Opening new doors: recent initiatives in open data at National Library of Scotland

Gill Hamilton
Systems Librarian
National Library of Scotland
George IV Bridge
Edinburgh, EH1 1EW
Scotland, United Kingdom
E-mail: g.hamilton[at]nls.uk

Meeting: **181 — *National libraries and open data: new discovery and access services* — National Libraries**

Abstract:

*In the past 3 years **National Library of Scotland** has undertaken a number of low cost initiatives to open up its metadata to wider audiences. The most significant of these has been the Library's use of the social media sites **Flickr** and **YouTube** to promote access to materials from its film, photographic and digitised collections.*

Taking what has been learned from the Library's interaction with social media and the benefits it affords; it is now building on these initiatives and working towards releasing metadata both as open data and as linked open data.

Introduction

Over the past 3 years **National Library of Scotland** has undertaken several low cost initiatives to expose its metadata to wider audiences. The Library has adopted the use of the social media services **Flickr** and **YouTube** to open up its metadata and resources and thereby achieve its strategic goals of providing wider access to collections to a broader, *non-traditional* audience.

Building on this experience the Library is currently working in collaboration with the **Open Knowledge Foundation** to explore how to prepare and publish some of its metadata as open data. By licensing the metadata as **Creative Commons CC.0** the Library hopes that others may re-use and re-mix it with other data sets to develop new services.

The Library is also investigating how it might develop its metadata in to linked open data. It has found that there are many challenges to creating linked open data including; development of appropriate skills, discovering URIs and understanding associated

technologies. However, some of these challenges can be overcome by working in collaboration with other like-minded organisations and groups.

Opening up the Library's metadata

Providing access to metadata beyond its traditional library catalogue is not new to **National Library of Scotland**. In the past the Library has provided access to, and shared MARC records with many services and organisations. In the late 1970s to early 1990s the Library hosted **SCOLCAP**[1], a shared cataloguing environment for Scottish libraries. Many SCOLCAP members benefited from the volume of metadata created by the Library enabling then to copy catalogue much of their collection rather than undertake original cataloguing. From the 1980s onwards the Library extended its sharing of metadata and delivered MARC metadata to such services such as **COPAC** [2] (a union catalogue of UK & Irish academic and special libraries), **SUNCAT** [3] (a union catalogue of serials of the UK research community), **WorldCat**[4] (the **OCLC** union catalogue of libraries from all over the world) and to other libraries and users on a one-at-a-time [5] basis via the Library's catalogue and z39.50 interface[6].

The **NLS Strategy 2008-2011**[7] had as one of its four core themes the aspiration of "*Widening access to knowledge*" and set Library staff the challenge to "*unlock knowledge so that is as widely available as possible to users and potentials users*". Its specific strategic goals included;

> *4.2 We will make a wide range of digital resources accessible through our website, in virtual learning environments and through other communications media which are effective in promoting the use and discovery of collections.*
>
> *4.3 We will improve remote access to our collections and expertise by responding to user enquiries, by finding new ways for our users to interact with us, and each other, to share their expert knowledge and by lending more items to libraries, museums and cultural bodies for display.*

[1] The Scottish Libraries Cooperative Automation Project (SCOLCAP) by Bernard Gallivan IN Catalogue and Indexing, no.98-99, Autumn/Winter 1990, p5-8.
[2] http://copac.ac.uk/about/ http://copac.ac.uk/search/
[3] http://www.suncat.ac.uk/description.html http://suncat.edina.ac.uk/F/?func=find-b-0
[4] http://www.worldcat.org/whatis/default.jsp http://www.worldcat.org/
[5] http://www.nls.uk/catalogues/guidance
[6] http://www.nls.uk/catalogues/z39-50
[7] http://www.nls.uk/about-us/corporate-documents/nls-strategy-2008-2011

The Library's **Digital Collections** team, which comprises digital collection management, digitisation and library systems, recognised that there were opportunities for delivering on these strategic goals by extending access to the Library's digitised collections to services beyond the Library's traditional website and catalogues. The team understood that by exploiting social media services such as **Flickr** [8]and **YouTube** [9] the Library could expose its digital collections to new and untapped audiences. However to achieve this it would be necessary to open and expose library metadata and resources more broadly than had been previously considered, taking the metadata beyond the library and information services environment. The Library identified that to gain the benefits of social media services issues of control, rights and maintenance would need to be addressed.

