

Anyone can say Anything

Improving library data with linked data

Owen Stephens
May 2015

I'm Owen Stephens and I am a librarian. I do consultancy which is usually related to how libraries can use IT to deliver great services.

In the world of linked data it is often said that "anyone can say anything about anything". This is both a huge challenge and opportunity for libraries wanting to exploit linked data. This talk will explore the 'open world assumption' of linked data, how it might benefit libraries and what approaches will allow libraries to take advantage of published linked data while trying to avoid problems caused by data of variable quality and veracity.

“With the Internet, we
each have our own
printing press”

Holbert, G.L. (2002). [Technology, libraries and the Internet: a comparison of the impact of the printing press and World Wide Web](#). E-JASL, 3(1-2).

What's the amazing thing about the web? What makes it different to what has gone before? One aspect is clearly the way in which information can reach many people in a very short time over long distances - allegedly Twitter travels faster than an earthquake <https://xkcd.com/723/>

Another is the way in which it reduces the barriers to publishing information - and people have been extremely quick to take advantage of this. In 1991 there was one web site. 14 years later there are around one billion (<http://www.internetlivestats.com/total-number-of-websites/>). That's a huge number and even with an estimate of 75% not currently in use, that's still 250 million websites. That one billion is made up of a huge range of information of all kinds being published by all kinds of people: businesses, charities, museums, libraries and individuals.

This is possible because we have a common system of addresses (URLs) and common standards on how to transfer information across the web (HTTP) and how to format the documents we publish (HTML)

“there may always exist additional sources of data, somewhere in the world, to complement the data one has at hand”

Learning Linked Data Project
<http://lld.ischool.uw.edu/wp/glossary/>

What is the *Open World Assumption*? It basically says - others may know things that we don't. That “there may always exist additional sources of data, somewhere in the world, to complement the data one has at hand”.

This is as opposed to the *closed world assumption*, which assumes that you know and control all the relevant data. An example might be a seat booking system in a theatre or cinema - in such a system it is reasonable to assume that every seat on the system that doesn't have a booking is currently free.

One of the things that typifies *open world* systems is support for making statements which negate information about things. If you operated a seat booking system with an open world assumption, for every seat that wasn't booked you'd need to make an explicit statement to that effect.



[s.n.]

A library catalogue as a whole generally operates on a closed world assumption - if there isn't a record for a book in your catalogue, you don't have that book. If a book isn't out on loan, it is in the library (well, or lost or stolen!).

However individual records in a library catalogue - the bibliographic descriptions - tend to work on an *open world assumption*. For example, if the publisher of a book is not known, you can explicitly state that it is not known. This used to be done with [s.n.] = sine nomine, but now done with the slightly more prosaic “publisher not identified”. This allows us to differentiate between ‘not known’ and ‘not recorded in this catalogue record’.

What has this got to
do with linked data?

With Linked Data you can...

- Create and publish unique identifiers for things you know about
- Make statements about things you know about using identifiers (your own, or other peoples)

In 2006 Sir Tim Berners-Lee published a note on Linked Data (<http://www.w3.org/DesignIssues/LinkedData.html>), this lays out the four things you need to do to have Linked Data. The first two of these are: 1) Use URIs as names for things; 2) Use HTTP URIs so that people can look up those names.

That is essentially to say use web URLs as identifiers for things. This means that every one of those 1 billion websites I mentioned earlier can not only be a place for hosting web pages, but also a place where identifiers can be created. If you have a website, you can publish your own linked data identifiers - it's just more URLs

Linked Data

- Enables you to say things about your own resources
- Enables you to say things about other resources
- Enables other people to say things about your resources

The other two things in that note on Linked Data are:

- * When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
- * Include links to other URIs. so that they can discover more things

These two points address how you use Linked Data to say things about your stuff or about other people's stuff. As long as that stuff is identified using Web URLs you can make statements about it using linked data.

