Humphrey Southall[*], Petr Přidal[**]

Old Maps Online: Enabling global access to historical mapping

Keywords: Federated map search; geographical searching; MapRank; DCMI BOX; MARC 034; metadata; Bounding Box.

Summary: Over the last thirty years, map libraries worldwide have scanned some hundreds of thousands of historical maps from their collections, and most are viewable online via the world-wide web. However, maps remain difficult to access because they are individually hard to find: a user needing a map of a given location needs to know which library or libraries to search, and library search interfaces generally require knowledge of map titles: there has been no "Google for old maps". We review previous attempts to create such a federated search portal and then explain how our new portal at www.oldmapsonline.org differs. One feature of our approach is that we prioritize rapidly achieving critical masses of content and usage over long-term sustainability. This means that we are assembling metadata from map collections essentially manually, maximizing the number of libraries who can participate. One longer term aim is to encourage use of sustainable web addresses for historical maps, Universal Resource Identifiers rather than URLs, which do not contain references to particular pieces of software, or reflect particular transitory arrangements of library web sites. A second long term aim is not to create a specialized automated metadata harvesting system for maps, but to ensure that the systems which major libraries are anyway putting in place for their overall collections do include the spatial coordinates needed to make content geographically discoverable.

Introduction

The two biggest limitations to traditional map libraries were firstly that few people lived close to one, and secondly that few people could find maps in them: cartographic historians could find maps in the conventional catalogues because they were actually interested in map titles and the names of cartographers, and the map librarians knew their collections well enough that they had to find maps for all the other users, who knew only the places they wanted maps *of*.

Both problems are potentially soluble through digitization. More obviously, once maps have been scanned they can be made available to anyone anywhere with Internet access, while modern online maps viewers are in some ways preferable to using a paper sheet. Perhaps less obviously, a digital map catalogue can be searched geographically, so users need only zoom in on a map to the place that interests them, and the catalogue software then returns a list of maps covering that point; and where the maps are historical, the visual interface can be extended to allow users to specify the date range to be covered with sliders. Such a visual interface to a historical map library is an example of faceted search (Schraefel *et al*, 2006), and the application of this approach to map libraries is surveyed by Fleet (2008).

However, making the searching of individual map library catalogues easier will not greatly expand their usage because relatively few of the potential users are used to using any particular library, or have any idea which library to look in for maps of a particular area. What is needed

[*] University of Portsmouth, Department of Geography, Buckingham Building, Portsmouth PO1 3HE, UK [Humphrey.Southall@port.ac.uk]
[**] Klokan Technologies, Bahnhofstrasse 9, 6340 Baar, Switzerland [info@klokantech.com]

instead is a single place on the internet – a portal – able to simultaneously search as many map libraries as possible. Many portals can coexist, as digital metadata is infinitely sharable, but where the maps themselves are all globally accessible the most useful search portal will be that covering the most collections, global rather than limited to maps held in one country.
This paper describes the Old Maps Online project which has created just such a portal. Our emphasis here is on the aims of the project and how we aim to achieve them, not on the technology *per se*.