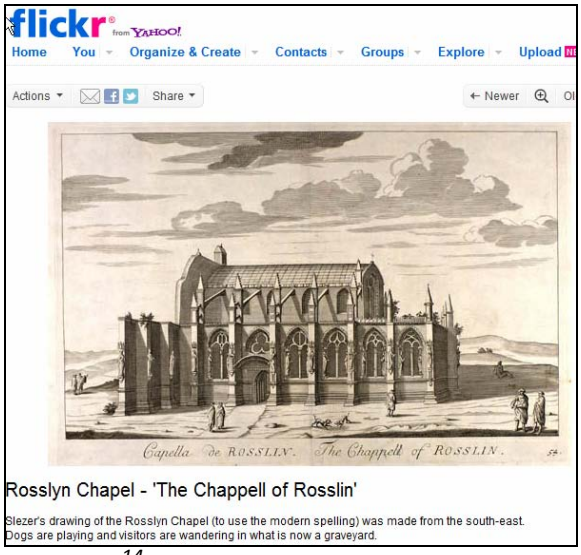
Opening data to social media

In 2008 the Library undertook an exploratory exercise with the social media services **Flickr** and **YouTube** to gain an understanding of the requirements, issues and opportunities for extending access to its collections to new and different audiences.

For the Library's **Flickr** account[10] a selection of approximately 40 examples from the Library's **Digital Gallery** [11] of digitised resources were identified for inclusion. These included digitised photographs, posters, maps, manuscripts and images from books. Short clips of about 15 digitised films from the Library's **Scottish Screen Archive** [12] were also highlighted for inclusion in the Library's **YouTube** account[13]. In both cases these sample resources and their associated metadata were loaded to **Flickr** and **YouTube** in a non-automated manner. The images and video were loaded one-at-a-time and the metadata was copied and pasted from the Library's databases.

The exploratory work with **Flickr** and **YouTube** was invaluable in exposing benefits and issues associated with engaging social media.

[8] http://www.flickr.com/
[9] http://www.youtube.com/
[10] http://www.flickr.com/photos/nlscotland/
[11] http://digital.nls.uk/
[12] http://ssa.nls.uk/
[13] http://www.youtube.com/user/NLofScotland



*Example [14] of a digitised print from National Library of Scotland's **Flickr** account*

*Example [15] of a film clip from National Library of Scotland's **YouTube** account*

Sustainability

An immediate finding from exploratory exercise was that the manual method used for the initial load of items to **Flickr** and **YouTube** was unsustainable and the Library quickly realised that if it wanted to expose large collections then it would need to find a more efficient and effective method for doing this.

