Copyright - Online Archives and Intellectual Property Rights

Dr. L. Chandra Sekaran  
Indira Gandhi National Open University, Cochin  
E-mail: lcskaran@gmail.com

Dr. N. Shani  
Professor & Head  
Akashaya Institute of Management Studies, Coimbatore  
E-mail: shaniresearch@gmail.com

Abstract

This case study gives an account of intellectual property and rights issues relating to an online research archive. It describes the licensing regime in place to protect the creators and publishers of the data, and provides a case study of actions taken to stop unlicensed re-publication of data by a third party. It concludes that clarity is essential in the protection of intellectual property rights and that effort taken at an early stage will be rewarded in the long term.

Context

The Arts and Humanities Data Service supports research learning and teaching in archaeology in the India by promoting good practice in the use of IT, by preserving important digital research archives for the long term, and by making a large amount of data available online.

At the time of writing, Arch Search (http://ads.ahds.ac.uk/catalogue/), our online catalogue of archaeological resources, provides details of some 600,000 sites, monuments and research projects in the UK and wherever UK-based archaeologists work. In many
cases, short descriptive records provide access to rich and detailed archives that may contain any form of digital object associated with archaeological research. Archives include very large quantities of text - theses, monographs and journal articles in digital form - as well as unpublished 'grey literature'. They include database files, images, CAD plans, geophysical and topographic survey, virtual realities, animations and statistical data. The vast majority of this data is supplied to us from external agencies who license us to distribute and preserve data on their behalf. This non-exclusive license specifies the conditions under which we can distribute data, and has provision for digital preservation but does not transfer ownership of copyright.

Access

Arch Search is free to users, and has few of the restrictions to access encountered in other online services. Archaeological information is the result of very many years of fieldwork and research - much of it funded through the public purse or explicitly to encourage and sustain a better understanding of the historic environment. It seldom has significant commercial value. Users of the data are not required to provide any personal details in advance, nor are they required to remember a password or login. There is no authentication system restricting access to registered IP addresses. Moreover, because we provide access to authentic archival materials, it would be prejudicial to the integrity of the resource were we to brand digital objects with copyright or other fingerprints. In some cases - especially print quality images - we present only screen grade images on the understanding that users can ask for higher quality originals: this is partly to protect copyright of depositors, but is also because these files can be very large and require high capacity bandwidth to access them.

This light regime places an extra responsibility to protect the intellectual property rights of material entrusted to us. Consequently, access to collections is covered by a generous but clear series of copyright, liability and access statements (see http://ads.ahds.ac.uk/cap.html and http://ads.ahds.ac.uk/copy.html). The copyright and liability statements form a common agreement that are shared across the AHDS. So, users are licensed to use resources for their own bona fide teaching, learning and research; they waive any liabilities that may result from use of the collections; and they are required to acknowledge the source of material. These statements are asserted repeatedly within Arch Search. Moreover, a copyright challenge for use of the system explicitly requires that users agree to these conditions. This challenge sets a cookie which expires after 15 minutes, thus ensuring that users have little ground for objecting that they are unaware of the access conditions.

Preservation through documentation

Digital preservation is a specialist and demanding task, of which rights management is only one aspect. It is particularly important for the discipline of archaeology that, through the process of excavation, systematically destroys its own primary evidence. In such
circumstances the preservation of notes of excavation - termed preservation by record - is of critical importance (Richards and Robinson 2000). Excavations in the crypt of Christ Church, Spital fields in London in the 1980s provide an example of the need for such recording, the problems that arise from the creation of such documentation, and complications that can arise in relation to intellectual property rights from this process.

The excavations were undertaken under the auspices of the Greater London Council on behalf of the Parish and Friends of Christ Church Spital fields who sought to turn the crypt of the church into a homeless shelter. Scientific interest in this unusual project ensured that a range of leading agencies and individuals collaborated to undertake the project and complete the painstaking laboratory research that followed. This included the Natural History Museum, the Nuffield Foundation, the Ministry of Works (subsequently English Heritage), as well as the architectural practice that was responsible for the conversion. Project staff were recruited on an occasional basis. Some stayed with the project for a long period of time, while sickness and attendance records suggest that some may have stayed for little more than a day.