Federated map search portals

Old Maps Online (http://www.oldmapsonline.org) is anything but the first project to try to create a search portal covering historical maps in multiple collections; we are not even the first project called Old Maps Online (Přidal and Žabička, 2008). Earlier projects focused specifically on finding maps have been of two types:
Firstly, projects creating an integrated portal for both searching and viewing, so that for all practical purposes the multiple paper map collections become a single digital collection. So far such projects have all been limited to the collections of a single country. Examples include Open GeoPortal (www.opengeoportal.org), linking a consortium of academic libraries in the US, and the Cartesius project in Belgium. Strictly speaking, Kartenportal in Switzerland should not be included here because it links to a separate site for viewing, but to only one such site, Swissbib, the metacatalog of Swiss university libraries and the Swiss National Library. In practice, such projects seem likely to be limited to libraries having close institutional ties, so individual portals are unlikely to ever cover a significant fraction of the world's maps.
Secondly, projects that are purely search portals and therefore provide access to multiple digital collections. Such systems can provide users with a comprehensive access route to historic mapping, a true "Google for old maps", but only if they cover a substantial fraction of the world's collections. This is our aim, and we know of three earlier projects which have attempted this.
First, the DIGMAP project funded by the European Union's eContent Plus programme (Borbinha et al, 2007; Borbinha et al, 2009; www.digmap.eu). DIGMAP ran from 2007 to 2009 as a funded project. It aimed to hold metadata for as many maps as possible, so was designed primarily for maps with only conventional metadata. It is consequently problematic as a tool for geo-spatial searching: much of the geo-spatial metadata was generated algorithmically by parsing map titles, so it can be profoundly inaccurate; and the user interface for geo-spatial searching requires users to actually type in the coordinates of bounding boxes, limiting the system to highly expert users. Although the DIGMAP portal lists access to 49,173 "resources", these are authors, dates and places. The system does list under "other resources" 5,680 maps, but the project's funding has ended so the collections are not now being extended.
Second, Cartomundi, developed by the Maison Méditerranéenne des Sciences de l'Homme (Arnaud, 2008; www.cartomundi.fr). Cartomundi differs from both DIGMAP and Old Maps Online in that it treats copies of maps not simply as individual items grouped into map libraries, but as instances of published maps, which are in turn often grouped into series. This is clearly a desirable feature. Unlike DIGMAP, Cartomundi has a business model for sustaining itself: that map libraries pay a subscription enabling access to premium features, payments justifiable in part by the assistance the system can provide with cataloguing because of the information it holds about the publication history of maps. However, it is unclear how many libraries can afford these subscriptions. Cartomundi currently says it holds 8,000 maps.

Thirdly, the *Maps of Australia* mapsearch portal (http://mapsearch.nla.gov.au) is federated in that it searches "over 100,000 maps of Australia held in Australia's libraries", but only three per cent of the collections covered have been digitized and these are primarily in the Australian National Library in Canberra, so as a portal to online mapping it is a limited resource.
Two other federated search portals should be noted although neither is focused on historical maps. Firstly, the National Geospatial Digital Archive was a collaboration between UC Santa Barbara, and Stanford University to create a federated network for the US to archive geospatial imagery and data (Erwin and Sweetkind-Singer, 2009). However, this is primarily a collection of "born digital" data and cartographic content in their Globetrotter portal (http://clients.alexandria.ucsb.edu/globetrotter) is limited to modern USGS mapping. Secondly, the Electronic Cultural Atlas Initiative (Buckland and Lancaster, 2004) envisioned an enormously ambitious distributed system linking diverse geo-referenced cultural content held by individual scholars via a central metadata clearing house implemented using TimeMap software developed at the University of Sydney (Johnson, 2005). The ECAI clearing house does exist online (http://ecaimaps.berkeley.edu/clearinghouse), and does retrieve some historical maps, but seems never to have actively solicited library metadata and most current content seems to derive from the University of Sydney's own collections.

The Old Maps Online portal

So how does the Old Maps Online project differ from these precursors? Firstly, we are not developing portal software from scratch, but rather using the MapRank Search software developed by Klokan Technologies GmbH and already in use by the Swiss Kartenportal system (Oehrli *et al*, 2011), the Moravian Library and the David Rumsey Collection. This has three major benefits. Firstly, we eliminated development risks. Secondly, the software had already been subject to extensive work to test and enhance usability, especially by the Kartenportal project. Figure 1 presents the Old Maps Online portal, and shows how we have kept the interface simple but highly interactive so as to make it accessible to the widest possible audience. Thirdly and most importantly, using existing software enabled us to launch the portal within four months of the start of our funding (November 2011), so we could spend most of our fifteen months working to add additional libraries to the collection, and promoting the portal to end-users.
A second major difference from DIGMAP is that we were clear from the outset that we would include only a quite narrow subset of all historical maps held by libraries: to be included in Old Maps Online, maps must have been scanned; the resulting images must be freely and directly accessible on-line; and the real-world coordinates of the map corners must be available for use in the portal. Imposing these requirements on libraries enable us to offer a very simple and attractive proposition to users: all you need to do to find relevant historical maps is use the interactive map in the portal to navigate to the real world location you are interested in; in fact, even this is not always necessary as our software can often identify a user's location from their internet address, and automatically takes them to that location in the portal. Further, once a user has found a map of interest, all they have to do to view it is click on our thumbnail image, and they are taken straight to the full image of the map on the relevant library site. An unavoidable limitation is that the different libraries we link to use a variety of viewer software, but in practice most viewers follow some standard conventions for zooming and panning.
Although Old Maps Online requires that every map be geo-referenced, enabling a map-based search interface, our approach is very different from the OpenGeoPortal project, which essentially