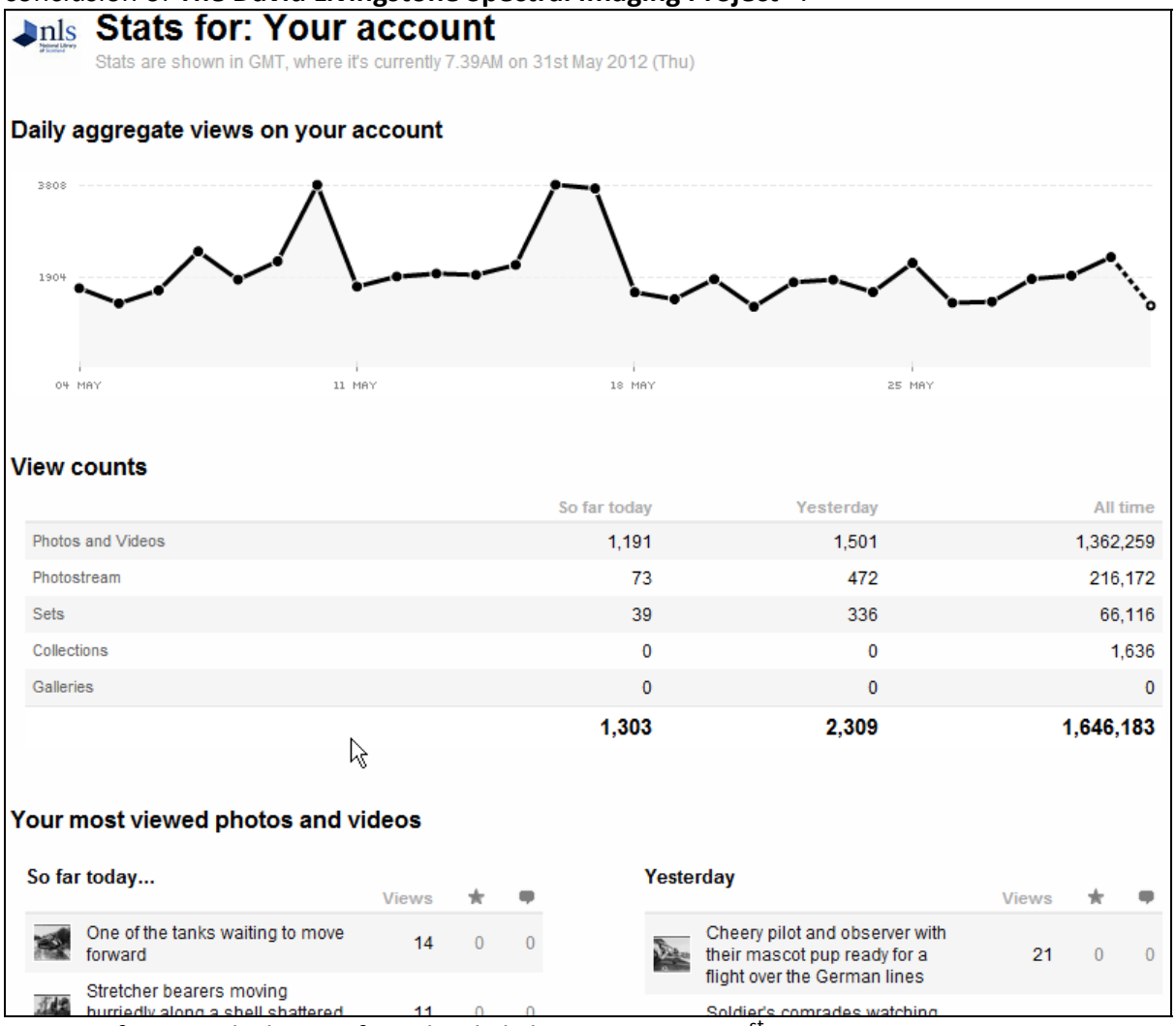
Flickr has an API [16] and the Library chose to develop a basic application based on the **Flickr API** [17] to automatically load both metadata and images. This enabled large collections such as the 2,000 images and associated metadata from the *World War 1 official photographs*[18] to be loaded rapidly without additional human intervention. Unfortunately, at the time, there was no equivalent API for **YouTube** so clips and their metadata continued to be loaded manually in small batches. A **YouTube** data API[19] has since been developed and the Library may seek to exploit this in the future.

Usage

To date the Library's **Flickr** account has seen 1.6 millions views on just 2,500 images, averaging about 1,500 views per day of the collection. The most popular image by far is *Soldiers watching him as he sleeps*[20] with more than 51,000 views. Occasionally there are daily peaks in usage, for example on **Armistice Day** on 11 November each year we see an increase in views of the *World War 1 official photographs* and in May 2012 we saw a peak

[14] http://www.flickr.com/photos/nlscotland/5372742544/in/photostream/
[15] http://www.youtube.com/watch?v=h3KDVbKU7is&feature=plcp
[16] http://en.wikipedia.org/wiki/Application_programming_interface
[17] http://www.flickr.com/services/api/
[18] http://www.flickr.com/photos/nlscotland/sets/72157624150609895/
[19] https://developers.google.com/youtube/
[20] http://www.flickr.com/photos/nlscotland/3012796098/

which we concluded may have been as a result of an announcement from **UCLA** about the conclusion of **The David Livingstone Spectral Imaging Project**[21].



Extract of National Library of Scotland Flickr statistics on 31st May 2012

On the Library's **YouTube** account there have been just under 300,000 views on 98 videos, the most popular being *Gas mask drill at Glasgow primary school*[22]. Like **Flickr**, **YouTube** also sees peaks in activity. For example **The Guardian** newspaper in its article *The coldest winters in the UK*[23] referred to the *Snow blizzard in Scotland 1947*[24] film and this saw views of the clip increase dramatically. The film continues to be one of the most popular in the Library's **YouTube** collection.

Peaks in activity are also caused by the Library highlighting and promoting items in **YouTube** and **Flickr** via its website news[25], email newsletter[26] and **Facebook** [27] and **Twitter** [28] accounts.

[21] http://livingstone.library.ucla.edu/
[22] http://www.youtube.com/watch?v=Ck1mxPsmz6M
[23] http://www.guardian.co.uk/uk/2010/jan/05/coldest-winters-britain-snow
[24] http://www.youtube.com/watch?v=wcv1mrL2qmo
[25] http://www.nls.uk/news
[26] http://www.nls.uk/news/email-newsletter
[27] http://www.facebook.com/NationalLibraryOfScotland

The level of usage on **Flickr** and **YouTube** far outstrip the equivalent usage of these items on the Library's websites and thereby demonstrates that the Library has gone some considerable way in achieving its strategic goal of providing wider access to its collections to a broader audience.

