Creating and Curating the Cognitive Commons: Southampton’s Contribution

Les Carr, Alma Swan & Stevan Harnad
School of Electronics and Computer Science
University of Southampton
Highfield, Southampton
SO17 1BJ UNITED KINGDOM
http://www.ecs.soton.ac.uk/

Four Revolutions in the Means of Production of Knowledge

Cognition is thinking, and the product of thinking is knowledge. Our species’ unique capacity for language empowered us to “mind-read” one another’s thoughts, creating and curating knowledge jointly through distributed, interactive cognition. The first and greatest cognitive revolution – the birth of language itself – took place about 300,000 years ago and created the Oral Tradition. Writing was the second revolution, 6000 years ago, creating the Written Record; print was the third, 600 years ago, creating the Published Archive. The fourth cognitive revolution is ongoing in our own era. With the invention of the Internet and the Web, humankind is on the verge of creating a Cognitive Commons – a global collaborative medium in which knowledge can be created and communicated at the speed of thought.

When language first evolved, it was its reciprocity and mutuality that gave it all its power, and that so dramatically hastened the permanent encoding of our language capacity into our genomes and brains: Language is fundamentally distributed and interactive. It would never have evolved if it had not been beneficial to our species survival and reproduction to share our thoughts. This is why language has been called a form of “reciprocal altruism,” acquired because it is also in the interests of our selfish genes. At its origin, thoughts were shared freely – within the family and within the tribe – to everyone’s benefit, and passed on from generation to generation through the oral tradition. Writing made it possible to preserve thoughts speaker-independently (“verba yolant, scripta manent”), but it was largely print that made thoughts – in the form of printed words -- into a commodity that could be bought and sold. There had of course already been secrets, including trade secrets, before print, but print made it possible – indeed necessary – to seek payment in exchange for words, in order to cover the considerable cost of disseminating the written word.

The possibility of charging “a penny for one’s thoughts” was not a bad development: it was a natural part of the cultural evolution toward specialization and division of labor, with the buying and selling of distributed commodities and services, rather than everyone’s having to be a self-sufficient master of all survival skills and goods. But there
were two negative side-effects of the writing and printing era: The first was in the domain of speed. Our brains were biologically adapted specifically to spoken (and gestured) language: the speed of thought is approximately the speed of speech, and it is highly interactive. Handwriting preserved a permanent record, but at the sacrifice of the interactivity of real-time dialogue. Print and mail increased the scope and reach of the written word, but its turn-around time was still vastly out of phase with the potential speed of thought. It is only with the possibility of email and instantaneous networked interaction – distributed and global “skywriting” – that cognition has regained its native potential to operate at the speed of thought for which our brains were optimised.

So the online era has overcome the first negative side-effect of the writing and print era: cognitive turn-around time. It is now on the verge of overcoming the second one, which concerns the access to – rather than the timing of – cognition: Both the illuminated manuscript era and the Gutenberg print era conferred the benefits of the new lapidary media at a cost: Handwriting had cost individual scribes considerable time and effort; and the production and distribution of printed work was also an expensive process that had to recover its costs by charging for access if it was to make ends meet. This constraint was not a problem for trade publication: Those who wrote in order to make their living through the sale of their words were well served by the natural access-barriers and access-tolls of the Gutenberg era. But there had always been special exceptions too: Many scholars and most scientists do not write in order to sell their words: they write in order to communicate their knowledge, reporting their research findings so that other scholars and scientists can access, use, apply and build upon them, in the reciprocal, cumulative and collaborative enterprise of learned inquiry. Yet these nontrade authors, who write for research impact rather than imprint income had to reconcile with the limitations on the scope, reach, uptake and impact of their findings that were necessarily imposed on all authors – trade and nontrade alike – by the real technological and economic constraints of the Gutenberg era.

In our PostGutenberg medium of online communication it is at last possible for this special nontrade authorship – scholarly and scientific researchers – to free their writings from the access-toll barriers that had existed since the beginning of the print era.

Open Access

Open Access (OA) means free online access. What made Open Access possible was the advent of the networked online medium: The Internet, and eventually the Web, empowered the authors of digital works to give them away free for all online if they wished.

The term “Open Access” was first coined by the Budapest Open Access Initiative (BOAI), sponsored by the Open Society Institute (OSI) in 2001. But the idea of providing free online access – and the provision of free online access – started much earlier than the BOAI and the adoption of the name “OA.” The inventors of Unix and the Internet – mostly computer scientists – had already been providing OA to their research papers by self-archiving them in anonymous FTP archives” since at least the 1970s. With the invention of the Web in 1990, websites soon became the preferred way of self-archiving papers. High energy physicists – who had already been systematically sharing their
papers on paper before the Internet, and then via email when it became possible – began self-archiving them in Arxiv, a centralised physics web archive, in 1991. Many individuals from other disciplines have since followed the lead of the computer scientists and the physicists.