treats historical maps as a class of geospatial data, accessed within an interface which will be very familiar to those used to GIS software such as ArcGIS or Quantum GIS, but arguably over-complex for those lacking that training. In Old Maps Online, there is no requirement that maps be geo-rectified by assigning multiple control points and then stretching and compressing the image to better fit it to real world geography. Neither is there any expectation that the library holding the image will be presenting it using specialised geospatial software; rather we expect them to be using standard image viewers, such as Zoomify or IIPImage. It is only our portal system which needs to hold coordinate data for the maps, and for older maps these coordinates are unavoidably fairly approximate.

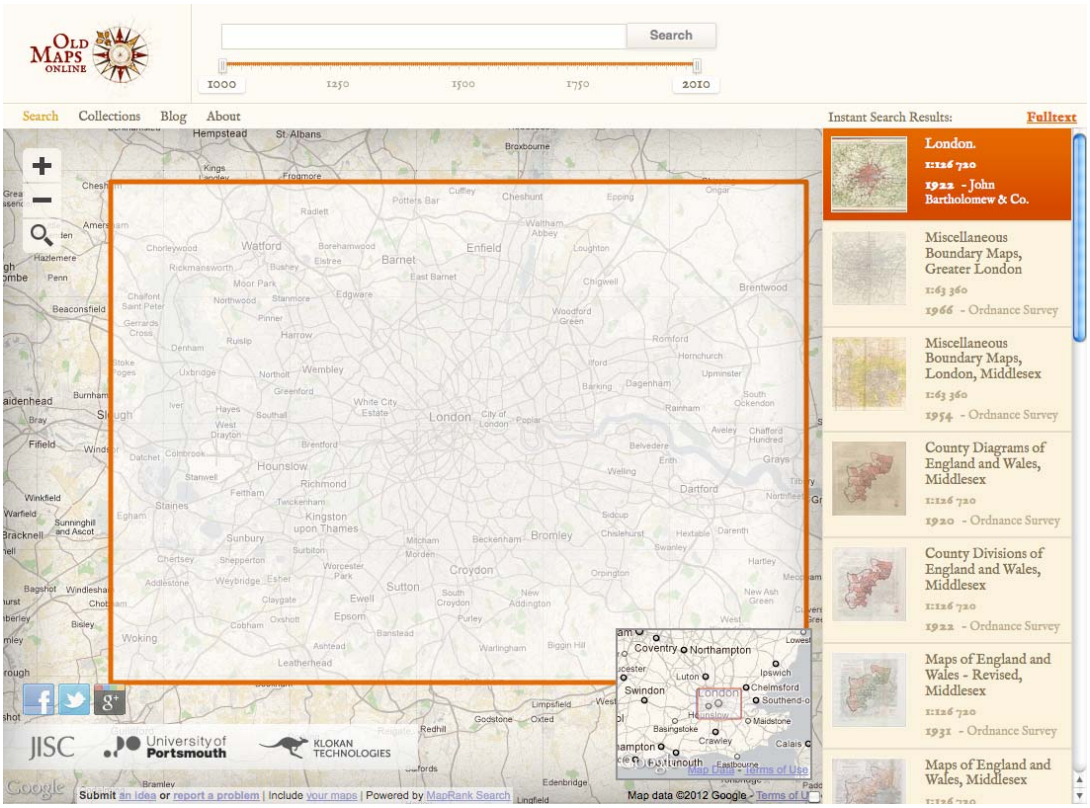

Figure 1: The Old Maps Online search portal. Clicking on one of the maps on the right brings up more details including a larger thumbnail image. Clicking on that image takes you to the map on the relevant map library web site.

