a book as open access costs approximately half of what it costs to publish a printed book.

OAPEN is dependent on funders to finance the OA publishing of books. OAPEN introduce the model to new areas and new publishers and funders through pilot projects. OAPEN is now hoping to introduce a pilot project in the Nordic countries, in order to include publishers from these countries in the OAPEN network of publishers. Ferwerda hopes, that the time may be right for this model in the Nordic countries, since these countries to a large extent already have a system of funding support for scientific monographs.

And more
Several other speakers gave highly interesting presentations during the two days. All these presentations may be viewed through the videos made available at the conference website: http://www.ub.uit.no/MC6

As Tromsø is located at almost 70 degrees north, the conference took place at the start of the two month polar night period in Tromsø. Unfortunately the weather did not allow participants to enjoy the spectacular northern lights, nor much of the beautiful blue daylight characteristic for this time of the year. Maybe next time.

Leif Longva University of Tromsø, Munin repository manager

OER – AN INTERNATIONAL TREND WITH SLOW DEVELOPMENT IN SWEDEN

Lasse Borelius, Alastair Creelman, Åsa Forsberg

Open Educational Resources, OER, are net-based resources, free to use and in many cases also free to remix and tweak. The global OER movement is very important. Development in Sweden however has been rather slow both with regard to open access publishing and sharing learning resources.

Within Swedish higher education there have been several notable initiatives by individual enthusiasts or pressure groups. University libraries are well aware of OER as a phenomenon and as a development based on principles closely related to Open Access and Open Source.

Internationally the OER movement is growing within all sectors of education, from K-12 to higher education. Some of the factors behind this development are the fact that education programs are increasingly globalised and that informal learning is valued and evaluated together with formal learning. Also in higher education general skills are growing in importance; in particular so-called 21st century skills such as information literacy, communication skills, digital literacy etc.

Important OER projects have been undertaken by large international organisations such as UNESCO and OECD. In July 2011 the organisation Commonwealth of Learning (COL) & Unesco published the Open Access publication "A Basic Guide to Open Educational Resources (OER)"*[1]. Unesco is in the process of revising their guidelines and arranged in the autumn of 2011 an online workshop to discuss the proposal of new guidelines, "Guidelines for OER in Higher Education".

The nations who so far have been more successful in integrating OER in education (t ex US, Australia, The Netherlands and the UK) have governmental authorities and other bodies that finance projects,

elaborate clear guidelines och promote OER in the country's academic community as a national priority.

A clear tendency today is that learning is more and more taking place in networks and digital social environments. In schools and universities OER is also becoming a formal resource. By using OER the traditional course content can be enriched with different perspectives and clarifications. OER offer more opportunities to collaborate between academic disciplines on a national and international level and teachers can gain both time and quality. Classroom time can be used for in-depth discussion and tutoring of pupils and students.

The use of open educational resources is not just a question of content and material. The use of OER is strongly related to new ways of viewing education and learning in general; and specifically to connectivism, a theory developed by Stephen Downes and George Siemens amongst others. The teacher's role as knowledge mediator is no longer valid since information and content are available everywhere by Internet.

Learning is taking place in networks where students collaborate on assignments and exchange experience. The teacher's role will rather be to be a mentor/tutor/inspirator/coordinator and the teacher's production of course material will diminish accordingly.

Open educational resources can also be both an incentive and a support to an increasing consciousness of the importance of evaluating sources. It is becoming increasingly important for both teachers and students to able to acquire a critical attitude towards Internet resources such as OER, a current example is the video Kony 2012[2]. OER can also be used as a resource in the process of acquiring a critical attitude to

---

[1] Available online
http://www.col.org/resources/publications/Pages/detail.aspx?PID=357

[2] Available online
http://www.youtube.com/watch?v=Y4MnpzG5Sqc

information sources, for both teachers and students in secondary and tertiary education.