Social media users tagging and commenting

Social media services have also lead the way in enabling their members to leave feedback in the way of comments and "tags". On **Flickr**, users have to date left over 400 comments on 248 items and tagged hundreds of items. **YouTube** has received 178 comments but no tags as it does not have this facility.

Comments include, for example, reminiscences, questions about the content or information about the content and sometimes corrections to how the Library has described items. On **Flickr**, the tags are particularly interesting because users often add tags which enhance the Library's metadata. For example in *Bullets from a German anti-tank rifle and a British rifle, France, during World War I*[29] the Library both described and tagged the photograph as containing a bullet. A **Flickr** member pointed out in a comment that it was in fact a cartridge, and then went on to add appropriate tags to the record. Potentially information derived from such comments and tags could be used to update and enhance the Library's source metadata however there are not currently resources in the Library to undertake the necessary checks for accuracy and updates to the metadata.

dbaldu (9 months ago)

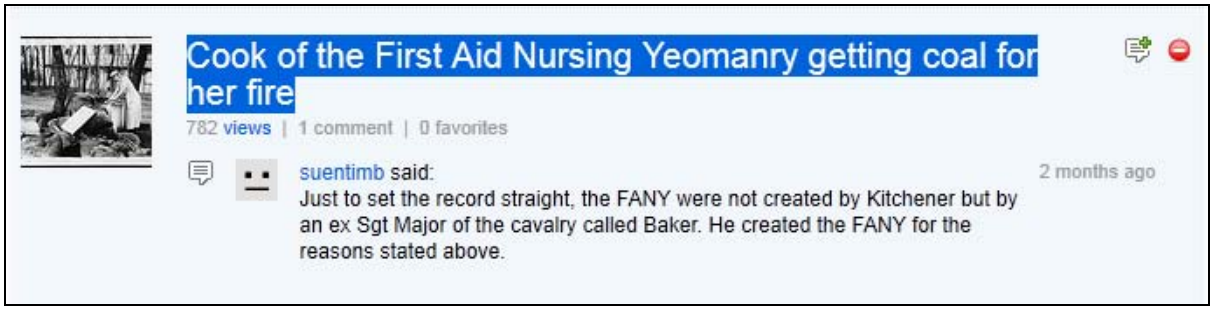
The cartridge (not bullet) on the left is the .303 British. The cartridge was used by British forces in rifles and machine guns from 1888 until it was supplanted by the 7.62 NATO cartridge in the 1950s. Also, the spelling of the German should be Panzerbuechse. (The "u" should have an umlaut, but "ue" is a correct substitute.) I don't seem to be able to create an umlaut in Flickr.

Tags (add a tag)

Photographic prints • France • Black-and-white prints (photographs) • Propaganda • War photography • Photojournalism • Bullets • World War, 1914-1918 Campaigns Western Front • ww1 • wwi • world war one • great war • 1914-1918 • great • war • The Great War • World War I • First World War • Short Magazine Lee Enfield • SMLE • Lee Enfield • Lee Enfield Rifle • Rifle Cartridge • Rifle Cartridges • Cartridge • Cartridges • .303 • .303 British • Ammunition • Mauser • Mauser 1918 T-Gewehr • Mauser M1918 • Mauser T-Gewehr • T-Gewehr • Anti-Tank Rifle • 13.2mm • 13mm TuF • 13.2 mm TuF • .303 Enfield • Mauser Tank-Gewehr • Tank-Gewehr • Panzerbüchse • Panzerbuchse

Examples of comments from a **Flickr** user that corrects and enhances the Library's description of *Bullets from a German anti-tank rifle and a British rifle, France, during World War I*[22]

[28] http://twitter.com/#!/natlibscot
[29] http://www.flickr.com/photos/nlscotland/4700028173/

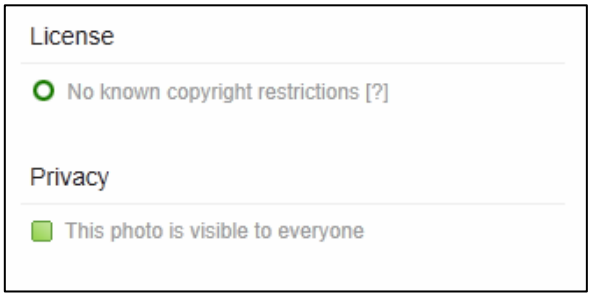


Example of a comment from a **Flickr** member that corrects and enhances the Library's description of *Cook of the First Aid Nursing Yeomanry getting coal for her fire*[30]

Licensing, copyright and rights

During the exploratory exercise the Library decided that it would only include items in **Flickr** and **YouTube** where the items were; out of copyright, had copyright waived or the Library held the copyright and had chosen to waive it. This decision was taken for practical reasons in that it is fairly straightforward to identify materials out of copyright by nature of their age. Furthermore, the Library was aware that images and films made available in **Flickr**, **YouTube** and similar services are often re-used and copied by their members. By supplying images and metadata free from copyright the Library went some way to protect others from inadvertently, or otherwise, violating copyright.