The third major difference from earlier projects is our "business model". Our funding application argued against "overly demanding sustainability models". Over the period of our funding, our aim is not to create a kind of business which then generates enough income to sustain itself indefinitely, but rather to create a resource which is *worth* sustaining. This means that our priorities are to expand the "collection of (map) collections" as rapidly as possible, and to develop as large a user base as possible actually within the project period. It is very helpful that, whereas the European Union emphasises "business models", Old Maps Online is entirely funded by the UK's Joint Information Systems Committee whose sustainability requirements are much simpler: we are required to operate the Old Maps Online portal for five years from the end of the project funding, which takes us to the start of 2018; but there is no requirement that the portal be updated. We are meeting this requirement mainly through advance payments from the project grant, to Klokan Technologies to maintain the software, and internally within the University of Portsmouth for server hosting within a private "cloud". The latter buys us only a fairly minimal server, but we

will be adding advertising from Google to the search results. We already have experience with this from our existing web site *A Vision of Britain through Time*, where Google advertising is bringing in about £9,000 per annum: if Old Maps Online proves as popular, advertising income will almost automatically scale up to pay for additional server capacity; and using cloud hosting means that provided we can pay for it it is easy to vary the computing hardware resources used in line with demand.

By not trying to solve all our sustainability issues, we dramatically simplify the task of adding content to the portal: for now, we are doing this essentially manually, and by any means necessary. We do not charge libraries for inclusion, and aim to provide them with considerable assistance with adding their metadata to our system. Our hope is that most of them can send us a spreadsheet with one row for each map. The absolute minimum is a map title; a date for the information on the map; a URL at which the map image is viewable; and the real world coordinates of the bottom left and top right corners of the map. We can optionally include most other conventional cataloguing information, such as author, publisher and place and date of publication. We hold coordinates internally following the WGS-84 standard, but can handle conversions from other systems provided libraries tell us what the system is. We also need to be told what image viewer software is in use, and some basic identifying information about the library, so we can for example display its logo where appropriate. The main restriction we impose are that maps be *directly* viewable at the URL supplied without payments or passwords, and that they must be *reliably* viewable at the URL. We do not expect this to be a problem with major libraries, but can be an issue for on-line map sites created essentially by individuals.

The Old Maps Online web site was launched at the New York Public Library in February 2012, with 60,837 high resolution maps from five collections: the British Library, the David Rumsey Collection, the Moravian Library, the National Library of Scotland and *A Vision of Britain through Time*. Additional metadata from and via the Rumsey Collection is being added in April 2012 taking the total to over 71,000 maps, while further contributions from the Harvard Digital Library, the New York Public Library and Retromap (Russia) should take the total past 100,000 maps, every one of them findable and viewable within a minute of arriving on the portal.

Improving map referencing

Two of our longer term aims are simply to expand the number of maps and map collections accessible through the portal, and to expand the number of users. However, we have two additional aims which are about influencing how maps are treated by libraries. The first is to improve the URLs used to access historical maps within library web sites, making them simpler and less prone to change, the ultimate aim being to enable historical researchers to provide citations which make sense to both computers and humans, and which reliably identify online maps not just at the date of publication but indefinitely. In assembling the metadata for the portal we are in effect carrying out a census of such URLs.