Though published in generous detail (Reeves and Adams 1993, Molleson and Cox 1993, Cox 1996), much of the research is now only available by inspecting the site archives. The notebooks, photographs, drawings and recording sheets in the archive provide a meticulous account of the excavation and thus represent the primary source. A long held desire to digitize a representative sample of this data was fulfilled in 2000 with the start of an online teaching and learning project that drew attention to the interdisciplinary nature of this type of study.

Yet digitizing this material presented a familiar problem for intellectual property rights (Dunning 2004). Specifically, it was far from clear who owned the copyright associated with the various files and documents that we digitized. It rapidly became clear that, while conventional archives are licensed via a single exchange of documents from the project team, in this case the project team no longer existed. The numerous agencies and staff involved in the excavations would have to be contacted and their approval sought individually. Amongst these, English Heritage, the Nuffield Foundation and the Natural History Museum had provided funding and in-kind support to the project. Each was asked to sign a standard deposit license allowing ADS to distribute and preserve the data (but retain their copyright in the data. For more on this process, with an example license see http://ads.ahds.ac.uk/copy.html). These three were joined by a fourth license provided by the Christ Church Parish and the 'Friends of Christ Church Spital fields' - a voluntary group that supports the work of the Parish. These were recognized as the most important 'inheritors' of the project and the data that resulted from it. In addition, the project team made strenuous efforts to contact all of the individuals involved in the project, to inform them of our plans and to seek their advice and approval, and to offer them a deposit license should they wish to sign. None in the end objected, but a consensus emerged that individual contributors would only be happy if the Parish and Friends of Christ Church Spital fields also approved.
Archaeology and Human Remains

Parallel to legal concerns over copyright were ethical and quasi-legal concerns over the explicit nature of some of the archive. Christ Church Spital fields was remarkable for the degree of preservation associated with many of the individuals recovered from the crypt. Soft tissue and fabric - seldom recovered from archaeological contexts - was present in significant quantities. This material became an important reference collection for forensic archaeology far beyond the context of eighteenth- and nineteenth-century London. Moreover identification of individuals through name plates and other inscriptions made it possible to cross reference forensic pathology with biographic details disclosed in historical sources. Consequently, much of the post excavation research concentrated on refining methodologies for analysing and identifying individuals from challenging archaeological contexts - such as may be found on battlefields or anonymous mass graves.

However, the handling and study of human remains for archaeological study is not unproblematic (for a discussion see Downes and Pollard 1999). Strict legal requirements exist, in part reflecting a number of ethical principles, such as a desire not to cause offence. It can be seen that digital images of partially decomposed human remains are appropriate tools in archaeological research and teaching. However, internet access raises the possibility that material can be lifted out of context and used easily for inappropriate purposes. Consequently the project team took particular pains to warn users about what they were likely to find in the archive. In this way the project team sought to avoid offence and to ensure that publication of the archive did not bring any of the agencies involved into disrepute. This desire was further amplified by creators of the data in their responses to the invitation to comment on the archive. In addition, senior independent advisors were asked for their opinions. All concurred that the archive was both fit for purpose and appropriate for use before it was released to the public.

Thus, after an extended process of licensing and review, the digital archive from Christ Church Spital fields was released under the standard terms and conditions that apply to the whole of Arch Search.