While developing its collection on **Flickr**, the Library became aware of **Flickr The Commons**[31] which had initiated by **Flickr** and **Library of Congress** to enable cultural heritage and knowledge organisations to declare images as having "*No known copyright restrictions*" and thereby stating clearly that the images moved to the public domain. The Library decided to participate in **Flickr, The Commons**. **Flickr** required the Library to prepare a rights statement[32] and sign an agreement of participation. Once the formalities were concluded all the Library's **Flickr** collections were converted from "*Some rights reserved"* to "*No known copyright restrictions*" and the Library joined the Commons community. [33] By placing resources and metadata in to the public domain in this way, the Library enables a global audience virtually unrestricted rights to use its resources and metadata. To date, **YouTube** has no equivalent of The Commons.



Open but not open enough

From its involvement with **Flickr** and **YouTube** the Library has learned of the benefits such services bring in helping achieve strategic goals for extending access to its collections and engaging with a wider audience. It has also helped the Library understand and address issues associated with rights on both its resources and metadata. However, although the

[30] http://www.flickr.com/photos/nlscotland/4699780031/
[31] http://www.flickr.com/commons/
[32] http://www.nls.uk/copyright/index
[33] http://www.flickr.com/commons/institutions/

Library has "opened" its resources and metadata up to these services, the services themselves remain "closed". For example the metadata supplied by the Library to **Flickr** is not made readily available by **Flickr** to others to use and build services.

To enable the Library to continue to build on its successes of extending and broadening access to its collections it became apparent that there should be engagement with the open data movement and explore how to publish metadata openly so it could be freely used by others for any purpose.

Open data

Open data [34] is a concept that data is made available in a way so that others can use and re-use it without restrictions from copyright and other control mechanism. As already mentioned the Library has been making its metadata available for many years however restrictions have been applied to the metadata such as; the one-at-a-time use of the Library's MARC records and metadata and resources being held in "non-open" environments such as the Library's website, **Flickr** and **YouTube**.

The Library has come to recognise that, as an organisation which is chiefly funded by tax payers it should, where possible, make its resources, data and metadata as widely and openly available as possible. Metadata sets such as the Library's main catalogue cannot be readily made available as open data because many records are derived from services that, to varying extents, license their metadata. For example records sourced from OCLC, The British Library and other libraries often have licensing restrictions associated with them. However, the Library is in a position to release metadata as open data for collections for which it undertakes original cataloguing. In these instances the Library creates the metadata and can therefore determine how to license that metadata. Furthermore, originally catalogued collections are perhaps amongst the most interesting and valuable to be released as open data as they often represent items that are unique to the Library. Originally catalogued items include materials like manuscripts, items that the Library has prioritised for description because of their Scottish content such as digitised maps of Scotland and digitised photographs of Forth Bridge construction and so on.

With advice and guidance from the **Open Knowledge Foundation** (OKF)[35] the Library has decided to release metadata from several originally catalogued collections as open data. The **OKF** advised that to achieve openness the metadata should be licensed, structured, documented, published and publicised. The OKF also advised that we should not overly fuss in the preparation of the data but rather "*Give us the data raw, and give it to us now*."[36]

Licensing open data

With regard to metadata licensing, the Library has an unwritten policy that will be formalised later in the year, that states that metadata wholly created by Library will be made available under **Creative Common 0 Universal - Public Domain Dedication** [37] (CC.0). This unwritten policy was developed by following the leadership taken **The British Library** in

[34] http://en.wikipedia.org/wiki/Open_data
[35] http://okfn.org/
[36] http://blog.okfn.org/2007/11/07/give-us-the-data-raw-and-give-it-to-us-now/
[37] http://creativecommons.org/publicdomain/zero/1.0/

publishing the **British National Bibliography** as linked open data under CC.0 [38] and **Europeana** with its CC.0 data exchange agreement. [39]

The Library will publish as open data the metadata sets pertaining to its digitised images of the *Construction of the Forth Bridge*[40] and the *World War 1 official photographs*. The datasets will be published by the end of June 2012.