The problems with current URLs could be demonstrated from many map library sites, but we will focus on the one site the project is unambiguously committed to improving, the *Vision of Britain* Map Library. For example, this URL accesses Wilkinson's 1812 map of *The British Isles,* scanned for us from the British Library collection (Maps 177.d.2.(15.)):

http://visionofbritain.org.uk/iipmooviewer/iipmooviewer.html?fileName=wilkinson_1812%2Fwilkinson_1812%3DThe+British+Library%3DR.+Wilkinson%3DThe+British+Isles%3DThe+British+Isles&x=68&y=59

Why so long, and why so obscure? The general problem with ours and many other sites is that the URLs are an unconsidered by-product of the work to set up the image viewer, a task almost always left to someone with specialist IT skills, not a librarian, and often done under considerable time pressure in the final weeks of a digitisation project. The above URL is more human-readable than some, as it incorporates the names of the map, the publisher and the library holding the paper copy. These are used by the viewer as labels, but the format in which they are passed is arbitrary and the replacement of spaces by "+" and what would otherwise be equals signs by "%3D" adds to the obscurity; it would be better to identify the map via a simple and potentially enduring identifier, most obviously an accession number.

The middle part of the address is deeply problematic. What does "iipmooviewer/ iipmooviewer.html" mean? For that matter, why do maps in the British Library have URLs like the following?

http://www.bl.uk/onlinegallery/onlineex/ordsurvdraw/c/002osd000000009u00103000.html

In general, the part between the institution/domain name and the identifier for the individual map contains two elements: one or more directory names identifying locations within the web site, in our case "iipmooviewer"; and something identifying the actual viewer software, in our case again "iipmooviewer" because we are using the open source IIPImage viewer (http://iipimage.sourceforge.net). The problem is that these elements are very likely to change whenever there is an overhaul of the institutional web site; for example, if the British Library decided to standardise its terminology on either "online gallery" or "online exhibition". Just one change of the directory name, directory location or the viewer software will result in users receiving "404 Page Not Found" error messages.

Policies addressing these issues already exist, and in the UK have already been endorsed by the government and by institutions such as the British Library. RFC3986 of the Internet Engineering Task Force (Berners-Lee *et al*, 2005) defines a general syntax for Uniform Resource Identifiers (URIs), which are "compact sequence[s] of characters that identifies an abstract or physical resource"; while Chief Technology Officer Council (2009) provides detailed guidance on implementing RFC3986 within the UK public sector, and is generally more readable.

The latter report sets out a set of general principles which any system of URIs should follow. Firstly, all URIs should actually work as web addresses; this distinguishes them from Digital Object Identifiers (DOIs), which are intended to be more generic. Secondly, "those public sector URI sets that are promoted for re-use should be designed to last for at least 10 years" (p.3), which should be an unambitious goal for memory institutions such as libraries, some of which have maintained sequences of accession numbers for multiple centuries, but means that addresses must continue to work over timescales on which server hardware and software are both likely to be completely replaced. For that reason, "a URI structure will not contain anything that could change" (p.4), which covers both server software and institutional structures: much UK government "open data" is now being published at data.gov.uk rather than on the web sites of individual ministries because ministries are frequently merged, split or renamed in periodic reshuffles. Memory institutions may be more durable, but how permanent are their internal sub-divisions? Maybe the most important guideline is simply "a URI path structure will be readable so that a human has a reasonable understanding of its contents" (p.4).

Our aim is to implement the above principles within *A Vision of Britain through Time*, as far as possible for both the individual sheets accessed via IIPImage and for the mosaics in our Web Map Server; and to use this experience to publish a guide to good practice for other map libraries to

follow. It is obviously not too hard to devise idealised systems of URIs, once the need for them is identified, but we also need to implement them with real world image servers. This work still lies ahead of us, but we can outline our general approach. In particular, any realistic strategy needs to work with the image servers that libraries have already installed and paid for, which means that parts of our guide will be specific to particular software; in particular, configuring sensible defaults so that only an identifier need be passed to obtain an initial view of the image. However, we see a large part of the solution as lying in the systematic re-writing of URLs by the overall web server before they reach the image server. With either of the two most widely used web servers, the open source Apache (Apache Software Foundation, 2012) and Microsoft's Internet Information Server (Yakushev, 2009), setting up URL re-writing requires "only" uploading a text file containing re-writing rules; but these rules depend on regular expression matching and are notoriously obscure. We cannot remove this obscurity, but we can provide a cookbook of examples tailored for map libraries. Much can be done by establishing some standard element that always appears between the overall domain name and the item identifier, such as "/images/maps/", which then gets re-written to the current implementation-specific path, such as "/iipmooviewer/iipmooviewer/". We suggest that the largest problems will be institutional not technical: curatorial staff deciding what is *really* permanent about their institution and digital collection, and then getting technical staff and contractors to implement it.