Research Data Online

The launch of the research archive, with the associated teaching and learning materials, has largely been a success. Patterns of use show that the archive has been among the most used resources within ArchSearch: in 2002-3 and again in 2003-4 the archive was the single most popular excavation archive available within the system (ADS 2004). The tutorial has been used in teaching in contexts that we previously could not have envisaged (Kilbride et al 2002), and we have received a number of requests for further data such as contact details for how to obtain access to the rest of the research archive. This represents a success within the operational plan of the organisation.
Late in 2003, routine monitoring of our systems showed an unusual amount of activity of use in the archive. Further analysis of this activity led to the identification of a website which had republished some of the more explicit - and thus potentially offensive - images from the archive. This new context in which the images had been embedded was a breach of the access agreement and copyright statements which users are required to accept before using these archives. Moreover, having seen this infringement it would have been negligent to allow this unlawful republication to continue. Given the legitimate concerns of data creators and those who licensed the data to us, it was imperative that we act swiftly and successfully.

**Action Plan and Procedure**

This is the first time since their creation that our Access Agreement and Copyright Statement had been invoked to protect the intellectual property rights of depositors. Though we hoped that the matter could be resolved without aggressive legal action, it was clear that we needed to imagine a 'worst case scenario', in which our host institution would have to take more formal legal action. Envisaging such a scenario we first had to be certain that the material was indeed copied from our web server and that the Access Agreement and Copyright statement had indeed been breached.

Both were straightforward. The project team had digitised the images from the original slides and still had them in our possession: the data could not have been obtained from any other source. The Access Agreement is explicit in only allowing users to pass on information to approved users - ie under the same terms and conditions by which they had been supplied to them - while the Copyright Statement offers a 'non-transferable' licence to users to use within their own research learning and teaching. Republication on a third party website is clearly beyond the scope of both documents. Moreover, the Access Agreement entitles us to terminate the licence they have to use the data, and makes it explicit that users risk legal action should they breach any of the conditions.

Having established this, we then approached senior managers to ensure that they supported our request to have the material removed - aware of the 'worst case scenario'. At this point we also consulted various legal authorities - in particular JISC Legal for their advice. This advice confirmed our view that we should begin with a relatively low key approach to the webmaster of the offending site, but also identified an alternative course of action which could be pursued through the website's Internet Service Provider (ISP). A recent case - Godfrey vs Demon Internet Ltd suggests that ISPs share liabilities in relation to material they publish (Charlesworth 2000).

The result of these consultations was to devise a simple decision tree that we should follow in the event of our requests being ignored or denied. This may be extensible to similar cases, so is published here should others find themselves in the same situation in the future. It is now our own agreed protocol of how to respond to future incidents of this kind, and might best be though of as a flowchart with 5 steps each of which take 14 days.
It ends in either the offending material being removed or in more formal legal action. One would hope that it is resolved before the end point, which is 70 days from the first email.

1. (‘Day 0’) Contact webmaster identifying clearly the areas of concern, identifying the terms and conditions breached, specifying changes to be made, requesting an immediate acknowledgement and requesting that changes be completed within 14 days from the dispatch of the first email.

2. (‘Day 14’) If the matter has not been resolved in 14 days, contact the webmaster again making clear the legal implications and our intention to proceed to court if changes were not completed within 14 days of the dispatch of the second email. Copy of this email to the Internet Service Provider.

3. (‘Day 28’) If the matter has not been resolved in 28 days, contact the ISP pointing out the infringement and requesting that they intervene to have the offending material removed from their domain within 14 days of the dispatch of the third email.

4. (‘Day 42’) If the matter has not been resolved by this point, write again to the ISP making clear the legal implications for them and our intention to proceed to court if changes were not completed within 14 days of the dispatch of the fourth email.

5. (‘Day 56’) If the matter has still not been resolved, obtain letters from the university lawyers to the webmaster and to the ISP making clear the infringement and requesting changes within 14 days of the postmark of the letter.

6. (‘Day 70’) If the matter has still not been resolved, takes further legal advice and begins legal proceedings.