Structured open data

By its nature all of the Library's metadata is structured. The two datasets that will be published as open data in June will initially be published as locally defined **Solr XML**[41] (see below). The Solr XML underpins the Library's Digital Gallery and is therefore to hand and requires no further development prior to release as open data. Ideally the Library would seek to release its metadata in a standard schema however this would take time to develop. And the Library has chosen to follow the **OKF's** motto "*Give us the data raw, and give it to us now*." The Library will consider republishing the metadata sets in standard schemas at a later date and may use **RDF**[42], **DC terms**[43] and **MODS** [44].

```
<field name="id">74570328</field>
<field name="projecttitle">Scottish Bridges</field>
<field name="title">Queensferry cantilever from end of approach viaduct
<field name="titlecontext">Scottish Bridges > Forth Bridge illustration
<field name="description">Photograph of the Queensferry cantilever from
<field name="sequenceno">12</field>
<field name="width">58.0</field>
<field name="height">43.0</field>
<field name="form_genrecontext">Images>Pictures>Photographs>Photographi
<field name="place">Forth Bridge</field>
<field name="placeTGNLatitude">56.1167</field>
<field name="placeTGNLongitude">-1.0833</field>
<field name="placeTGNType">bay</field>
<field name="subject_content">Box beams</field>
<field name="subject_content">Bridges (built works)</field>
<field name="subject_content">Cantilever bridges</field>
<field name="people_organisation">Forth Bridge (South Queensferry, Scot
<field name="yearType">Date created</field>
<field name="year">1887</field>
```

Extract of Solr XML for *Queensferry cantilever from end of approach viaduct*

Imageimage of *Queensferry cantilever from end of approach viaduct*[45]

Documentation and open data

OKF stresses that it is important to document the metadata sets as they may be used by others outwith the library and information domain. Preparation of the documentation has been a simple exercise to complete because the Solr XML was previously described and documented for **Serials Solutions** as part of the Library's projects to integrate Digital Gallery resources in to the Library's **AquaBrowser**[46]and **Summon**[47] resource discovery services.

38 http://www.bl.uk/bibliographic/datafree.html#lod
39 http://creativecommons.org/weblog/entry/29133
40 http://digital.nls.uk/scottish-bridges/pageturner.cfm?id=74464117
41 http://en.wikipedia.org/wiki/Solr
42 http://en.wikipedia.org/wiki/Resource_Description_Framework
43 http://dublincore.org/documents/dcmi-terms/
44 http://www.loc.gov/standards/mods/
45 http://digital.nls.uk/scottish-bridges/pageturner.cfm?id=74570328
46 http://discover.nls.uk/
47 http://nls.summon.serialssolutions.com/

Publishing open data

OKF recommends that the Library should publish its metadata sets to **The Data Hub** [48]. There are already 76 bibliographic datasets available on The Data Hub including The British Library BNB, CERN library, Project Gutenberg, Cambridge University Library and Europeana.

Publicity for open data

OKF suggests that once the metadata sets are published that the Library should blog, tweet and talk about its work so that others may come to learn about the data sets. The Library has agreed to collaborate on the writing of an **OKF** blog entry about the Library's experience of preparing its first datasets as open data.

I hope to report on progress with publishing metadata as open data at the IFLA 2012 Congress in Helsinki.

Linked open data and the Library

As highlighted in the *Review of the evidence of the value of the linked open data approach: final report to JISC* [49] linked open data (LOD) is currently a set of maturing approaches and technologies and many organisations lack or have limited skills, knowledge and experience to develop LOD. In 2011 this was very much the case in the Library; we had a broad understanding of LOD but lacked the skills to prepare metadata as linked open data. To address this, in spring 2011 colleagues and I undertook exploratory work to; develop our understanding of LOD, learn the basics of RDF, test what we learned by developing RDF records serialised as XML and then identify areas in which we needed to develop further.