An important component in this development is that OERs should be tagged with metadata. Many teachers prefer to be able to search OERs according to criteria such as previous knowledge required, level, learning outcomes, main target group etc. Many enthusiastic teachers lose interest when they realize how difficult it is to make targeted searches for material.

Another important component is the need to spread knowledge about the use of Creative Commons as a complement to traditional copyright law.

To mark up the learning resources according to metadata standards and to supply a Creative Commons licence clearly stating how the material can be used are both important when ensuring that the resources can be used in a sustainable way. But they are also means of retrieving relevant material and material that can be remixed and tweaked in order to either increase the quality, to adapt it to the needs of a specific target group or to update it according to new scientific findings.

The digital development in society influences all sectors and professional roles. Factors like accessibility, mobility, interactivity and individual adaptation facilitate communication, collaboration and creativity. Students must learn to use the full potential of the net as an integral part of their education in order to acquire relevant skills for their professional lives.

Unfortunately a conservative attitude towards ICT resides still in further education and the gap between an increasingly digital economy and a basically analogue educational sector is widening.

The labour market of the future will require quality and innovation, independance and entrepreneurship, and the challenge for higher education is to create an environment that stimulates such characteristics.

Therefore it is extremely important that the Swedish universities encourage innovation and new thinking in teaching and learning.

A further problem for Sweden is the concept that teachers own the right to their material rather than the university (known as the *teacher exception* or in Swedish *lärarundantag*), and clarity in this issue demands attention from university leaders and authorities.

OER is not a separate question, OER must be seen as a integrated in a much more extensive process of change in higher education.

Individual universities can develop strategies for OER but it is necessary to have national coordination to minimise unnecessary duplication and the creation of parallel processes. Just as with Open Access it is necessary to have directives and strategic initiatives in order to change attitudes. In Sweden OER is still a grassroots movement, and without the full commitment of university managements and government authorities the use of OER will be fragmented and ad hoc.

A workgroup in the network ITHU (IT in Higher Education) has carried out projects to spread knowledge about OER in higher education. In order to move further on it is now necessary that main bodies such as the National Library of Sweden, the KK-foundation, HSV, SULF and SUHF now take initiatives and provide funding. The initiatives should not be limited to the knowledge about OER, it is crucial to also discuss questions such as the right to use learning resources versus the "teacher exception" and find a solution together with the teacher unions. A national Swedish OER network, including above mentioned organisations and the universities, could pursue these issues on a wide front!

Lasse Bourelius Blekinge Institute of Technology, Head of Unit Education Development, LearningLab and CareerCenter

Alastair Creelman Linnaeus University, Sweden,

Åsa Forsberg Lund University, Sweden

OPEN ACCESS – LOOKING TEN YEARS BACK AND A FEW YEARS AHEAD
Jan Hagerlid

I have been invited by the editor to reflect on the development of open access over the latest ten years. When writing this I am about to retire from active service at the National Library of Sweden and from my position as coordinator of the OpenAccess.se programme. The occasion makes it tempting to look back and try to draw some conclusions.

Ten years back
My own interest in the open access issue started already in the middle of the 90-ies but became stronger around 2002. There were a number of breakthroughs during just a few years. We saw the three B-declarations - Bethesda, Budapest and Berlin - and the Open Letter from Public Library of Science. Interoperability was taken to a new level with the Open Access Initiative Protocol for Metadata Harvesting. The number of open access journals grew fast and commercial players entered the field. A Swedish contribution was the Directory of Open Access Journals at Lund University which was being developed from 2002, mainly with the support of the Open Society Institute but also from the National Library of Sweden.

I was invited to write a chronicle for a Swedish research journal (Dagens forskning) in 2002. I wrote that "it is strange that universities still accept the rules of the game for the publication of scientific journals. First universities pay the salaries of their researchers as authors, reviewers and editors of these journals. No compensation is given from the publishers. Then universities via their libraries pay soaring subscription prices to get access to the same journals." In dialogue with the editors the chronicle was given a flashy title, "The system should have perished by itself ". My analysis at the time was that the publishing system was so absurd and irrational that it had to break down soon, now that a viable alternative was being built. The PloS Open Letter had brought the discussion to a much wider audience, in mass media and in journals like Science and Nature. Somewhat later the Berlin declaration on Open Access Knowledge in the Sciences and Humanities testified to the intention of major

organizations within research in Europe to work for open access.

