

NSF/JISC REPOSITORIES WORKSHOP

Professionally Indisposed to Change

Ken Hamma

We recently went through an exercise to see how difficult it would be for Getty collections to expose non-bib cataloging information (metadata) in the OAI Protocol for Metadata Harvesting and digital surrogates (resources) in a standard File Transfer Protocol for an aggregator to harvest. The OAI harvesting worked in the usual way and a simple URL reference in each XML record triggered the subsequent FTP download and association of resources with metadata. After a few bumps this worked well and, we are convinced, is a simple low-cost alternative (until ORE) to the cumbersome and resource intensive models for contributing to aggregate resources that were part of earlier efforts like AMICO, the Art Museum Image Consortium. (In the end, a single export routine will makes information and images available to OAI exposure and to publication on getty.edu, and makes indexing available to search engines.)

If one wants eventually to have an environment in which it is easy to find, for example, all paintings by Rembrandt Peale and associated archival materials, the strategy of making that possible has to envision the participation of many small and under-funded institutions. Because of that we assumed it is critical to place the bar for cataloging and technical investment as low as possible without impairing simple online discovery. We also began to explore models where key aggregators like OCLC would host open OAI servers as well as publicly available services like vocabulary assisted searching, where application vendors would provide an export mapped to the XML schema, and where big institutions would provide services to smaller ones in regional or intellectually affiliated consortia.

This provides a model for the useful exposure of potentially all repositories of unique works in museums, archives, libraries, which are among the key data repositories for scholars in the humanities. Moreover it does so at the network level and in a way that is predictably available to any institution or individual interested in exploring and aggregating for any purpose, from large scale service providers (ARTstor) and portals (EUBAM), to very large scale resource aggregators (WorldCat), to very small and individual research projects or single classroom teaching.

Why won't this work today.

1. What OAI refers to as resources most archives and museums refer to as assets, and assets are expected to make money. This has a long tradition in museums, which require even scholarly publications to pay significant image licensing fees, and perhaps not as long a tradition in archives from which the

success of a PBS documentary inevitably elicits groans of wanting to share in the money. This is also completely understandable in institutions that, aside from a very few, are habitually under-funded and under-staffed. It is, indeed, an explicit government expectation in many countries that national museums “use” their collections as assets in this way. All of this leads to severe gate keeping at odds with anything that has a scent of openness.

2. Even with (sometimes especially with) the research community there is a mistrust of intentions and a feeling among museum professionals (archivists are getting over this but still struggle) that they must guard over the proper use of resources (assets). This covers a wide range of behaviors from knowing who has got a resource to knowing how it is being used, e.g. not being published with inappropriate cropping or insufficient color management. It is clear to everyone that no such controls truly existed in the analog world, but the ease of high quality digital distribution inevitably leads directly to another reason for gate keeping.
3. For authors a different sense of ownership attaches to even a modest set of cataloging information. This sensibility is strongly grounded in the traditional print model of collection catalogs in which the authors’ names and reputations (and in academic settings tenured career opportunities) are inextricably bundled with the entire scope of information, no matter how trivial, no matter how rich. Many people recognize, once pushed, that ensuring attribution and authenticity in a digital environment is sufficient, but the environment itself is not sufficiently well built out that there is much trust. Exceptions exist.
4. Institutions acting in the same area exhibit an extraordinary lack of imagination. Deliberate limitations on quality and quantity in available data are seen to ensure the viability of traditional modes of scholarly communication. (This can be either a principled decision or one on funding.) And these traditional modes seem very satisfying especially when the network goes down. Never mind that the network has not failed for decades; we fondly remember the days when it predictably did on a weekly basis and that reinforces concerns expressed about the future of those bits and bytes.

A statement released by the Association of American University Presses in February of this year contains the blinkered observation: “The increasing enthusiasm for open access as a model for scholarly communication, which grew out of the pressure to relieve the financial burden on libraries of maintaining subscriptions to STM journals, presents new challenges and new opportunities for university presses. In its pure form, open access calls for an entirely new funding model, in which the costs of publishing research articles in journal are paid for by authors or by a funding agency, and readers can have access to these publications for free.” What body other than the university or research center itself a) employs the researchers, b) provides subventions for press runs of fewer than 500, and c) comes up with the

budget for the library to buy the subsidized but still expensive books for the researchers they employ? If funding is not the real issue – there is a lot of money in that system – could it be that we are expecting too much in the way of critical path decisions on institutional policy and intellectual property from publishers comfortable with or desperate to preserve a traditional model, and librarians just sufficiently well funded to be comfortable with their own and publishers'/distributors' traditional roles?

5. And because we are not sure we want to, for all the reasons above, it is difficult even in well-funded institutions to find the money for item level cataloging and digitization of large bodies of resources. Smaller institutions don't even get this far, stopped in their tracks by the cost of owning technology and managing commercial applications.

Flavors of these issues can be found nearly everywhere one looks, and I don't mean to be pointing a finger specifically at humanists, let alone those I work with. This is just one arena where the characteristic of openness – a characteristic that efficient networks and interoperability assume – makes opportunities seem scary.

I began thinking about this short text not with our OAI exercise but with a position statement from the Wellcome Trust. This won't be new to a JISC/NSF audience, but I would hazard that it and similar position statements have not yet appeared on the horizon at, say, NEH or NEA.

The mission of the Wellcome Trust is to foster and promote research with the aim of improving human and animal health. The main output of this research is new ideas and knowledge, which the Trust expects its researchers to publish in quality, peer-reviewed journals.

The Wellcome Trust has a fundamental interest in ensuring that the availability and accessibility of this material is not adversely affected by the copyright, marketing and distribution strategies used by publishers (whether commercial, not-for-profit or academic).

With recent advances in internet publishing, the Wellcome Trust seeks to encourage initiatives that broaden the range of opportunities for quality research to be widely disseminated and freely accessed.

The Wellcome Trust therefore supports unrestricted access to the published output of research as a fundamental part of its charitable mission and a public benefit to be encouraged wherever possible.

Specifically, the Wellcome Trust:

* expects authors of research papers to maximise the opportunities to make their results available for free and, where possible, to retain their copyright

* will provide grantholders with additional funding to cover open access charges levied by publishers who offer this option and can meet the Trust's requirements

* requires electronic copies of any research papers that have been accepted for publication in a peer-reviewed journal, and are supported in whole or in part by Wellcome Trust funding, to be deposited into PubMed Central (PMC) or UK PMC once established, to be made freely available as soon as possible and in any event within six months of the journal publisher's official date of final publication

* affirms the principle that it is the intrinsic merit of the work, and not the title of the journal in which an author's work is published, that should be considered in making funding decisions and awarding grants.

http://www.wellcome.ac.uk/doc_WTD002766.html

What can NSF do?

Convene federal funding agencies on policies for access to research and its by-products created with public money.

Support the development of open source software and web service models as an element in reducing the cost of technology particularly in institutions that by nature have no affinity for it.

Help envision models that clearly indicate digital products will have a chance of lasting as long as Sumerian clay tablets and with as little effort.

Consider what role for NSF if far-reaching cyber infrastructure issues for the humanities resolve first not in research and teaching but around existing repositories of unique materials that may well reveal themselves, once digitized and available at sufficient scale, as platforms also for scientific investigation.