Myself and Miriam Kaulbarsch, a placement student studying for a BA in Library Management at **Fachhochschule Potsdam**, went through an iterative learning process to develop several RDF XML files based on a few metadata records from our collection of digitised *World War 1 Official Photographs*. The steps we took were:

- learn basic RDF from the **w3schools.com** [50]
- take a single full metadata record and attempt to match its attributes to properties from **DC terms**. For example match the **title** column from the database to the **dc:title** property and then express the record as a set of RDF triples using those properties. (see extract below)
- identify and record what metadata attributes that could not be matched to **DC terms** for later consideration
- manually match the attribute values to their URIs. For example match subjects, names, languages against familiar library namespaces such as the **Virtual International Authority File** [51] (VIAF), **Library of Congress Subject Headings** [52](LCSH), **MARC relator codes** [53] and **MARC language codes** [54] in order to find the

[48] http://thedatahub.org/
[49] http://ie-repository.jisc.ac.uk/559/1/JISC_Linked_Data_Review_Oct2011.pdf
[50] http://www.w3schools.com/rdf/default.asp
[51] http://viaf.org/
[52] http://id.loc.gov/authorities/subjects.html
[53] http://id.loc.gov/vocabulary/relators.html
[54] http://id.loc.gov/vocabulary/languages.html

corresponding URIs. For example the LCSH subject heading for **fur coats** became *<dc:subject>*[*http://id.loc.gov/authorities/subjects/sh85052491*](http://id.loc.gov/authorities/subjects/sh85052491)

- manually match more attribute values to their URIs against namespace outwith the library domain such as **GeoNames**[55] and xsd:dateTime[56]
- finally we submitted our test records for external review & critique to **Gordon Dunsire**[57]an independent consultant with some expertise in this area