Harvesting metadata

The long-term justification for the above work to promote simpler and more persistent identifiers for online historical maps is to make it easier for users to cite them, and therefore make maps more central to their scholarship. However, making such identifiers less likely to change has obvious practical benefits for a project relying on manual harvesting of map metadata, and with funding to employ staff to do this for only a little more than a year. Our limited capacity to update the metadata also explains our requirement that maps be "*reliably* viewable": we assume that libraries will be able to meet this requirement but are cautious, for example, about collections created by individual researchers and hosted on essentially personal areas within university web sites.

Even with such caution, and even if libraries make map URLs less implementation-dependent, we doubt that we can sustain *Old Maps Online* as currently operated much beyond the five years required by our funder. However, and provided we achieve critical masses both of map collections and users, within that five years we hope to play a significant role in achieving our final aim: an automated system for harvesting geo-spatial metadata for historic maps, making portals such as Old Maps Online self-sustaining. Here again we see the challenge as primarily institutional, not technical.

The technology for such harvesting has existed for some time, most obviously via the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH; Van de Sompel, 2004); and indeed the DIGMAP project discussed above implemented just such harvesting. The problem is that even where libraries are exposing their metadata for harvesting, in most cases they are not exposing the coordinate data required to support geographical searching; specifically, they are not populating field 034 in the MARC encoding, which maps to DCMI Box in the Dublin Core element set. For example, the DIGMAP web site lists 8,406 "resources" as having been harvested using OAI-PMH (http://portal.digmap.eu/index/index.html, accessed 1st April 2012); but 7,548 of these come from the National Library of Australia; and our examination of that library's metadata

as currently exposed via OAI-PMH does not show any maps with metadata in a form we can use. Similarly, the Library of Congress offers bounding box metadata for only 1,832 maps, and the French Bibliotheque Nationale offer it for 736 maps. Given that our manual harvesting has already assembled geospatial metadata for over 70,000 maps, something is clearly profoundly wrong with automated harvesting: the numbers of maps currently accessible make sense only as a proof of concept, but OAI-PMH was being established ten years ago and by now we should be well beyond proof of concept.

One way forward would have been to organise a sequence of international meetings where map librarians came together in touristic cities around the world to make plans for making geospatial metadata available for many more maps. It is unlikely that we could have obtained funding for such a project from a UK-only source, but anyway such approaches have been tried before and failed, at least in terms of metadata availability rather than touristic experiences. In the specific case of metadata for maps, there is not just the general problem of institutional inertia but the particular problem that, with some specific exceptions like the Rumsey Collection and *Vision of Britain*, maps libraries are usually embedded within larger libraries, within which cataloguing systems and web services are provided centrally: getting map librarians excited about DCMI Box can be only a first stage, they must persuade the overall heads of their libraries that this matters.

Our ultimate aim, therefore, is to assist them in this by creating a portal designed not for cartographic historians or GIS specialists, but for everyone fascinated by past places, notably the army of amateur local and family historians; and a portal containing as much of the world's cartographic heritage as possible, populated by any means necessary. Such a system should lead to a large enough rise in usage of online historical mapping at the participating libraries that even head librarians start to notice. In the really long term, we hope those head librarians will start to realise that geographical search interfaces work well with many kinds of geo-referencable content besides maps: topographical drawings and photographs; travel writing; archival documents linked to places, which were the focus of the QVIZ project (Aucott *et al*, 2009); even sound recordings of dialect speech or birdsong.