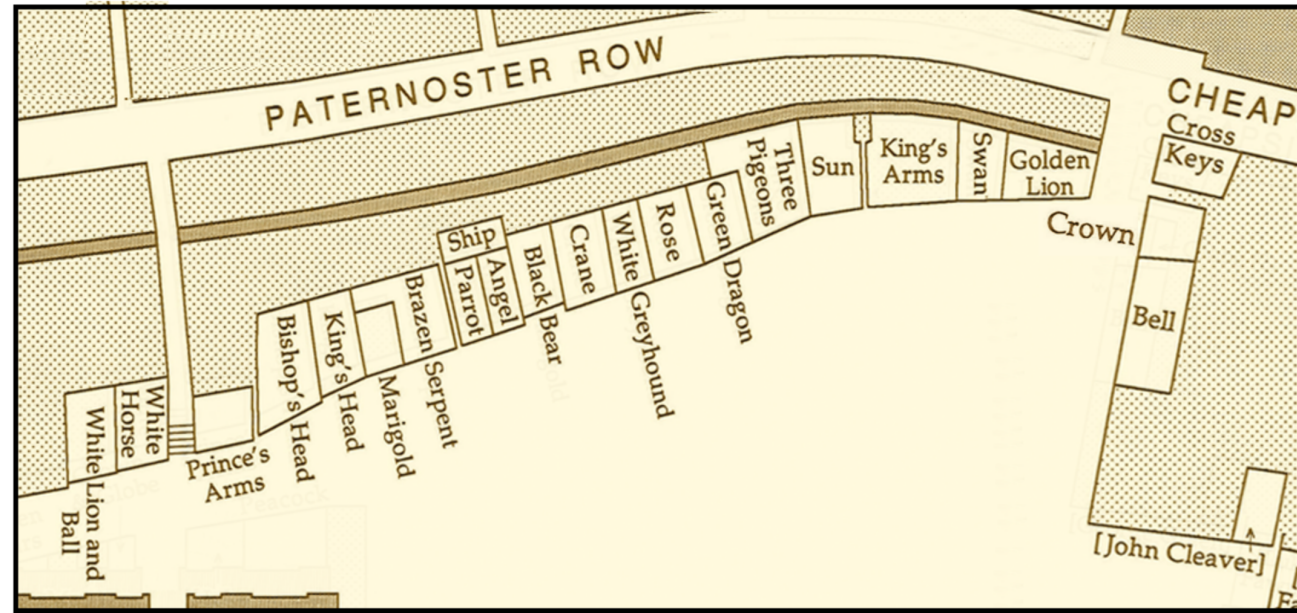
So in the same way that the web has made it possible to publish documents about things in a highly interoperable way (using HTML), Linked Data is a lingua franca for publishing data - it makes it possible for data from many sources, published by different people and organisations, to interoperate.

Mathieu has already talked about how the OU can use linked data to allow statements about resources across different systems in the OU. I want to talk about the implications of making it possible for other people to say things about your resources.

Printed by M. Flesher, for R. Dawlman
and I. Rothwell, and are to be sold at
the signe of the Brazen serpent, and
Sun in Pauls Churchyard

In March I attended the Early English Books Hackfest at the Bodleian Library in Oxford. The event was to celebrate the release of over 25,000 texts from the Early English Books Online (EEBO) project into the public domain. As is usual at such events, at the end of the day a number of people demonstrated or talked about the projects they had worked on during the day. There were a whole range of amazing projects, but one of the things that struck me was that several relied on the descriptive metadata, not the actual full text. Sadly, although the public domain texts do include some descriptive metadata, the underlying MARC records that describe the collection are not in the public domain.

One project I was particularly interested in was looking at colophon statements. Colophons are typically statements of who published a book and when and where it was published. However the colophons for material in EEBO often contain more information than this, extending to information about where the item was to be sold - such as the one here.



In the early 17th Century Paul's Cross churchyard (which is the site now occupied by St Paul's Cathedral) was full of bookshops. These have been documented by Dr Peter Blayney in "The Bookshops in Paul's Cross Churchyard. Occasional Papers of the Bibliographical Society, 5. London: The Bibliographical Society, 1990".

Just think what it would be like if Peter Blayney, or someone else, was able not only to do this research and record it, but publish it in a way that was intimately linked to the catalogue records. Linked Data makes this possible. If the English Short Title Catalogue (ESTC) which documents works published in the relevant period was published as linked data. Each work with it's own linked data identifier. Then anyone, including Dr Blayney could publish data which added to the knowledge that ESTC already represents, and link it to those records from the ESTC.

For the first time there is a way for experts anywhere in the world, no matter who they are to enhance library data in a way that can be captured and used - by them, by the libraries, by others.

You are not the (only) expert

Library data already operates on an open world assumption. It is time for us to embrace the full meaning of this - that we are not the experts on the resources we hold, and that potentially there always data in existence elsewhere that complements ours.