Too optimistic?
But the system did not break down in the years thereafter. We did not see a quick transition to an open access model. So had I got it wrong? When you are deeply involved in a process and you have realized how irrational the present system is, it is easy to underestimate the inertia of social practices and structures established during a long period of time. Science has among its basic values not to accept something new until it is solidly proven. The tight connection between the present scientific journals and the model for career advancement is a very strong conservative force. Authors don't lose any income in an open access model but they might feel they risk their careers. It is interesting to compare with the music industry. Here the free flows on the web is directly threatening the very livelihood of the creators, but still we now see a rapid transition to a new model (Spotify etc) where costs are covered in a new way and access is, if not completely free, yet immensely widened for users.

Look at the bright side
But let's look at the bright side; the glass is rather half full than half empty. Changes were not as rapid as I had expected but they have still been very substantial, especially when seen in a ten year's perspective.

Open access has moved from a fringe discussion in smaller circles of researchers, librarians and publishers to a level where the model in principle has the support from practically all major stakeholders within research; universities, research funders and to a degree also governments. In Europe the EU early took a positive interest and gradually has strengthened its support. Publishers have moved from an attitude mixing aggressive opposition and downgrading of the importance of open access to a stance where they generally take the issue seriously and in many cases actively take part in open access developments. It is hard to avoid the conclusion that the political battle is

won! It seems very unlikely that we would see a general retreat from these positions.

The developments on the ground are lagging somewhat behind but all growth figures are impressive.[1] [2] [3] It is not a question whether we will get a transition to a state where the open access model dominates, but only (no small questions!) how **fast** it will come, what **means** are most effective and what **shape** it will take.

In the new strategy[4] of the OpenAccess.se we write: "The conditions constantly improve for reaching a breakpoint in the next couple of years, when more than half of the yearly production of Swedish, publicly funded research publications will be freely available. After that everything points to an even faster development towards open access." This is not just wishful thinking. There are numerous examples of technological and social developments where you get this acceleration when passing a breakpoint.

What means?

What means are most effective? I think most people would agree that we need things like the following:

- Clear and coordinated open access policies from Higher Education Institutions (HEIs), research funders and governments,
- Reliable and well presented information to researchers and other stakeholders about open access
- An infrastructure with user-friendly and efficient services to researchers, including repositories at HEIs, national OA journals,

national search service, and data centres etc
- Economic agreements and solutions that facilitates publishing in OA journals, or at least create a"level playing field"

Should we choose between Green and Gold?

However, there is some disagreement on whether we should stress the "Gold road" or the "Green road". The strategy of the OpenAccess.se is even-handed on the issue. "The Programme shall support both publishing in open access journals and parallel publishing in open archives. For a foreseeable future these two roads will run parallel." I find this the only sensible position at the present stage. A strategy has to be based on a realistic appraisal of the facts. Open access journals are growing rapidly in numbers and in volume of articles as shown by Laakso et al.[5] But still, according to their estimates the share of OA articles of all articles published in peer review journals in 2009 was only 7.7 %. In another article by the same group of researchers the total OA availability of articles in 2009 was studied.[6] The total OA availability was 20.4 % of which 8.5 % was Gold and 11.9 % was Green. What more, they show very clearly how the authors preferences for Gold or Green differs widely between major subject areas. Researchers in Medicine and Biology lean heavy towards Gold whereas researchers in all other field prefer Green. These are studies with a carefully developed and well presented methodology. It is likely that the general OA share should have risen at an even quicker pace since 2009, but not that the shares of Green and Gold should have changed dramatically. This holds even though the OA journals probably have got some extra boost from the success of mega journals like PloS ONE with followers.

