PROG
43,1

Map metadata: essential elements for search and storage

Ashley Beamer
Glasgow School of Art, Glasgow, UK

Received 11 September 2008
Revised 23 October 2008
Accepted 31 October 2008

Abstract

Purpose – The purpose of this paper is to develop an understanding of the issues surrounding the cataloguing of maps in archives and libraries. An investigation into appropriate metadata formats, such as MARC21, EAD and Dublin Core with RDF, shows how particular map data can be stored. Mathematical map elements, specifically co-ordinates, are explored as a source of optimal retrieval.

Design/methodology/approach – This paper is based on both the personal experiences of map cataloguers as well as previous literature on map retrieval elements, metadata formats and map retrieval systems.

Findings – The difficulties behind map cataloguing do not lie in metadata file formats but rather in maps themselves, staff and budget. They also lie in the lack of map-appropriate retrieval systems and the lack of co-ordinate search capabilities.

Practical implications – The practical implications of this work reflect the necessity for strong map-retrieval systems and strength of available metadata formats to store essential map data for retrieval. Future map cataloguers should secure appropriate systems for retrieval and include geographical location information, specifically numerical co-ordinates.

Originality/value – This paper provides insight into current issues in map data and the file formats currently used for storing this data. It also investigates current map-friendly systems in use by libraries and archives.

Keywords Libraries, Maps, Databases, Cataloguing

Paper type Technical paper

1. Introduction

As a cataloguer of maps in an archival setting, the author's own experiences have shown that this resource type raises very particular issues. Map materials are notorious for their complicated nature and the difficulties they pose for both cataloguers and systems developers.

Through discussions with colleagues it seems that map cataloguers feel cataloguing and metadata standards are insufficient for creating full map records. However, it may be the case that insufficiencies lay elsewhere. This paper looks at the fundamental reasons behind map cataloguing difficulties and record retrieval. It examines which map data are most important and which electronic metadata formats are sufficient and appropriate for representing this data. Doing so will lead to an increase in both access and interoperability between domains, and will be integral to strengthening user service. This is particularly true for libraries when using a map-friendly OPAC, thereby recognising extensive map access points over traditional bibliographic resources.

The author was formerly the Project Collections Officer at the Royal Scottish Geographical Society. All URLs in the references were checked on 26 September 2008.

Program: electronic library and information systems
Vol. 43 No. 1, 2009
pp. 18-35
© Emerald Group Publishing Limited
0033-0337
DOI 10.1108/00330330910934084

As many of the map collections found in libraries and archives are historical in nature, this paper focuses on paper-based materials rather than modern electronic Geographic Information System (GIS)-based maps and other cartographic materials.

2. Map elements

Firstly, it is beneficial to understand the distinguishing features of the maps themselves and the elements that are important to include in a catalogue record. In addition, the map elements discussed in the following are particularly important, as they are the ones persistently raised by map cataloguers as most problematic. Geographical area coverage is examined in particular. It is worth noting how these characteristics differ from more traditional bibliographic resources.

Overall, map cataloguing often requires some detective work. It is occasionally the case that with early maps there is very little information other than the map image itself. Thus there is often no direct information about the cartographer or publisher, and sometimes no title. It may also be the case that the date is omitted from the map, requiring the cataloguer to obtain an approximate date by comparing locations and map features to other maps where the date *is* recorded. Illustration may also act as a clue. Occasionally the cataloguer can determine the date of the map's publication based on other maps by that publisher (Rockwell, 1999, p. 50).

Map titles can be complicated as there is often more than a single title present. It is generally suggested that the title should represent the "strongest" element on the map and/or the title that best reflects the subject and the geographical location represented should be chosen; however, this is a subjective decision. Furthermore, titles are not always obvious and can be written in obscure locations on the map. As titles often provide map subject information, it is also important to note here that subject access is integral to map searching.

Another problem with titles is that place names change over time. When cataloguing a country (including those no longer in existence), it is difficult to know what name users will search for. Alternate spellings and any newer names should also be entered in the record in an appropriate note field. When not using a gazetteer, geographic locations should be catalogued by what is expected that the user will use to search. Occasionally the location is difficult to determine: the map may represent a country no longer in existence, may be inaccurate or may not state the covered region at all. Thus, describing the area covered is not as straightforward as it seems and, as discussed in the following, it is more accurate to use other methods for electronic geographical cataloguing.

One issue that tends to concern map librarians more than archivists is the issues of map sheet series'. In an archival setting it is most common for the individual map to be entered into the system referenced to the series of which it is a part in another title field if provided. The series may also be entered providing the scope for that series, particularly in cataloguing software such as CALM. Therefore it is possible for the series description to be retrieved as well as each individual map within that series. For libraries, series information can be added in the 4xx field as well as in subfields of 245. This is a particular problem for library cataloguers, however not one which is inspected in great detail here.

2.1 Mathematical data elements

Scale is an essential mathematical map component and is described by the Representative Fraction (RF), or the map distance in relation to ground distance. A map of the UK with a scale of 1:10,000 is a large-scale map and may have to be shown on numerous sheets as one sheet cannot contain all of the information. For interpretation's sake, 1:10,000 translates to 1 km. on the ground to 10 cm on the sheet map. A small-scale map with a scale of 1:50,000,000 could easily fit onto one sheet (Farrell and Desbarats, 1984, p. 22). Scales can be a source of confusion for cataloguers as they are occasionally indeterminable and at other times a map can be drawn at several different scales.

Map scales can vary greatly in the manner they are presented (if they are presented at all in early maps) and can be time-consuming to interpret and convert to a map database's chosen scale format. Online converters are available to alter the scale into the standard factor. Graphic scales may have to be measured before they are converted. If there is more than one scale, the most prominent should be entered first if it can be determined. If the scale is not shown, another similar map (of the same area) can be used to deduce the scale.