A “Subversive Proposal” to make all refereed journal articles free for all by self-archiving them online originated from the University of Southampton in 1994. The proposal also identified the way to cover the cost of publication if OA self-archiving eventually made subscriptions unsustainable: fees for publishing individual articles instead of subscription fees for accessing them. The first OA journals began appearing in 1989; most were either the online versions of subscription journals or they were subsidised online-only journals.

Meanwhile, the 1994 Subversive Proposal to self-archive went largely unheeded: For the following decade, the rate of author self-archiving continued to hover at about 15-20% of yearly refereed research output. The proportion of articles published in OA journals was even lower. Providing centralized archives like Arxiv for other disciplines (e.g., CogPrints for the Cognitive Sciences, created and hosted by University of Southampton in 1997) likewise failed to increase the rate of OA self-archiving. Nor did the creation of the American Scientist Open Access Forum in 1998 – hosted by the Sigma Xi Scientific Research Society and moderated from the University of Southampton.

In 1999, the Open Archives Initiative (OAI) developed a metadata-tagging protocol for the purpose of making all Open Archives “interoperable.” This means that depositing locally in any individual OAI-compliant archive would be equivalent to depositing centrally in one global, seamlessly searchable Open Archive. In 2000, free software (EPrints) was designed at the University of Southampton (by adapting the CogPrints software to make it OAI-compliant and generic) to make it possible for all universities to create their own OAI-compliant Open Archives (which soon came to be called, instead, “Institutional Repositories” [IRs]). Many IRs were subsequently created, worldwide – their growth has been monitored by the University of Southampton’s Registry of Open Access Repositories (ROAR) since 2001 -- but IRs remained near-empty because 85% of researchers still were not self-archiving.

In 2001, Steve Lawrence published a paper in Nature reporting that OA articles in computer science are cited significantly more than non-OA articles. Many subsequent follow-up studies – collected in a growing bibliography of “The effect of open access and downloads (‘hits’) on citation impact” created and hosted by University of Southampton -- confirmed that this “OA impact advantage” was also present in every other scholarly and scientific field tested. But even the OA advantage was not sufficient to induce the 85% of non-self-archiving authors to do so.

**Mandating Open Access**

It had already been proposed since 1998 in the American Scientist Open Access Forum that universities and research funders should mandate OA self-archiving (i.e., make it a requirement, as a natural extension of the requirement to publish-or-perish). Having first proposed it, the School of Electronics and Computer Science at Southampton University
(UK) also went on to become the first in the world to adopt an OA self-archiving mandate, in 2002, and offered it as a model for other other institutions in the OSI Eprints Handbook in 2003.

The University of Southampton has provided and hosted a Registry of Open Access Materials Archiving Policies (ROARMAP) for registering and tracking the growth of OA mandates since 2003. The first university-wide OA mandate was adopted by Queensland University of Technology (Australia) and the first European university-wide mandate by University of Minho, (Portugal), both in 2004.

Likewise in 2004, the UK Parliamentary Select Committee on Science and Technology recommended that universities and research funders should mandate OA – again at the urging of University of Southampton. In the same year, the US House Appropriations Committee recommended that NIH should mandate OA.

The Wellcome Trust became the first research funder in the world to mandate OA in 2005. In the same year, NIH adopted an OA request instead of a mandate; that policy failed and was upgraded to a mandate in 2007. The UK government failed to act on the Committee’s recommendation, yet within a few years all seven of the UK Research Councils nevertheless followed it, each adopting a self-archiving mandate of its own.

A further incentive to mandate and provide OA was provided by the fact that the outcome of the UK Research Assessment Exercise -- in which peers review and rank the research publications of all departments of all UK universities every six years -- turns out to be highly correlated with the citation metrics that OA has been shown to increase. The University of Southampton has been strongly promoting the development of OA metrics to track, evaluate and reward research usage and impact, creating Citebase as a model for a scientometric engine for research evaluation and navigation and IRStats for gathering IR usage metrics.

Two international, cross-disciplinary author surveys in 2005 reported the most fundamental strategic and practical finding about why OA growth had been so slow: Although most authors do not self-archive, over 90% of them indicate that they would self-archive if their funders or institutions mandated it – over 80% of them indicating they would do so willingly. Outcome studies from Arthur Sale in Australia have since confirmed that within two years of mandate adoption, compliance rates are indeed over 60% and well on the road toward 100%. ROARMAP shows that the number of mandates is approaching 200 worldwide and now includes Harvard, MIT, UCL and ETH Zuerich, as well as the European Research Council and the European Commission.