The only possible conclusion is that a successful open access strategy must base itself on a combination of Green and Gold. If the focus earlier was on Green the balance today should be more even, but not lean over to fast to Gold. Of course the Green road is a transitional model in the sense that it is dependent on articles first being published in traditional Toll Access journals. But it contributes to create a pressure on traditional publishers to change. Why otherwise would a publisher like Elsevier put so much energy into

[1] Björk B-C, Welling P, Laakso M, Majlender P, Hedlund T, et al. (2010), Open Access to the Scientific Journal Literature: Situation 2009. PLoS ONE 5(6), www.plosone.org/article/info:doi/10.1371/journal.pone.0011273

[2] Laakso M, Welling P, Bukvova H, Nyman L, Björk B-C, et al. (2011), The Development of Open Access Journal Publishing from 1993 to 2009. PLoS ONE 6(6): http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0020961

[3] Morrison, H. Dramatic Growth of Open Access Series. The Imaginary Journal of Poetic Economics. http://poeticeconomics.blogspot.com/2006/08/dramatic-growth-of-open-access-series.html

[4] Strategy for the OpenAccess.se programme 2011- 2013 http://www.kb.se/dokument/Om/projekt/open_access/2011/OA-strategi_eng_final.pdf

[5] Laakso et al (2011), ibid.

[6] Björk et al (2010), ibid.

blocking institutional open access mandates?[7] Let us be clear that there can be no institutional or funder OA mandates without the Green road. Finally, repositories at Higher Education Institutions already have many different functions and could certainly evolve further.

The shape of things to come
So much for the pace of changes, what about the shape? The challenge is to manage a transition to an Open Access model, with full regard to other important demands that can and should be addressed to the scholarly publication system:

- It must guarantee and stimulate a continuous increase of scientific quality
- It must be economically sustainable for authors, universities and funders
- It must be efficient and flexible for both authors and users
- It should guarantee long term access to publications (and data)

The open access movement has a strong focus on the access issue and rightly so. But now when we no longer dream about a distant goal, but rather explore the practicalities of a transition to a model for scientific communication where open access is taken for granted, we have to discuss all the other aspects of the system we would like to see.

Open access in Sweden
I will make a few comments on the way work for open access has evolved in Sweden. For a more detailed story I refer to a few earlier articles on Swedish open access developments.[8] [9]

---

[7]Hagerlid, J. (2012), Elsevier tries to block institutional OA mandates. Open access i Sverige. http://openaccess.kb.se/?p=637

[8] Hagerlid, J. (2006), Open Access in Sweden 2002-2005, ELPUB2006. Digital Spectrum: Integrating Technology and Culture - Proceedings of the 10th International Conference on Electronic Publishing held in Bansko, Bulgaria 14-16 June 2006 / Edited by: Bob Martens, Milena Dobreva. ISBN 978-954-16-0040-5, 2006, pp. 135-144. http://www.informatik.uni-trier.de/~ley/db/conf/elpub/elpub2006.html

[9] Hagerlid,J. (2011), The role of the national library as a catalyst for an open access agenda: the experience in Sweden", Interlending & Document Supply, Vol. 39 Iss: 2, pp.115 - 118 http://www.emeraldinsight.com/journals.htm?issn=0264-1615&volume=39&issue=2

The Swedish universities individually and collectively through the Association of Swedish Higher Education (SUHF) have expressed a principal support for open access, starting already in 2004 when SUHF signed the Berlin Declaration. The Swedish Research Council signed the Berlin Declaration in 2005 and adopted an OA mandate in 2009. So far five other research funders - governmental and foundations - have adopted OA mandates. Of the four governmental research funders only one, VINNOVA, has not as yet decided on an OA mandate. The situation would seem ripe for the government to adopt a national open access policy and hopefully this will come as a part of the Research bill that will be presented to the parliament in autumn 2012.