Map projections are another mathematical aspect of large-scale maps. As the earth is an oblate spheroid, it is impossible to represent it on a flat surface without presenting a distorted view of at least some part of the mapped area. In order to compensate for this distortion, mapmakers have come up with various projections that allow for a workable understanding of the earth when represented on paper. In a typical equatorial projection for example, the equator will be perfectly represented but towards the poles the features will look proportionately much larger than they actually are. When the map is of a smaller area of the world, projections are not as problematic.

2.2 Geographical area: co-ordinates

Williams (2005), map cataloguer of the National Library of Scotland, argues that cataloguing standards whether bibliographic, generic or archival do not reflect an understanding of how individuals search for map resources. As most map searchers do not know the particular map they are searching for, most bibliographic access points are virtually meaningless. When it comes to features like author, map scale, date, map subject and type one must have a good idea of what they are searching for, for example, a 1950s large-scale ordnance survey map of Lanarkshire. Map geographical coverage on the other hand is generally the map feature that is known above all others.

As every map reflects a spatial location or area, the communication of this information is vital in reflecting map content (Perkins, 1991, p. 224). Locations may be represented by co-ordinates, identified on the spheroidal earth by lines of latitude and longitude, or meridians and parallels. The Prime Meridian runs through Greenwich, England but other longitudinal lines can be used instead. If this is the case, it is recommended that the meridian should be converted to the modern day Prime Meridian and the original meridian mentioned in the notes section. Longitudinal and latitudinal co-ordinates are generally measured in decimal degrees: where the latitudinal degree is up to 90° North or South and the longitudinal degree is between 180° East and 180° West. An alternative, more traditional, system is to use degrees, minutes and seconds: a minute is 1/60th of a degree, and a second is 1/60th of a minute.

In both cases, co-ordinates should be given from West to East and North to South respectively.

The metadata formats investigated in the following are able to cope with the various elements discussed previously. For the purposes of this paper, only tags relating to geographical location and mathematical data are shown.

3. Metadata

The strength of a given archival system or library and information system (LIS) for cataloguing maps is ultimately determined by the accuracy with which it can represent a map's metadata. However, the utility of even a "perfect" map cataloguing system to an external user or institution is only as good as the metadata formats that the system can accept or produce. Therefore it is instructive to examine common archival and library metadata formats and other general metadata formats for the purposes of interoperability as well as data storage.

Metadata is commonly referred to as data about data. It is essentially electronic cataloguing; it provides an electronic record for an item or a group of items. There are many different types of metadata including administrative, preservation and technical metadata, as well as descriptive and structural metadata, which are most relevant to the typical librarian or archivist.

For the purpose of outlining the different metadata file formats, an early map of 1595 has been chosen as an example. The map (which is originally part of an atlas) is given in Figure 1 along with some basic description.

The main goals of most modern metadata are to provide general rules and common relevant fields, for interoperability. International standards are imperative for the ease of exchange of electronic data between institutions. The following standards include the same core information fields, (or define cross-walks between them) and have either native XML schema or provide simple mappings to XML, and therefore promote information exchange. I have chosen to examine:

- MARC21 as a generic standard based on AACR2.
- The Encoded Archival Description (EAD) as a specific archival standard based on the General International Standard Archival Description (ISAD(G)) and compatible with the Rules for Archival Description (RAD).
- Dublin Core/Resource Description Framework (RDF) as a universal metadata format.

3.1 MARC21

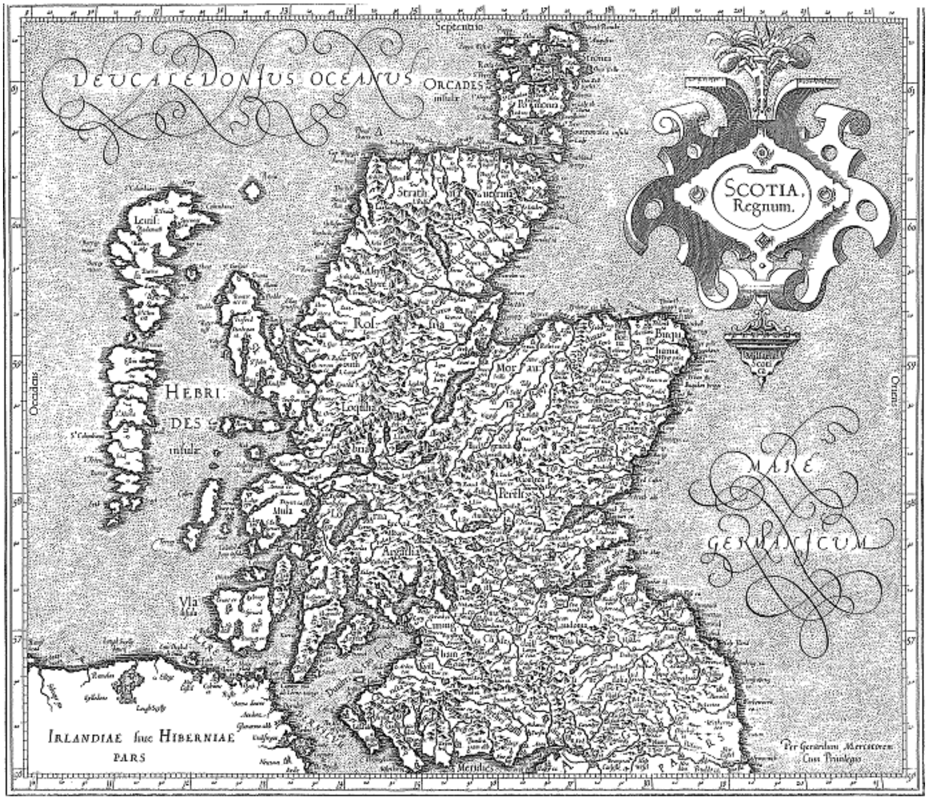
In June 2004 the British Library adopted MARC21 over UKMARC as its electronic cataloguing standard. Several fields are pertinent to map cataloguing and are therefore very relevant when examining MARC21 conversions and mappings to other file formats.