**Publisher Agreement**

We have reached agreement with participating publishers to include each journal’s back issues in a non-profit organization own to define the "back" issues? A "moving wall" approach was devised - we start with the very beginning of a journal's run (in some cases well back into the last century), and let the publisher tell us how far forward to go. For example, if the publisher specifies a moving wall of five years, we include all issues up to five years ago, and move the journal forward in a non-profit organization one year at a time, maintaining the five year lag. One purpose served by the moving wall is to provide assurance to libraries that they can rely on the archive into the future. For those publishers who are mounting an electronic version of current issues of their journal, we will work with a "fixed wall," ending the run of issues included in a non-profit organization at the point where the publisher's electronic issues begin. However, should the publisher ever decide to drop any issues from its own electronic archive, we require that those issues then be included in a non-profit organization.
Researchers ultimately will not be content to search only journals published more than five years ago, or to have to submit separate searches for recent and older materials. The capability to conduct a search in either the current or archival database of a journal, and to have that search seamlessly carried out in the entire published run of the journal, will be of enormous scholarly value. It is our intent to work with publishers to develop effective linkage between the archival issues and the electronic versions of current issues to make this possible.

From a copyright law standpoint, it is not our contention that publishing a journal in a non-profit organization constitutes the creation of a new work. Our approach is image-based, and our goal is faithful replication of all issues of the journal, from cover to cover, including advertising material, membership lists, or any other miscellaneous content. As specified in our publisher agreement, the publisher owns the copyright rights in the digitized images that we create. In order for a non-profit organization to serve as a reliable archive, however, we must offer permanent access rights to libraries. Therefore, we ask the publisher to grant us a perpetual license to the digitized archive that we have created. If the publisher were to withdraw from a non-profit organization this perpetual license is limited to the right to provide the archive only to those libraries that were participating in a non-profit organization as of the date the publisher terminates our agreement.

The non-profit organization publication license is nonexclusive - publishers can license others to digitize the back files of their journal if they so choose. It is also royalty-free, although to the extent that our revenues from participating libraries exceed our costs we will make distributions to participating publishers. Finally, we do not ask that publishers commit to us for a fixed term. Instead, we agree that either party may cancel on six months' notice. This cancellation provision imposes a present discipline upon us; it means that we must continue to be responsive to publishers' needs as well as to those of libraries and scholars.

**Library Agreement**

Probably the hardest issue that we faced in drafting a library license agreement was how to address the difficult question of interlibrary loan (ILL). We had listened to publishers tell us that ILL is rampant and out of control, how libraries are not obeying the "rules," (As per Section of the Copyright Act and the related Guidelines), that the procedure undercuts current subscriptions to journals and is basically a huge disaster. We had also listened to libraries tell us that ILL is a right granted from on high, is limited and controlled, and that everyone is following the rules scrupulously. We had to propose some compromise or neither side would have been willing to work with us. With the help of friendly hardliners from each camp, we molded and negotiated a clause that offers something to both sides. Libraries are permitted to engage in ILL for two years with materials that are printed from a non-profit organization but not with electronic copies. Libraries are required to keep records of their use of a non-profit organization materials in ILL. At the end of two years, we will all catch our breath, examine these records, and
see just how extensive ILL really has been. In the meantime, publishers will have had access to the digitized archive of their journal through a non-profit organization as well. We will provide access to publishers in the publicly accessible area of our web site. It is possible that provision by publishers of materials directly will obviate the need for ILL over the long haul.

Other aspects of our library agreement also reflect our goal of finding workable compromises between competing interests. In an effort to make the a non-profit organization database as useful as possible, we offer libraries a broad definition of "Authorized User." The agreement is basically a "site license," meaning students, faculty and staff can access the database remotely, but we also include in the definition of those who may use a non-profit organization anyone physically present in the library. In this way, we accommodate public repository libraries.

We hope to work with libraries collaboratively in many ways. For example, there is much to learn about the usage of the archival issues of these journals, and much of this learning will require cooperation and input from libraries. We intend to work together to establish effective links into the database to increase its usefulness. We will undoubtedly be cooperating as technologies evolve and means of protecting access to the database change.

The non profit organization user agreement is standard in many ways, consisting of the typical online shrink wrap terms with which we have all become familiar. In a bow to the needs of today's students and researchers who may wish to carry a copy of something on diskette home for the weekend, we offer the right to make one electronic storage copy of any journal article in addition to one printed copy.