Top-down or bottom-up
This is all very good, but there is a danger that the process is too much top-down. Some researchers might feel that open access is just another bureaucratic imposition that only serves to hinder their research and make life more complicated. This kind of reaction came clearly across in a petition from 83 Swedish chemists in autumn 2010 who strongly criticized the SRC open access mandate.[10]

On the other hand we get reports from universities that many researchers appreciate their repositories, especially when they can see the increased impact of their own work. There are also impressive growth figures for Swedish open access journals. The number seems to have grown every time I check in DOAJ, today (8 of March 2012) it is 54.[11] These researchers, who are positive to open access, are not publicly visible in the same way as some of the opponents, like the chemists in the petition.

The OpenAccess.se has put a lot of energy in promoting strong open access policies at universities and research funders. Could we also find new ways to support bottom-up open access initiatives? One way would be to organize workshops for researchers active in new or potential open access journals. Another might be to find ways to connect to the growing number of science bloggers. When you blog about

---

[10] Hagerlid, J. (2011), Vetenskapsrådet förtydligar sina open access-krav. Open access i Sverige. http://openaccess.kb.se/?p=200

[11] Directory of Open Access Journals. http://www.doaj.org/doaj?func=findJournals&uiLanguage=en&hybrid=&query=Sweden

your research it would be an advantage to be able to link to works that are open access.

Perhaps this is essentially a question about the way the Swedish open access movement defines itself. When we organize meetings most of the people attending are involved in repositories at universities. We see few, if any, editors of Swedish open access journals or, for that matter, Swedish researchers involved in international open access journals.

The achievements and challenges of repositories

A major achievement during these ten years is that e-publishing at universities have grown from scattered initiatives to a stable network of repositories covering practically all Higher Education Institutions and having their content harvested to the national SwePub service. Repositories from the start only held open access material but are now integrated with publication databases containing metadata for the total (in principle) research output of a given institution. Increasingly these repositories/publication databases bring several different kinds of benefits to their host institutions: visibility, profiling, impact, evaluation and reporting. I suspect that evaluation and reporting always will be highest on the agenda of university leaders and thus define institutional priorities. This makes it even more important that repository managers, libraries and researchers instead stress access, impact and visibility. If these sometimes contradictory ambitions are handled wisely, the wide role of repositories might give them a very strong position within their host institutions.

Repositories could develop in a number of directions. Should they create functions that usually characterize university presses, like quality selection and marketing, when engaging in publishing of monographs and journals? Access to research data has attracted a rising interest during the last few years. Should repositories link to research data in a systematic way and should they also host (some) research data? Can Open Educational Resources, that are constantly evolving, be handled in a meaningful way by repositories? Will we see for real a development where authors first publish in repositories but the quality selection then takes place in international research portals, overlay journals or whatever you like to call them. There seems to be no lack of exciting challenges and opportunities for repositories.

Jan Hagerlid, Senior executive officer at the Department for National Cooperation at the National Library of Sweden from 1992. Chief librarian at the Swedish University of Agricultural Sciences 1997-2002. From 2002 back at the National Library of Sweden, working mainly with e-publishing and open access. Coordinator of the OpenAccess.se programme 2006-. Involved in EU project OpenAIRE, in Confederation of Open Access Repositories (COAR), in the NORDBIB programme and in IFLA working Group for Open Access.

SHARING RESEARCH DATA AT THE NORDIC LEVEL
- A vision for the future Nordic cooperation from Finnish perspective

Marjut Salokannel

Current situation in Finland

In Finland open government data and data produced by public funding has been on the foreground during the past years. At the governmental level there have been several working groups dealing with these issues, one of which has specifically concentrated on opening-up publicly funded data for research purposes. A continuation of this work is carried on by a project of the Ministry of Education and Culture.[1]

At the international level research funders around the world are requiring that the results of the research they have funded have to be publicly accessible. This refers both to the publications and the underlying data. In Finland the national research funding agency, the Academy of Finland, is not there yet, but is merely recommending that research articles be openly available. The Academy does require that the application for funding includes a data management plan. When we look at the general picture with regard to the availability of research data in Finland there are big differences depending on the field of research. In natural sciences and biomedicine open access to research data is the norm and the data is being deposited in the international depositories whereas in humanities and social sciences, with the exclusion of economics, open research data remains still a rarity.