A MARC 21 record for the map might consist of fields as shown in the following (please note that the scale:

034 1#$a a $b 1393920 $d W0030000 $e E0060000 $f N0560000 $g N0620000.

043 $a e-uk-st.

245 $a Scotia Regnum $c Gerard Mercator $h [map].



Map:
Title: Scotia Regnum
Size: 13.75in x 15.75in
Scale: approx 22 miles : 1 inch (taken from Atlantis Pars Altera rather than from the map itself)
Year: 1595
Projection: Mercator
Region: Scotland, UK, N56-62, W3-E6 (Mercator's Prime Meridian, passing through Fuerteventura in the Canary Islands, was at 14 degrees west. The longitudinal coordinates here have been modified to place the Prime Meridian in its modern day location through Greenwich in London.)
Author: Gerard Mercator

Map found in Atlas
Title: Atlantis Pars Altera
Year: 1585
Published in: Duisberg

Source: RSGS (1973, opposite p. 84)

Figure 1.
Mercator Map 1595

255 $a Scale [ca. 1:1,393,920]. 1 inch = 22 miles.; $b Mercator Proj. $c (W 30°–E 60°/N 560°–N 620°).

342 $b decimal degree.

500 $a Mercator's Prime Meridian on the map passes through Fuerteventura in the Canary Islands 14 degrees West of the Prime Meridian. The longitudinal coordinates here have been modified to place the Prime Meridian in its modern day location through Greenwich in London.

In MARC21, co-ordinates may be provided in decimal degrees, as shown previously in fields 034 and 255. The 034 field uses a fixed machine-readable representation in decimal degrees whereas the freetext 255 notes field may represent the area in any format the cataloguer desires. A fuller example showing the use of minutes and seconds is as follows:

[255 $c E 32°30'07" – E 34°30'12"/N 35°30'11" – N 35°00'00")].

Although there may be doubts as to the strength of MARC21 in recording cartographic information, it has been modified to reflect mathematical data important in the extraction of a record.

It is important to consider the title of the map as searchers may search using the terms "Scotland", "Scotia", or "Scotiae" as this is how Scotland is occasionally spelled in early maps. Alternate spellings and other place name issues can be dealt with via subject headings discussed in the following. Notably, the 043 field exhibits the Library of Congress MARC geographic area code, which in this case represents Scotland. Other map related MARC fields are referred to throughout this paper.

Importantly, MARC21 is widely used and can be converted into various other data formats including Metadata Object Description Schema (MODS), Open archives Initiative (OAI), RDF, ISAD(G), SPECTRUM, USMARC, Z39.50, Dublin Core and EAD, and vice versa (Day, 2002).

3.2 Encoded Archival Description

EAD is an XML-based and SGML-based document type definition (DTD) maintained by the Library of Congress and the Society of American Archivists. The EAD metadata scheme is particularly strong with regards to describing structure. The data is structured by nested tags to which attributes may be attached in order to further specify their meanings. Extensible Markup Language, or XML, is a simplified subset of SGML, and the EAD document type definition was written to be compatible with both XML and SGML.

The analogue encoding attribute, ENCODINGANALOG, can be used in the <ead>, <eadheader> and <archdesc> tags to indicate the use of another descriptive encoding system for that element's content. These systems might include MARC, ISAD(G) or Dublin Core. For example, the <eadheader> elements may be mapped to Dublin Core whereas the description of the finding aid for a particular fonds may be mapped to MARC. Furthermore, EAD can incorporate MARC21 information in its attributes thereby strengthening the coding and the information contained within it.

The top-level <archdesc> tag contains an attribute that describes the level of the finding aid. This can be one of: collection, fonds, class, recordgrp, series, subfonds, subgrp, subseries, file, item, and other level. It is here that the highest level must be recorded and described (fonds, class, series, etc.) and all child levels are then described using nested < c > tags (subseries, file, item, etc.) which are numbered based on their level.

EAD has been developed to permit the description of more complex archival materials such as cartographic materials. Using the < materialspec > or material specific labels tag for example, one can add fields pertaining to essential details of the map. These attributes can include scales, projections, co-ordinates and other relevant details. The following EAD script exhibits the data elements from the before mentioned Mercator map example.

< arcdesc level = “collection” relatedencoding = “MARC” >

< did >

< dsc >

< c01 >

< did id = “scotia_regnum” >

< materialspec label = “Mathematical Data” >

< materialspec label = “Scale”

encodinganalog = “255$a” > 22 miles: 1 in < /materialspec >

< materialspec label = “Projection”

encodinganalog = “255$b” > Mercator < /materialspec >

< materialspec label = “Coordinates”

encodinganalog = “255$c” > W3°-E6°/N56°-62° < /materialspec >

< /materialspec >

< /did >

< controlaccess >

< head > Subjects: < /head >

< geogname encodinganalog = “651” source = “tgn” > Scotland, United Kingdom < /geogname >

< /controlaccess >

< /c01 >

< /dsc >

< /did >

< /archdesc >

It is both a strength and a weakness of EAD that it is so general as to have the need of incorporating an outside metadata type to specify elements. Here the

< materialspec > tag shows the encoded MARC21 fields as well as the appropriate mathematical data. The 255 tags are provided so users can search scale, projection, and co-ordinates. If the 034 field was used, each co-ordinate would need to be entered under a separate < materialspec > entry, for example 034$a, 034$b, and so on. The < geogname > tag source attribute specifies the Getty Thesaurus of Place Names (TGN). Place names are discussed further in the following.