Next Steps

These initial decisions enable us to move forward with publishers and with libraries. But the environment is fast-moving, and our decisions must be constantly tested, challenged, and modified as necessary. We are now turning to new questions and issues, and are working to devise and test possible answers.

One interesting aspect of a non-profit organization is that our software development and production offices are located on university campuses. There are a lot of positives in this relationship: we benefit from the heat of the research environment, and students and researchers are able to work on a real-world endeavor. We remain committed to working within this community. At the same time many interesting questions arise due to this arrangement, from IP ownership, to devising an effective management structure, to imposing a production schedule on an academic setting.

On the technical side, we decided to combine the use of images to display the pages of journal literature with the use of a text file created using OCR to facilitate searches of the material. Key factors contributing to this decision were the need to display non-textual
material accurately, the importance to both libraries and publishers of faithful replication of the journals, and the economic realities we face in providing access to hundreds of years of back files. While we believe that this approach offers the best of both the image and text worlds, we recognize that this choice does not come without some costs. Our technology group is working hard to reduce, if not eliminate, the performance differential that exists between images and text. Every effort is being made to reduce download and response times for use of the database. JSTOR has developed helper applications to allow users to print articles on their printers at a variety of resolutions, including an archival quality 600 dpi version. An extremely promising development is the availability of new compression technology from Cartesian Products Inc. that appears to reduce the size of the non-profit organization database by a factor of 6 to 7 times as compared to Group IV TIFF.

From our inception, it has been anticipated that the non-profit organization would be a resource of great value not only in the United States, but around the world. Given the state of the Internet, and the difficulty encountered by users overseas in trying to access U.S. sites, it will be necessary to develop mirror sites in geographically dispersed locations. This will present both technical and legal challenges. Our original database was built at the University of Michigan, and we now have a mirror site located at Princeton University. We are developing and implementing processes for keeping the data synchronized and for balancing the load between the two sites. The legal terms by which we will accomplish provision of a non-profit organization to non-U.S. libraries present interesting issues given the vagaries of IP regimes internationally and the unsettled nature of online law generally. We hope to develop relationships with individuals and entities abroad to help us stay abreast of changes in the legal, political and regulatory landscape. The non-profit organization is off to an auspicious beginning. We are taking steps into new territory, technical as well as legal. At the same time, as a small organization in an industry that is changing by the nanosecond, we must also practice patience. We are trying to learn as much as we can about the issues of importance to us, and to make educated decisions based on collective experience. As Elton White, one of a non-profit organization board members puts it, we strive "never to get our spending ahead of our learning." Success will depend upon our ability to remain open to change; to continually refine our decisions as we learn.

**Conclusion**

In the end we were pleased not to have to proceed beyond the first stage in our decision tree. The offending material was removed within a few hours of the first contact being made and the matter was closed.

At first inspection, threatening legal action seems an excessive and risky policy to resolve a breach of copyright. In this case, however, we were confident in our approach for a number of reasons, mainly to do with clarity. We had a very extensive documentation trail that showed conclusively where copyright resided. The Access Agreement and Copyright Statement were well drafted so the nature of the infringement was clear. We
could spell out in simple terms the infringement that had occurred and what action we required. In fact these modifications required relatively trivial amounts of work: so we did not have to make onerous demands. In addition, one suspects that the relatively generous access conditions also worked to our advantage. It could hardly be argued that we were restricting access to information or protecting special intellectual privilege. All of these gave us confidence to act decisively.

Thus, clarity in intellectual property rights management was core to the protection of open access, web-based research materials. The effort at the start associated with licensing materials and ensuring that users are aware of conditions of use has proven to be time well spent. Our experience suggests that protecting IPR needn't require complicated mechanisms of authentication and authorisation: especially in contexts where resources have limited commercial value. But without clarity no regime - however sophisticated - is likely to be viable.

Bibliography