Data sharing at the Nordic level

Even if the general approach with regard to opening up research data follows the internationally established practices of each research field, there are some issues that are specific to the Nordic countries. One of these issues relates to the vast amounts of register data contained in different population based registers which are almost identical in all the Nordic countries. These registers are an immense source of raw data for epidemiological research, population based research,

economics and other fields of research where combining and analysing microdata from different sources forms the basis of research. Similarly datamining different sources of data in bioinformatics and genomics gives us new knowledge of diseases and basis for developing therapies.[2]

Unfortunately, these data reserves are currently being underutilized, since using the microdata contained in these registers for research purposes poses constantly difficulties, in particular with regard to combining data from different sources. In Finland the most difficult bottleneck has been the data stored within the National Statistics Authority.

In the following we shall discuss the core legal issues in respect to accessing research data, that is the extensive copyright protection and data protection law. Finally we raise the question of how it would be possible to establish, at the Nordic level, a secure federated environment for a seamless use of register data from different sources irrespective of the place of residence of the researcher, and a way to combine this data with other data, such as data in biobanks or patient records.

Copyright issues with regard to using data for research purposes

The Nordic copyright laws have a common history dating from the beginning of 1960's. Since then there has been some minor variations in the national laws but the basic principles remain very similar. One of these basic issues is the extensive protection afforded by the law to all sorts of collections of data resulting from a combination of the EU *sui generis* database right and the Nordic catalogue rule which extends the protection also to those databases that do not fall under the copyright protection or *sui generis* database protection. From researcher's perspective the situation

---

[1] Exploitation of Public Sector Information" -working group in the Ministry of Traffic and Transport, two working groups in the Ministry of Finance ("Interfaces of public sector basic information resources" and "More efficient use of public sector information resources"), a Ministry of Finance
working group for amending the Statistics Act and a cross-sectoral steering group by the Ministry of Education and Culture "Information resources for Research".

[2] See also "Sharing registry data for health research in the Nordic countries – a proposal for increased collaboration ", Report from the Nordic Task Force for Access to national data repositories, available at
http://ki.se/content/1/c6/13/68/01/Sharing_Registry_Data_for_H ealth_Research_in_the_Nordic_Countries.pdf (last accessed 14th June 2012).

is aggravated by the fact that, unlike in most other countries in the world, there exists no provision in the law, which would provide legal security for research uses of copyrighted materials. For example, the Finnish Copyright Act does not contain an exception in the law, which would make it possible to use protected materials, including databases or catalogues, for research purposes. This sets Finnish researchers in an unequal position in relation to their foreign counter parts and makes it difficult for researchers residing in Finland to take part in international collaborations.[3]

Digital technology has made it possible to use data produced by public funding by the government and research institutions in ways that have not been previously possible. This has even given birth to new fields of scientific research such as bioinformatics and biostatistics, which use data mining and analytics as their principal tool for creating new knowledge out of existing data. In this respect the Nordic countries enjoy an especially advantageous position in form of their over century old population based registries which are built on the same basis in all the Nordic countries.

Data mining and text mining in particular have proven to be problematic in terms of the copyright law. In the United States the *fair use* doctrine makes it possible to use data mining techniques for research purposes but in Europe this is not necessarily the case. In particular in countries, such as Finland, where there is no statutory exception providing for research use of copyrighted material, undertaking data and text mining currently call for careful legal analysis if future legal problems are to be avoided.