3.3 Dublin Core and RDF

RDF and Dublin Core (DC) are two related standards that are gaining acceptance. They are not complete solutions in themselves, but instead each provides part of a complete metadata description language.

RDF defines how resources are described, and DC provides an appropriate set of properties with which to describe those resources. Briefly, RDF describes a resource using "triples" like this:

{subject} – {predicate} – {object}

This can be read as "the {subject} has a property {predicate} whose value is {object}". Both the subject and object can be resources being described, and so RDF can be used to build tree-like descriptions where resources have defined relationships to each other rendering it archive-appropriate.

RDF does not define any concrete predicates and it is up to the user to provide them. The Dublin Core Metadata Initiative (DCMI) provides such a set of terms, and is often used in combination with RDF to provide a standard method of metadata description. The set of 54 DCMI Metadata Terms includes 15 base terms along with specialised versions of each. For example, the "extents" term has specialised counterparts "spatial" and "temporal".

The combination of RDF and DC to form a simple metadata description language is understood by many library and archival systems (Day, 2002). However, the simplicity of DCMI Metadata Terms means that it cannot capture the same level of detail about individual items as can MARC21, and it lacks the archival specificity of EAD. It is not impossible to use only DC, but some institutions choose to introduce some extra terms for their own specific use (Banski, 2002, p. 133). As such the metadata is less suitable for interoperability purposes.

If we create a URI for the sample map:

www.example.org/ScotiaRegnum

the map's mathematical and geographical metadata in RDF and DC can be represented as follows:

< rdf:RDF xmlns:rdf = "www.w3.org/1999/02/22-rdf-syntax-ns#"

xmlns:dcterms = "http://purl.org/dc/terms/" >

< rdf:Description

rdf:about = "http://example.org/ScotiaRegnum" >

< dcterms:title > < rdf:value > Scotia

Regnum < /rdf:value > </dcterms:title >

```
< dcterms:spatial >
< dcterms:box >
< rdf:value >
name = "Scotland, UK";
northlimit = 62; southlimit = 56;
westlimit = -3; eastlimit = 6;
projection = "Mercator"
 < /rdf:value >
 < /dcterms:box >
 < /dcterms:spatial >
 < /rdf:Description >
 < /rdf:RDF >
```

Note that we do not describe the scale explicitly here. The DC Box Encoding Scheme refers to a region of space rather than a map, and so does not explicitly allow a scale to be represented. To include the scale, we could add an additional "spatial" element as follows:

```
< dcterms:spatial >
< rdf:value > Scale = 22 ms:1in < /rdf:value > < /dcterms:spatial >
```

The DC is able to incorporate various syntax encoding schemes. There are several schemes appropriate for use with the "spatial" term, including two specifically provided by the DCMI: the point encoding scheme represents a specific location as latitude and longitude values, along with the name of the location, and the box-encoding scheme (shown in the previous) which represents the four extents of a map from the equator and the Prime Meridian at Greenwich, along with the name of the region. The Treasury Board of Canada Secretariat describes several other suitable schemes that may be incorporated in the DC spatial element to describe location. These include: The Canadian Geographical Names Data Base, Regions of Canada, The ISO 3166 – Country Codes, Canadian Postal Codes, The Getty Thesaurus of Geographic Names (TGN) (TBCS, 2007). Regardless of scheme chosen, a controlled vocabulary (such as the Thesaurus of Geographic Names) is recommended for plain-text location names in order to ensure consistency of spelling and place-name titles.

4. Real issues behind map cataloguing

The three metadata formats chosen are able to deal with the storage of important map elements particularly mathematical elements and place names. Other difficult map elements can also be represented by these formats including creators, date ranges, map subject, separate title information, resources type and item level information for maps in an archival collection.

Fleet (1994), deputy map curator at the NLS, asserts that maps are able to fit into most generic metadata schemes without any major problems. In particular, EAD is, and MARC21 can be, XML-based, which is well supported and allows for simple mapping between different XML-based formats. There are already a number of crosswalks, mapping MARC21 to a variety of common metadata formats. It may therefore be argued that as long as a system understands a format that is mappable to MARC21 or EAD, then information held by that format can be readily migrated to other systems.

XML file formats are also notable because they allow the use of other XML data formats to be integrated through the use of namespaces. Namespace encoding allows XML-based DTDs to be merged into one file, including DC via RDF, MARC XML, ONIX XML, and EAD. In utilising open-source systems, one is able to incorporate a range of XML-based files via namespacing as a means to enhance data storage and management for purposes of future interoperability. Furthermore, following the element and attribute definitions developed by the Conceptual Reference Model provides an increased likelihood for further interoperability.

As early as 2000, the Interoperable Metadata Dublin Core Testbed Project has determined that RDF built on XML is the recommended syntax for encoding metadata for the purposes of interoperability between cultural heritage organisations (Dunn, 2000).

It is possible to map data from one format to another, although not always straightforward as EAD is a hierarchical format where MARC21 is not. Furthermore, some may argue that metadata standards do not reflect map searching as maps have not been at the forefront of their development. This is particularly true of MARC which has been in use for many years and is still being modified for historical resources. Regardless, each of the previous formats are sufficient for cataloguing maps even if it is not necessarily easy to decide which one is best for librarians and archivists both. It is primarily the database systems that do not reflect how users search for maps. The database must be user and map-friendly by providing relevant search fields. Before investigating map-friendly systems, the other pertinent issues behind the difficulties surrounding maps should be discussed.

4.1 Full map records

Aside from the difficulties in extracting map data, other challenges in map cataloguing lie in the lack of full records. The cataloguer must ask "how will an individual search for this map?" or "which map elements are most important in this map?" The extraction of the map data while keeping in mind the nature of a user search is of significant importance for any map cataloguer. Cataloguers dealing with this complicated resource must have the proper training and background knowledge of maps and map searching.