Conclusion

The majority of digital library projects are of course concerned simply with scanning, and making the resulting images available online. There is an obvious need for projects which go beyond this, and particularly for projects which make the ever-growing body of digital content more easily findable. However, most such projects are handicapped by funders requiring them to advance the *technical* state of the art. One risk is that projects developing new software will simply fail, but just as useless are projects that create novel but unusable interfaces, or that spend years simply reinventing the wheel. Even where projects are entirely successful in developing great new interfaces, that tends to be their final act before the project funding ends: the web site keeps running, but there is nobody to add content or promote it to users.

We were deeply fortunate with Old Maps Online to be able to base our portal on existing software, so that we can concentrate on what comes next: creating a metadata collection spanning not thousands of maps but hundreds of thousands of maps; establishing a user base far larger than the specialist audience for traditional map libraries; and then promoting changes to library practices making maps easier and more reliable to cite, and to enable systematic harvesting of library metadata including geo-referencing. It will be clear that unless and until this final aim is

achieved *Old Maps Online* cannot be sustained indefinitely, but we believe that our first priority should be to build something *worth* sustaining.

References

Apache Software Foundation (2012). *Apache mod_rewrite*. Apache Software Foundation (http://httpd.apache.org/docs/2.4/rewrite, accessed 2 July 2012).

Arnaud, J.L. (2008). A New Website for Cartographic Resources — CartoMundi. *Liber Quarterly* 18 (2), 271-5.

Aucott, P., A. Von Lünen and H.R. Southall (2009) Exposing the history of Europe: the creation of a structure to enable time-spatial searching of historical resources within a European framework. *OCLC Systems & Services*, 25 (4), 270-286.

Berners-Lee, T., R. Fielding and L. Masinter (2005) Uniform Resource Identifier (URI): Generic Syntax. Internet Engineering Task Force, *Request for Comments 3986* (http://tools.ietf.org/html/rfc3986).

Borbinha, J., G. Pedrosa, D. Reis, J. Luzio, B. Martins, J. Gil and N. Freire (2007) DIGMAP – Discovering Our Past World with Maps. *Lecture Notes in Computer Science*, 4675, 563-6.

Borbinha, J., G. Pedrosa, J. Luzio, H. Manguinhas and B. Martins (2009) The DIGMAP Virtual Digital Library. *e-Perimetron*, 4 (1), 1-8.

Buckland, M., and L. Lancaster (2004) Combining place, time, and topic: The Electronic Cultural Atlas Initiative. *D-Lib magazine*, 10 (5).

Chief Technology Officer Council (2009) *Designing URI Sets for the UK Public Sector*. Cabinet Office (http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf, accessed 2 July 2012).

Erwin, T., and J. Sweetkind-Singer (2009) The National Geospatial Digital Archive: A collaborative project to archive geospatial data. *Journal of Map and Geography Libraries* 6 (1), 6-25.

Fleet, C. (2008). Life beyond the paper graphic index: Evaluating new geographic retrieval technologies for the future map library. *Liber Quarterly* 18 (2): 255-70.

Johnson, I. (2005) Indexing and delivery of historical maps online using TimeMap. *National Library of Australia Magazine*, 15 (4), 6 pp.

Oehrli M., P. Pridal, S. Zollinger and R. Siber (2011). MapRank: Geographical search for cartographic materials in libraries. *D-Lib magazine* 17.

Přidal, P., and P. Žabička (2008) Tiles as an approach to on-line publishing of scanned old maps, vedute and other historical documents. *e-Perimetron*, 3 (1), 10-21.

schraefel, M.C., M. Wilson, A. Russell and D.A. Smith (2006) mSpace: improving information access to multimedia domains with multimodal exploratory search. *Communications of the ACM*, 49 (4), 47-9.

Van de Sompel, H., M.L. Nelson, C. Lagoze and S. Warner (2004) Resource Harvesting within the OAI-PMH Framework. *D-Lib magazine* 10 (12).

Yakushev, R. (2009) *Using URL Rewrite Module 2.0*. Microsoft Corporation (http://learn.iis.net/page.aspx/664/using-url-rewrite-module-20, accessed 2 July 2012).