Green and Gold Open Access

OA self-archiving has come to be called the “green” road to OA (or “Green OA”), to distinguish it from OA journal publishing, which is called the “gold” road to OA (“Gold OA”). The most frequent misconception about OA is that OA only means Gold OA (publishing). In fact, the fastest and surest road to OA is the green road of OA self-archiving, for two fundamental reasons: (1) providing green OA is entirely in the hands (and interests) of the providers of the research itself, the global research community, and
(2) green OA can be mandated -- whereas gold OA is in the hands of the publishing community and cannot be mandated.

Hence green OA needs to come first, and it needs to be universally mandated, by institutions as well as funders. It has been a great strategic mistake to wait instead for Gold OA. If, despite all the benefits, most authors are not providing green OA spontaneously of their own accord, at no cost, and without having to abandon their journal of choice, then they certainly will not provide gold OA, for an additional cost, and having to publish in a gold OA journal instead of their journal of choice. Nor will their institutions have the money to pay their authors’ gold OA publishing costs while those funds are still tied up in paying for journal subscriptions. Nor can institutional journal subscriptions be cancelled while the journals’ contents are still not otherwise accessible to the institution’s users. Moreover, the asking price for gold OA publishing is still much higher than it needs to be, while journals are still producing print and online editions.

If universal green OA mandates are adopted first, then, if and when the resulting universal green OA makes subscriptions unsustainable as the means of covering publishing costs (because institutions cancel their journal subscriptions), the natural effect will be to induce journal publishers to cut costs, downsize, and convert to gold OA; and then the self-same annual windfall savings from the institutional cancellations will be available to pay the institutional authors’ costs, per article, of gold OA publishing. Those costs per article will, however, be substantially lower after universal green OA has made subscriptions unsustainable, because journals will no longer need to provide the print or online edition: All access-provision and archiving will have been offloaded onto the distributed network of green OA IRs. Journals will only provide the service of peer review, and the institutional savings will be more than enough to cover its costs.

If, rather than mandating green OA first and waiting for green OA mandates to propagate globally and to have their natural effects, institutions instead commit some of their scarce available funds to paying pre-emptively for gold OA -- and at the current asking price -- they will get very little OA in exchange for their money and they will reinforce gold OA publishing’s current asking price and current modus operandi while failing to grasp the universal (green) OA that is already within their reach. Consortial institutional “membership” commitments (like SCOAP3) – intended to bargain down journal prices in exchange for their converting to gold OA – are unsustainable, because, unlike subscriptions, they can be cancelled at any time by individual institutions without losing access (because the journals have converted to gold).

Hence the only scaleable, sustainable and certain means of attaining universal OA is to mandate green OA first, and to convert to gold OA only if and when universal green OA makes subscriptions unsustainable. That means institutional cancellations force journals to downsize to providing the peer review service alone while at the same time releasing the institutional subscription cancellation funds to pay for it.

This scenario is independently confirmed by the Houghton Report, an economic analysis focussed on publishing costs. Its conclusion is that universal gold OA publishing will eventually save institutions money, but that by far the biggest benefit/cost ratio can be gained from mandating green OA today.
Institutional and funder OA mandates need to be **convergent and collaborative** rather than divergent and competitive: institutional deposit followed by central harvesting (not direct central deposit for funder mandates vs. institutional deposit for institutional mandates).

**The Only Obstacle is Over-Reaching**

Apart from the two ways of providing OA (green OA self-archiving and gold OA publishing) there are also two forms or degrees of OA: “**gratis**” OA is free online access and “**libre**” OA is free online access plus certain further re-use rights (which may include republication or remixing in derivative works). *Both gold OA and libre OA are premature and cannot be mandated*; but universal green, gratis OA will prepare the ground for universal gold OA and increasingly widespread libre OA.

Copyright is not an obstacle to universal OA self-archiving mandates; copyright reform will come as a **consequence**, not a **precondition**, of universal green OA. The **majority of journals** (including almost all the top journals) already endorse OA self-archiving of the author’s refereed final draft, immediately upon acceptance for publication.

For the articles in the minority of journals that do not yet endorse immediate OA self-archiving, if the author wishes to honor the publisher embargo, the paper can be deposited in the IR immediately upon acceptance anyway, and access to it can be set as **Closed Access** instead of OA. IRs have a semi-automated **“email eprint request button”** that allows any user to request – and the author to provide – an individual copy of a Closed Access deposit for research purposes through just one click each. This is not yet OA; it is “Almost OA,” but it will soon hasten the end of OA embargos.