Williams perceives cataloguing not as following a set of rules, but rather as an artform that requires contemplation and testing. It is important to perceive what data the searcher will utilise to chase the item. Williams created the Cartographic/Ownership/Describe/Explain system for map cataloguing, developed to assist cataloguers in extracting relevant data for record creation (Williams, 2005). This includes the cataloguer asking:

- Is the resource Cartographic in nature?
- Is there Ownership of the item?

- How can I best Describe this item? – including mathematical elements.
- How can I best Explain what this item portrays?

According to Williams, although there are areas of MARC21 that are not *perfect* for maps, the standard is generally quite efficient. Problems lie in the lack of trained staff, budget cuts and cost-saving exercises. For example, partial or "bare-minimum/cataloguing is insufficient for map retrieval" (Williams, 2005, 230).

When it comes to maps, cataloguing only a few access points is not sufficient, nor realistic for retrieval. Current pressures on cataloguers to produce shorter more "cost-effective" map records result in lost and useless files. It is necessary to create as full a record as possible thereby increasing the chances of the map record being located, particularly in cases where the searcher has little search data to work with. In a survey of North American map cataloguing libraries, only 26.4 per cent reported that full records were created for at least 70 per cent of the maps needing original cataloguing (Davis and Chervinko, 1999, p. 25). Inputting the correct and full data record once means that users are able to find and utilise the information time and again (Andrew and Larsgaard, 2002). With full records, any retrieval problems lie in the database used. The next section outlines the strengths and weaknesses of a proprietary maps database.

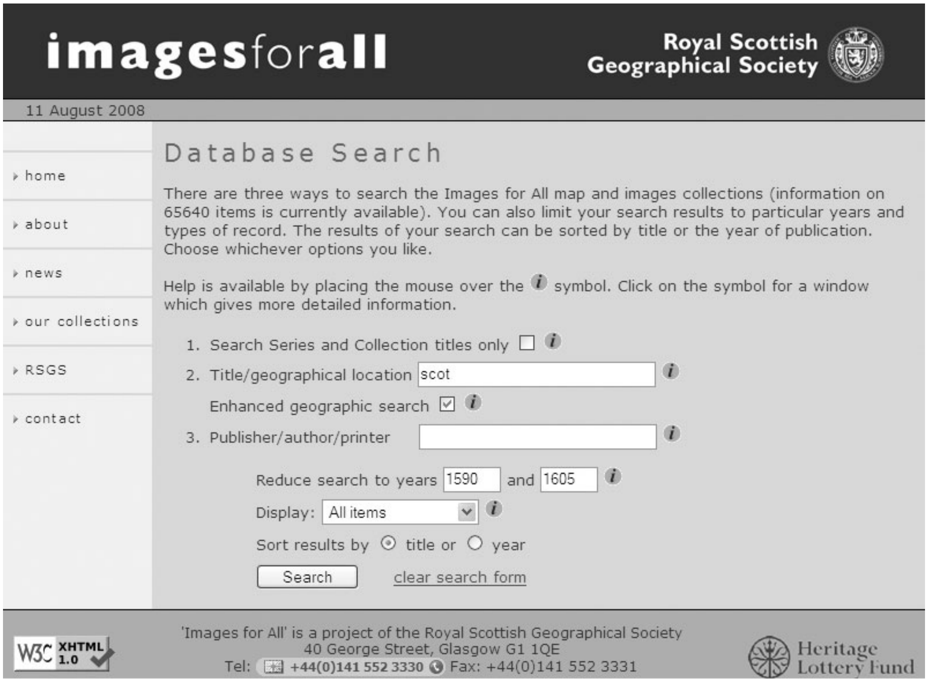
5. RSGS map systems: a proprietary database

The Royal Scottish Geographical Society (RSGS)'s Images for All project has purchased an Oracle database which it has modified according to its specific map retrieval needs. The database allows the archivist to catalogue to series level as well as item level. As it is able to export data in both EAD and MARC, the data can be easily migrated into different systems if, and when, necessary. Furthermore, map donations can be easily imported if they are held in these formats. Although the system is not ideal for map retrieval, it is arguably sufficient for the needs of the Society.

A significant problem with the database is that it is designed to search only particular fields. Therefore separate title or description information entered into the notes fields will not be found by an individual searching for that information. As such, many of the previous metadata elements may be largely useless in a user search.

Figure 2 illustrates several strengths of the system. When searching for the Scotia Regnum map one may use the truncation feature, a date range, as well as an enhanced geographical location search for Scottish maps not containing the words Scotland or Scotia(e).

The RSGS uses the Parsons Classification system to identify geographical locations of maps. This system is used by the UK Ministry of Defence to describe all countries and territories currently or previously in existence. When entering the term "Scotland", the enhanced geographical location search will offer results where that term appears in the title field as well as maps that begin with C18, the Parson code for Scotland. Therefore sub-codes of C18 will also be extracted revealing maps of more specific Scottish areas. The system does not work the other way around and as such will not pull up hits for Scotland when the search term "Edinburgh" is entered. Although this is a strong tool for geographic searching, allowing the searcher to choose numerical co-ordinates allows for the creation of their own specific location search. For example, searching Belgium and Germany in the Parsons Classification system would require



Source: RSGS (2008)

Figure 2.
Sample screenshot of system features

the searcher to search for Europe as a whole resulting in a longer results list. Co-ordinates searching on the other hand can result in a more customised search and is the strongest tool for map retrieval. Unfortunately, the RSGS system is unable to store co-ordinate data as it provides no co-ordinate fields on the administrative or the user ends. A database geared specifically for maps ideally should provide search capabilities for all of the mathematical map elements, particularly co-ordinates. Such coordinate-based systems are discussed in the following.