Data protection and data sharing

Data mining raises also questions relating to possible data protection issues if it makes it possible to identify persons by combining data from different sources, such as public population based registers. The EU Commission has tackled this issue in its proposal for a general data protection regulation in which it subjects processing of data for research purposes under specific rules.[4] The proposed regulation makes it possible to create a cross-border framework for data sharing in e.g. register and biomedical research, by requiring the de-identification of individual data when it is used for research purposes. Data which have been de-identified may be used for research purposes in terms of the

regulation also without the consent of the data subjects.

The proposed regulation permits processing of personal data for statistical and scientific research purposes

1) with the consent of the data subject for one or several specific purposes;
2) processing is necessary for the purposes of historical, statistical or scientific research under the following conditions:

   a) these purposes cannot be otherwise fulfilled by processing data which does not permit or not any longer permit the identification of the data subject; and
   b) data enabling the attribution of information to an identified or identifiable data subject is kept separately from the other information as long as these purposes can be fulfilled in this manner. (Article 83)

In addition, the regulation provides that when the data processed by a controller do not permit the controller to identify a natural person, the controller shall not be obliged to acquire additional information in order to identify the data subject for the sole purpose of complying with any provision of the Regulation. (Article 10) This provision provides relief for researchers and research institutions acting as data controllers when research is being conducted with de-identified data.

Future vision for data sharing in register based research at the Nordic level

For researchers it is of paramount importance to be able to combine data from different sources, whether it is from different registers, biobanks, or patient records. Researchers also want to pool this data with their colleagues in other countries and use it in collaborative research projects. The best way to protect the privacy of the persons whose information is contained in such data is to anonymize the data. With the advances of technology complete anonymization is becoming increasingly difficult and we shall have to use rough estimates with regard to the probability of identification of data subjects. Moreover, in scientific research there may be exceptional instances when the identification of a certain person could be necessary.

If register data is encrypted in a way that it does not permit identification of data subjects but allows the decrypting of the codifiers in certain cases

---

[3] *Marjut Salokannel,* Using Copyrighted Works for Research Purposes, IPRInfo 2/2009, available at http://www.iprinfo.com/lehti?action=articleDetails&a_id=662&id=43 (last accessed 14th June 2012).
[4] Proposal for a regulation of the European parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, COM(2012) 11 final, 25.1.2012.

(pseudonymisation), this would allow linking of personal register data with data in different registers without compromising the privacy of data subjects. The Finnish data protection ombudsman has proposed of instituting a specific personal research identifier for each person through which different data may be combined from different sources. The identifier would replace the social security number. The next step for Nordic purposes would be the creation of Nordic research identification number which could be used throughout the Nordic countries thus abolishing the need for national identifiers.

When visioning future Nordic cooperation in data sharing we must keep in mind that Denmark and Sweden already have well functioning on-line environments for using register data for research purposes. From researcher's point of view it is, however imperative that data can be combined from different registers with, e.g., molecular data and data in biobanks, and used in collaboration with other researchers across national borders, which is not possible with closed on-line systems. Cross-border data-sharing is, however, precisely what the EU data protection regulation aims to achieve. Sharing encrypted or pseudonymized data for research purposes is an answer to this question. At the Nordic level, this principle could be applied with regard to register data when combined with common anonymization service which could provide the data subjects with specific research identification codes which would replace the national social security numbers.[5]

From the researchers' perspective what is needed is the possibility of mining, analyzing and combining data from different sources, including from different countries, and to be able to exchange and collaborate in international research settings while using this data. We should be able to make available the Nordic register based data for research purposes in such a way that it is possible to link and combine the microdata with microdata in other registers across borders and, ultimately, with data residing in biobanks or even with patient records in such a way that it does not compromise the privacy of individuals whose data is stored in these registers. How this can be achieved in legal, technical and practical terms requires, however, further study.

Marjut Salokannel is LL.D. Docent Helsinki Collegium for Advanced Studies

---

[5] See also the report commissioned by the Nordic Council of Ministers *Sandberg, M.*, Reinforced Nordic collaboration on data resources, TemaNord 2012:514.