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
    <rdf:Description rdf:about="dod:74548028">
        <dct:isPartOf rdf:resource="dod:74462370"/>
        <rda:titleProper xml:lang="en">Drivers of the First Aid Nursing Yeomanry
        <rda:summarizationOfTheContent xml:lang="en">Dressed in a rather exotic
        <rda:cataloguersNote xml:lang="en">The photograph's original caption rea
        <rdf:Statement rdf:value="D.725" />
        <rdarole:photographer>John Warwick Brooke</rdarole:photographer>
        <dct:subject rdf:resource="lcsh:sh85052491#concept"/>
        <rda:contentType rdf:resource="rdaContentType:1014"/>
        <rda:mediaType rdf:resource="rdaMT:1003"/>
        <rda:carrierType rdf:resource="rdaCarrierType:1018"/>
        <rda:colourOfStillImage rdf:resource="rdaCollStillImage:1001" />
        <nls:recordCreation>
            <rdf:Bag>
                <rdf:li rdf:resource="creatorName:Flora Lee" />
                <rdf:li rdf:resource="creatorID:18" />
                <rdf:li rdf:resource="creatorRole:descriptive metadata record au
                <rdf:li rdf:resource="http://purl.org/dc/terms/created#07/02/200
            </rdf:Bag>
        </nls:recordCreation>
        <rda:termsOfAvailability>Creative Commons 2.5 Scotland</rda:termsOfAvail
        <rda:publishersName rdf:resource="viaf:132075128"/>
        <dct:rightsHolder rdf:resource="viaf:132075128"/>
        <rdarole:collector rdf:resource="viaf:88873527" />
    </rdf:Description>
</rdf:RDF>
```

Extract of RDF for *Drivers of the First Aid Nursing Yeomanry in their fur coast*

[55] http://www.geonames.org/
[56] http://www.w3.org/TR/xmlschema-2/#dateTime
[57] http://www.gordondunsire.com/

Image of Drivers of the *First Aid Nursing Yeomanry in their fur coats*[58]

We continued through several more iterations exploring and learning;

- what attribute values were not appropriate to have URIs. For example the title was not appropriate to have a URI assigned because the title of a digitised photograph is in fact the label of the photograph
- what attribute values were appropriate to have URIs but for which there were no namespaces. For example we had metadata about who had catalogued an item and if we required to record this in the RDF we realized we should establish an National Library of Scotland namespace for cataloguers' names and their associated URIs
- more appropriate properties from other library namespaces such as the IFLA published **ISBD namespace** [59] and the Joint Steering Committee [60] published RDA namespace[61]. The properties in these namespaces give, in some instances, improved specificity over the properties of DC terms.

Outcome of LOD learning process

From undertaking this exercise we learned many things about preparing metadata as LOD however we still had many questions about how to take things forward.

We had learned that the Library's data was well suited to creating rich RDF triples because the Library follows international standards for description and uses where it can international standards for vocabularies (LCSH, TGMII, LCNAF/VIAF). Also we understood that we could mix properties from different namespaces to suit our needs and did not need to constrain ourselves to a single namespace. For example we could mix DC terms with ISBD, RDA, XSD. We realised that as a result of this exercise we had the beginnings of a metadata application profile for some of the Library's digitised collections and finally we knew we could export metadata from Library databases and transform it into RDF.

[58] http://digital.nls.uk/first-world-war-official-photographs/pageturner.cfm?id=74548028
[59] http://iflastandards.info/ns/isbd/
[60] http://www.rda-jsc.org/
[61] http://rdvocab.info/

What we had still to learn was how to automate the process of discovering URIs from literal attributes values. For example how could we "learn" the LCSH URI for LCSH term **Fur coats** without having to look it up manually? The Library's digital collections are sizeable and we do not have resources to undertake URI resolution by hand. Furthermore, we understood that having "learned" a URI either by manual or automatic methods then we should record it to enable its re-use but the existing structure of the Library's databases would have to be modified to hold this new data. Finally, we knew that the final goal would be to publish LOD but we still had no idea what this entailed.

Moving forward with LOD

The issue with automating the discovery of URIs was addressed by a chance meeting of the **Free Your Metadata**[62] team at the **DCMI conference** in September 2011. **Free Your Metadata** is collaboration between the **Multimedia Lab of Ghent University**[63] and **MaSTIC** at **Université Libre de Bruxelles**[64]. As part of their research the team seeks to understand the barriers to cultural heritage organisations publishing LOD. They develop tools, techniques and teaching and training aids to help organisations overcome these barriers. From discussions with **Free Your Metadata** we learned that the **Google Refine** [65] tool has a URI reconciliation service whereby a text literal can be compared against a specified namespace and return corresponding URIs for matches. For example **Google Refine** could be instructed to send the text string **Fur coats** to the LCSH namespace and it would return the URI **sh85052491**. This went some way in helping us understand how we might automatically "learn" URIs. Having received assistance and advice from **Free Your Metadata** the Library in turn was able to reciprocate and gave the team a dataset, licensed as CC.0 so that they could freely use it to educate and demonstrate to others how structured metadata could be transformed into linked open data. The Library still has some issues to address with URI resolution and hopes to continue its collaboration with the **Free Your Metadata** team.

The collaboration with **Free Your Metadata** helped the Library understand that it is a "*data rich*" organisation. And while the Library still lacks the full compliment of skills to develop its own LOD it can actively participate in the LOD movement as an open data publisher and provider of data to others. We have found that organisations and groups such as the **BBC**, the **The Scottish Government**, **Europeana** and **Free Your Metadata** are very interested in using and repurposing the Library's metadata. In exchange they are prepared to share their outputs.

The **BBC** is currently undertaking a prototype service, the **Digital Public Space**[66], which will test if it is feasible to build a service that will give members of the British public access, in a single service, to the nation's cultural, heritage and social assets online. The Library has contributed a subset of the metadata that forms **Scottish Bibliographies Online**[67]. In

[62] http://freeyourmetadata.org/
[63] http://multimedialab.elis.ugent.be
[64] http://mastic.ulb.ac.be/
[65] http://code.google.com/p/google-refine/
[66] http://www.bbc.co.uk/blogs/bbcinternet/2011/10/digital_public_space_idea.html
[67] http://www.nls.uk/catalogues/scottish-bibliographies-online

exchange the BBC has involved the Library in the project and invited us to comment on the application profile it is developing.

The Library has a *Memorandum of Understanding* with **The Scottish Government** to collect, provide access to, and preserve all government electronic publications. The Library has recently been approached by colleagues at the Government asking that we collaborate as they are interested in the metadata associated with these electronic publications. They want to explore conversion of the metadata to linked open data so that it may be linked with other government data. In exchange, our Government colleagues will offer advice about publishing LOD and offered space on their triple store.

It is obvious that the Library has some way to go before it is in the position to prepare and publish LOD in an efficient manner. However, from the experience we've gained from our modest internal research and our involvement with other interested and expert organisations we hope that in due course we will be in a position to publish National Library of Scotland linked open data. In the meantime, as a first step the Library will continue its work in publishing metadata sets as open data.

In conclusion

From its experience with social media, the Library has learned that by exposing metadata and resources to non-traditional services it can deliver its collections to a broader audience. At this stage of the Library's development with open data and LOD it is not yet possible to give determine if the use and re-use of its metadata will have a similar impact. The Library also does not yet have evidence of the development new services or new relationships/links with other data sets. However, despite this the Library is still committed to releasing as much of its metadata as possible as open data and linked open data and to that end, is a signatory[68] to the JISC open metadata principles[69].

Author
Gill Hamilton is Systems Librarian at **National Library of Scotland**. Her specific areas of interest and expertise are in metadata transformation and integration, resource discovery and process/workflow development. Gill was previously a committee member and Information Coordinator for the **IFLA Information Technology** section (2002-2009).

[68] http://discovery.ac.uk/businesscase/signatories/
[69] http://discovery.ac.uk/businesscase/principles/