Although the RSGS database shows the search strengths mentioned previously, it does not feature a thesaurus and therefore searching the term "Scotia" does not retrieve items containing title information, "Scotland". An appropriate thesaurus example is one featured by the Consortium of European Research Libraries (CERL) which produces place names printed between 1450 and 1830 and is therefore useful for maps and other historic materials.

As we shall see in the following, digital map images are also invaluable to user retrieval.

6. System recommendations

A map system must be sufficient on both the administrative side as well as the user end. On the admin side, the system must be straightforward, easy to use and provide all of the appropriate fields for the generation of a full metadata record for maps. From the user side, the system must have the proper map-search capabilities. Most OPAC systems are largely geared towards searching bibliographic resources and are not able, or configured, to represent MARC21 fields created specifically for maps. OPAC systems must be developed to specifically include map retrieval or offer a separate search interface for maps and other cartographic materials. Note: Whenever possible, bibliographic resources can be catalogued to be accessible based on geographic searching. Place names or even co-ordinates can be attached to the records of relevant resources so that searches based on geographical locations retrieve not only cartographic materials but other information sources as well.

As it is often the case that map searchers do not know what they are really looking for until they see it, it is best for catalogued record results to be retrieved with a digitised thumbnail of the map attached. The user might then have the option of clicking on the thumbnail to access a larger JPEG version of the scanned map. A system that allows a zooming function (such as Google maps) after the map has been scanned at a high resolution is ideal. Particularly for archives and for libraries with collections of old maps, the digitisation of early maps means that the original is less likely to be perused by interested users. As custodial institutions such as libraries and archives have a central role in the preservation of the documentary heritage, they are obligated to provide digital surrogates (Feather, 2004, p. 6). These surrogates are a key part of the system and provide access to information while the original is kept safe. Of course it must be considered that many institutions do not have the financial or temporal resources to create digital surrogates of their map collection.

A strong user database provides access to digital surrogates and would:

- be free/open-source (whenever possible/suitable for the institution);
- be created for both archival and library management purposes;
- be metadata-modification capable (i.e. addition of metadata fields not already known to the system);
- be easy to use for both administrative users and end users;
- recognise common XML-based formats such as EAD as well as MARC and Dublin Core;
- be able to store and display digital surrogates, including thumbnails, JPEGs and TIFFs (when feasible); and
- provide appropriate search fields for optimal retrieval. Such fields would include, title, subject (possibly from a drop-list if appropriate), date, date range covering map periods, co-ordinate searching (see the following), place-name searching (see the following), all individuals involved in the map's creation, and corporation searching.

It is the last element where the difficulty in searching such a complicated resource lies. Catalogues are generally designed for searching by author, title, or subject. Database system fields must include the relationship between maps and other resources, as well as maps within other resources. Also, there must be search fields that account for the roles of different individuals and corporations in the making of a map, including surveyor, engraver, author, draftsmen and publisher (which may be accounted for in the administrative side 1xx and 7xx MARC21 fields). Search fields must include geographical location headings, as well as map subject, which can be chosen from a list if the map collection falls under a suitable number of subjects and types. These would include routes and trail maps, ordnance maps, railroad maps, nautical charts, topographic maps, geologic maps, historical maps, political boundary maps, utility maps, road maps and so on. Specific subject names can be acquired from the Library of Congress Subject Headings, the Subject Cataloguing Manual or the Map Cataloguing Manual (Moore and Hall, 2001, p. 20). Fields must include dates for both the date the map was made as well as relevant time periods or date ranges for the period in time that the map represents. This is due to boundary changes around the world and throughout time due to political or natural events. When it comes to maps, co-ordinates and place names must become the point of focus as these elements are where optimal retrieval lies and this is often the most unambiguous information that the user has. Many libraries "are in the process of developing their spatial searching capabilities through the implementation of GIS capabilities, and by providing additional numerical data" (Gonzalez, 2007). The ability to search by co-ordinates is essential.

6.1 Co-ordinates

The ability to search for maps via geographical location distinguishes it from traditional bibliographic access points (Fleet, 1994, p. 227). Furthermore, as other elements are not particularly accurate for optimal searching, mathematical data can be regarded as a strong, unchanging point of reference. They cannot be mis-spelled and do not have different meanings. It should be noted that although this is true, map co-ordinates can be entered mistakenly and standards must be put in place, which are able to deal with other meridians used as the prime as well as alternate co-ordinate systems. It may be rather unrealistic for cataloguers of a large map collection to convert mathematical elements manually. A strong database that allows co-ordinate searching must also understand the different co-ordinate search capabilities including bounding box co-ordinates and the pinpoint, or centroid, co-ordinate. Co-ordinates whether polygonal, box-shaped or a single point or centroid are the strongest tools for located geographic spaces covered by maps. Map cataloguers historically have provided "bounding (box) co-ordinates", or a set of four points that surround some geographic region as shown on a map (Andrew, 2003, pp. 94-95). Quite a number of sheet maps identify the four co-ordinate points representing the interior area that has been mapped. An appropriate system would retrieve items where those points were described, it would retrieve items within the box points and most importantly it would retrieve points within a specified radius of the point(s) searched (Buckland and Lancaster, 2004). With this type of search, particularly one that is able to locate areas within or near specified co-ordinate(s) boxes, there is less need to rely on other data such as titles or even place names. Gonzalez suggests that, "providing the latitude and longitude with complete degree, minute and second information reduces the margin of error and provides for the closest approximation to the true measurement of the area on a map being represented" (Gonzalez, 2007).