OA’s primary target is refereed scholarly and scientific journal articles – 2.5 million articles per year, published in the planet’s **25,000 peer reviewed journals**, across all disciplines, languages and nations – because every one of those articles is, without exception, nontrade writing as described earlier: an author give-away, written solely for research uptake, usage, applications and impact, not for income from sales. Other forms of digital content – books, textbooks, magazine/newspaper articles, music, video, software – are not author give-ways, written for impact rather than income. Here again, the growth of OA to refereed research articles is likely to encourage providing more OA to these further forms of content too, but it is again a great strategic mistake to treat trade and nontrade content as if it were the same sort of thing, under a vague notion of “open access to knowledge.”

Another increasingly important form of research content is **research data** – but providing immediate OA to data cannot be mandated because researchers must be allowed a fair period of exclusive time to mine and analyze the data they have gathered. Researchers can also be encouraged – but not required – to provide OA to their pre-refereeing preprints; this must remain a matter of author choice. There is scope, however, for research funders to mandate that as a condition of funding the data on which a peer-reviewed research paper is based must be made OA **once the paper has been accepted for publication** (with due exceptions for the timing of serial articles all based on mining one data-set).
The optimal green OA self-archiving mandate is the “Liège model,” which designates depositing all papers accepted for publication in the IR as the (sole) mechanism for submitting them for institutional performance review and for national research assessment. Policy guidance for institutions and funders worldwide about designing OA mandates is being provided by Enabling Open Scholarship (EOS), Open Access Scholarly Information Sourcebook (OASIS), and SPARC Campus Open Access Policies.

Creating and Curating the Cognitive Commons

It is now an ironic historic fact – but a true and undeniable one – that in order to create the Cognitive Commons it has proved necessary to require researchers to do with their nontrade publications what was already overwhelmingly in their own best interests (as well as those of humankind as a whole) to do. Future historians will have to explain why researchers did not do it spontaneously of their own accord -- as we at Southampton had at first naively assumed that they would quickly do, as of the mid-1990s – once shown how it was feasible, and given the means (OA IRs) to do it. Biologically, and by way of analogy with the origin of language, it does seem rather like having to force mothers to share their knowledge (let alone their food) with their progeny, except that, if anything, making one’s research findings OA for uptake by one’s fellow-researchers has even more direct and palpable rewards in the here-and-now for the researchers than does the successful transmission of one’s selfish genes to later generations.

So perhaps the reason contributions to the Cognitive Commons had to be mandated at the formative stage is more closely related to why it is that people become scholars and scientists (rather than salesmen, soldiers or politicians) in the first place: After all, even publishing had to be mandated by researchers’ institutions and funders (“publish or perish”), otherwise a good deal of research may never have been published at all, the researchers’ curiosity having been satisfied by simply having conducted the study and then put the results in a desk drawer.

Well, in the PostGutenberg era, even publishing one’s findings for those users whose institutions can afford to access them via subscription is no longer enough. Perhaps the best way to encapsulate Southampton’s contribution to the creation and curation of the Cognitive Commons is to have helped extend the Gutenberg era’s “publish or perish” mandate to the PostGutenberg era of “self-archive to flourish.”

Our graphic illustrates how a series of Southampton milestones have helped OA IRs, OA IR contents, and OA mandates to grow since the posting of the Subversive Proposal in 1994, the creation of the Eprints IR software in 2000, and the adoption and promotion of OA mandates as of the ECS Southampton mandate in 2002.

REFERENCES


HYPERLINKS OF OA INITIATIVES

Arxiv http://arxiv.org/
Citebase http://www.citebase.org/
CogPrints http://cogprints.org/
Enabling Open Scholarship (EOS) http://www.openscholarship.org/
EPrints http://www.eprints.org/openaccess/
Internet History http://www.isoc.org/internet/history/brief.shtml
IRStats http://wiki.eprints.org/w/IRStats
Open Access Scholarly Information Sourcebook (OASIS) http://www.openoasis.org/
Open Access Timeline http://www.earlham.edu/~peters/fos/timeline.htm
Open Archives Initiative (OAI) http://www.openarchives.org/
Open Society Institute (OSI) http://www.soros.org/
Registry of Open Access Repositories (ROAR) http://roar.eprints.org/
Registry of Open Access Materials Archiving Policies (ROARMAP)
http://www.eprints.org/openaccess/policysignup/

ROMEO http://romeo.eprints.org/stats.php

SCOAP3 http://scoap3.org/

Sigma Xi Scientific Research Society http://www.sigmaxi.org/

SPARC Campus Open Access Policies http://www.arl.org/sparc/advocacy/campus/

UK Parliamentary Select Committee on Science and Technology
http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsetech/399/39903.htm


UK Research Councils (RCUK) http://www.rcuk.ac.uk/access/default.htm

ULRICHS http://www.ulrichswbo.com/ulrichsweb/

UNIX Timeline http://www.unix.org/what_is_unix/history_timeline.html

World Wide Web History http://www.w3.org/History.html