Unfortunately, users would not necessarily think to utilise co-ordinates in a search. Co-ordinate-searchable databases can link searchers to instruction pages that provide information about co-ordinates, provide conversion capabilities from decimal degrees to degrees/minutes/seconds (dms) and vice versa. It is also important that the cataloguer choose a particular manner in which to represent the co-ordinates and keep to this standard, such as the decimal degrees or dms.

Another option would be for the system to allow the searcher to choose the required geographic location by clicking on a map that is able to zoom in and define the region, resulting in the retrieval of maps within that region. This type of search is used by The Geological Society, which also provides co-ordinate searching however does not provide digital surrogates.

A difficulty with co-ordinates is that many early maps do not provide the co-ordinates covered. However the cataloguer is often able to infer from the information on the map which region of the world is shown and can provide co-ordinates accordingly. Although the co-ordinates can be calculated manually, this is often a difficult task and can take a great deal of time. Arguably the cataloguer must determine if the record is weak enough to warrant such calculations. Ideally, co-ordinates should always be included in a map record in the case of any future implementation of systems that have such co-ordinate searching capabilities.

EAD and MARC21 (and DC as well) are capable of including co-ordinate information and making this field mandatory will ensure that the map standard is raised and standard systems can be introduced to enhance map retrieval capabilities. Ideally then, the database's administrative side will provide a resource type descriptor so the system knows which fields must be entered.

6.2 Place names

Location information in the form of place names is another of the strongest methods of map retrieval. Cataloguing based on place names should describe the location recorded from the specific to the general, covering the most ground, so to speak, for the highest number of searchers. One problem with location names is that they are subjective. Also as geographical locations have changed dramatically over time the searcher may need a historical knowledge of that location. It is therefore necessary to utilise thesauri capabilities by using an authorised list such as the Library of Congress Geographic Subject Headings or the Getty thesaurus/gazetteer for geographic names.

Geographical subject headings provide chronological subdivisions. Each map represents a particular temporal aspect and therefore subject headings can provide a manner in which to label and search, for example a map featuring routes followed by soldiers in the Civil War can be provided with a "Civil War" subject heading. A gazetteer brings together place names as well as co-ordinates and geographical features. It is possible for online catalogues to be linked to online gazetteers therefore ensuring that searches are optimised. This is recommended as consistency in place-name descriptions is difficult and with a large number of maps, unrealistic. Geographical location, even when using modern place name schemes such as the Getty thesaurus of place names or postal code schemas for a given country, for example, do not lead to the definite retrieval of relevant information.

EAD offers such a means of incorporating thesauri within the < geogname > tag and using the "source" attribute (as shown in the Mercator example mentioned previously). The MARC 6xx$z and 651 fields allow for the usage of the geographic subject headings. The 651 field therefore allows the cataloguer to incorporate different types of thesauri used to describe place names. The second indicator reveals the thesauri choice for example "0" represents the Library of Congress Subject Headings. The number "7" allows a thesaurus not listed to be incorporated into the record (Library of Congress, 2008). Furthermore as seen in the previously mentioned MARC example, the 043 field provides space for the geographical scope note. MARC fields such as 667 allow for previous geographic names to be added; names which may have change due to political events. As we have seen in the previous, DC is also able to embed thesauri schemas while entering relevant place names.

Of course while it is possible to embed references to thesauri or the names of relevant encoding schemes into the previously mentioned metadata standards, this embedded information is only useful if it can be understood and utilised by an appropriate system.

6.3 DIGMAP

The reality of an actual map retrieval system can be quite different from the ideal. Notably, the large-scale DIGMAP project that began in 2006 promotes access to digitised maps. Its testbed stages incorporate the maps from various European countries associated with The European Library (TEL) and TEL's vision for the European Digital Library 2010 initiative (DIGMAP, 2007).

The system can be searched via geographic boundaries and a multilingual geographic thesaurus based on a gazetteer developed for the Alexandria Digital Library and map authority files (Borbinha *et al.*, 2008, p. 2). If the searcher does not know specific search terms, DIGMAP prototypes allow searchers to choose authorship/publisher, date and index terms via alphabetical lists. These lists allow the individual to choose from geographic locations and dates as well as map types and subjects from the thesauri. The DIGMAP thesaurus is made up for both a gazetteer and an authority file, which provides information about the creator. Together these "manage concepts, geographic coordinates, names of places, areas and persons, as also related historical events (with dates or time intervals)" (Borbinha *et al.*, 2008, p. 3). Importantly once a given map is retrieved, various sections of the map may be investigated closely. The primary prototype areas allow the searcher to zoom into the map for clarity and visual inspection. The success of the project comes from its recognition of the relationship between time, place and subject, and how the different combination of these results in map groupings for a successful search.

The key strength of the project is that it is open source in nature thereby providing the ability to access catalogued data and advocate open information exchange for interoperability. This idea promotes the idea of national and international standards of data storage and promotes the development of appropriate metadata systems. DIGMAP uses REPOX, an XML-based infrastructure that manages the metadata subsets (Borbinha *et al.*, 2008, p. 8). By retrieving the desired map and viewing its XML record, one may observe the use of DC as a means of storing the map data. The system also accepts UNIMARC and MARC21 as well as other metadata formats before the DC conversion. Many maps are given <spatial> data thereby allowing future developments to incorporate co-ordinate searching. It also provides map relationships by utilising the DC term <isPartOf>, crucial for map archivists where the relationships between items in a collection are of primary importance. The DC coding also allows the incorporation of creators, contributors, dates, language, shelfmark, and various other elements useful to a search. Searching by date is particularly important in early maps, as this is generally an element that is known and fixed. Importantly, DIGMAP allows uploaded maps to be georeferenced automatically based on both their metadata as well as their image content.

DIGMAP is setting a precedent for map access and is one that should be met by institutions that safeguard early map collections around the globe. The DIGMAP project including its user interface features is currently under active development.

7. Conclusions

The results of this paper do not point to major insufficiencies in cataloguing and metadata formats. MARC21 and EAD are sufficient with regards to the map retrieval elements that can be stored by them, particularly geographic area coverage. Even DC, embedded in RDF, a common format not originally designed for the library cataloguing domain, is a suitable metadata companion to map retrieval and access. Importantly, most modern metadata standards include crosswalks to allow conversion between formats, essential in format elaboration and interoperability. The formats are also able to store relevant information pertaining to digital surrogates. Instead, map cataloguing issues lie with the lack of appropriate map-related systems, the lack of knowledgeable cataloguing staff, and the lack of full map records.

Clearly, some problems in cataloguing maps are in fact due to the complexities of maps themselves. Map data can be either confusingly sparse or overwhelmingly plentiful. Institutions which are primarily focused on other resources, such as books or journals, often do not consider the extra needs of map resources. This is reflected in the lack of map appropriate systems and library OPACs must evolve to broaden their retrieval capabilities. If the system has the correct search fields, MARC21, EAD and DC/RDF are all formats capable of storing and exchanging full map catalogue records, including co-ordinate information.

When the correct database system is implemented, optimal searching capability relies on full map records created by trained professional cataloguing staff; only then is map retrieval and user service enhanced. From that point onwards, the work will lie among map cataloguing professional, to determine how best to extract often obscure, albeit relevant data from this complicated information resource.

References

Andrew, P. (2003), *Cataloging Sheet Maps: The Basics*, Haworth Information Press, New York, NY.

Andrew, P.G. and Larsgaard, M.L. (2002), *Maps and Related Cartographic Materials: Cataloging, Classification, and Bibliographic Control*, Haworth Information Press, New York, NY.

Banski, E. (2002), "Implementation of Dublin Core at the University of Alberta libraries", *OCLC Systems and Services*, Vol. 18 No. 3, pp. 130-8.

Borbinha, J.L., Pedrosa, G., Luzio, J., Manguinhas, H. and Martins, B. (2008), "The DIGMAP Virtual Digital Library", paper presented at the International Workshop on Digital Approaches to Cartographic Heritage in Barcelona, Catalunya, 26-27 June.

Buckland, M. and Lancaster, L. (2004), "Combining place, time, and topic: the electronic cultural atlas initiative", *D-Lib Magazine*, Vol. 10 No. 5, available at: www.dlib.org/dlib/may04/buckland/05buckland.html

Davis, H. and Chervinko, J. (1999), "Map cataloguing and classification: the basic who, what, and where", in Andrew, P. and Larsgaard, M.L. (Eds), *Maps and Related Cartographic Materials: Cataloging, Classification and Bibliographic Control*, The Haworth Press, New York, NY.

Day, M. (2002), *Metadata: Mapping Between Metadata Formats*, UKOLN, Bath, available at: www.ukoln.ac.uk/metadata/interoperability/

DIGMAP (2007), "Discovering Our Past World With Digitised Maps (Webpage)", Programme eContentplus, available at: www.digmap.eu/doku.php

Dunn, H. (2000), "Collection level description – the museum perspective", *D-Lib Magazine*, Vol. 6 No. 9, available at: www.dlib.org/dlib/september00/dunn/09dunn.html

Farrell, B. and Desbarats, A. (1984), *Guide for a Small Map Collection*, Association of Canadian Map Libraries, Ottawa.

Feather, J. (2004), "Introduction: principles and policies", in Feather, J. (Ed.), *Managing Preservation for Libraries and Archives: Current Practice and Future Developments*, Ashgate, Aldershot.

Fleet, C. (1994), "Comparing automated map catalogue systems: a pilot study based in the National Library of Wales", *Program: Electronic Library and Information Systems*, Vol. 28 No. 3, pp. 223-37.

Gonzalez, J.A. (2007), "Problems that arise when providing geographic coordinate information for catalogued maps", *Coordinates, Online Journal of the Map and Geography Round Table*, Series B No. 8, available at: www.sunysb.edu/libmap/coordinates/seriesb/no8/b8.htm

Library of Congress (2008), *MARC21 Format For Bibliographic Data, Field List*, Library of Congress, Washington, DC, available at: www.loc.gov/marc/bibliographic/ecbdlist.html

Moore, S.M. and Hall, L.M. (2001), *Map Cataloguing: Learning The Basics*, available at: www.sunysb.edu/libmap/basics.pdf

Perkins, C.R. (1991), "The automation of map library routines: problems and potential", *Program: Electronic Library and Information Systems*, Vol. 25 No. 3, pp. 223-40.

Rockwell, K. (1999), "Problem areas in the descriptive cataloging of sheet maps", in Andrew, P. and Larsgaard, M.L. (Eds), *Maps and Related Cartographic Materials: Cataloging, Classification and Bibliographic Control*, The Haworth Press, New York, NY.

RSGS (1973), *The Early Maps of Scotland*, 3rd ed., The Royal Scottish Geographical Society, Edinburgh, Vol. 1 (authored by a Committee of the Royal Scottish Geographical Society).

RSGS (2008), *Images For All (Webpage)*, The Royal Scottish Geographical Society, Perth, available at: www.geo.ed.ac.uk/rsgs/IFA/home.html

TBCS (2007), *IMRC – Geographic Coverage Sub-group: Usage Guidelines (Webpage)*, The Treasury Board of Canada Secretariat, Ottawa, available at: www.tbs-sct.gc.ca/im-gi/mwg-gtm/gcs-scg/docs/2005/optional-fac-eng.asp

Williams, P. (2005), "Where do I start? A cartographic cataloguing code", *The Cartographic Journal*, Vol. 44 No. 3, pp. 227-30.

Corresponding author
Ashley Beamer can be contacted at: ashley.beamer@googlemail.com