I see three key ways in which this could be exploited by libraries

Improved displays

[My Music example]

Copac Full Records

[Help](#) [Login](#)

[First](#) [Previous](#) [Next](#) [Last](#) [Brief Records](#)

Record 1 of 1.

Main author: [Handel, Georg Frideric](#), 1685-1759, [cmp.]

Title details: [I will magnify thee \[music\]](#) = : Ich will dich erhöhen : HWV 250b / [music by] Händel ; piano reduction based on the Urtext of the Halle Handel edition by Andreas Köhs.
[[I will magnify thee](#) HWV 250b. Vocal score. German & English]
[[Ich will dich erhöhen](#)]
[[Anthems für die Chapel Royal.](#)]

Series: [Handel, George Frideric](#), 1685-1759. [Works. 1955. Bd. 9](#) ; Ser. 3 ;

Published: Kassel ; New York : Bärenreiter, ©2005.

Physical desc.: 1 vocal score with piano reduction (v, 45 p.) ; 27 cm.

Identifier: EAN: 9790006527236

ISBN: 9790006527236 (pbk.)

Music plate: BA 4264a Bärenreiter

ISMN: M006527236

Notes: For solo voices (SATB), chorus (SATB), and orchestra; acc. arr. for piano.

"Supplementary edition to: Georg-Friedrich-Händel, Hallische Händel-Ausgabe ... Series III, volume 9: Anthems für die Chapel Royal" - T.p. verso.

"Bärenreiter Urtext" - Cover.

Pref. by Gerald Hendrie in English with German translation.

Words from various Psalms, in English with German translation.

Subject: [Chapel Royal \(St. James's Palace\)](#)

[Anthems.](#)

Added by the
'Composed'
Bookmarklet based
on information from
the MusicNet Codex
and the BBC

New Search

You can keep a record of your searches and save references by logging into Copac.

See information for "Haendel, Georg Friedrich" in other places



Image courtesy of the [BBC](#)

- [Records in the British Library Catalogue](#)
- [Records in the British Library Catalogue](#)
- [Records in COPAC](#)
- [Records in COPAC](#)
- [Records in COPAC](#)
- [RISM](#)
- [On the BBC](#)
- [Wikipedia](#)
- [IMDB](#)
- [The Grove Dictionary on Music](#)



http://www.meanboyfriend.com/overdue_ideas/2011/07/compose-yourself/

This was a demonstration application I built using linked data. It enhances the display of library catalogue records related to composers – in this case it adds a portrait of the composer from the BBC, and links out to more information elsewhere.

Data correction

[William Gibson example]

 100 1 _ [1a Gibson, William, 1d 1948-](#)
 100 1 _ [1a Gibson, William, 1d 1948-](#)
 100 1 _ [1a Gibson, William, 1d 1948-](#)
 100 1 _ [1a Gibson, William, 1d 1948-](#)
 100 1 _ [1a Gibson, William, 1d 1948-](#)
 100 1 _ [1a Gibson, William, 1d 1948-](#)
 100 1 _ [1a Gibson, William, 1d 1948-](#)
 100 1 _ [1a Gibson, William, 1d 1948-](#)
 100 1 _ [1a Gibson, William, 1d 1948-....](#)
 200 _ 1 [1a Gibson, 1b William, 1f 1948-2008](#)



National Library of Portugal
PTBNP-32722

This is data from VIAF - the Virtual International Authority File. VIAF has become a key Linked Data resource due to the number of people and organisations it identifies. However, in this example I want to highlight how that data can bring to light inconsistencies in our data, and it should trigger the question ‘have we got this right?’

How do people tell you you’ve got things wrong in your catalogue at the moment? What if they could publish corrections (or at least their view of things) in a way we could use gradually to improve our own data?

Here we can see that unlike all other libraries in VIAF, the National Library of Portugal thinks that William Gibson, the science fiction author, died in 2008. I suspect this would come as a shock to him!

If we published our authorities as Linked data it would open up the possibility of

Improved search and
discovery

Pr

Printed by M. Flesher for R. Dawlman and L. Faune at the Brazen Serpent in Pauls Church-yard (1)
Printed by M. Flesher for R. Dawlman at the Brazen Serpent in Pauls Church-yard (3)
Printed by M. Flesher for R. Dawlman at the Brazen Serpent in Pauls-Churchyard (1)
Printed by M. Flesher for R. Dawlman, and L. Fawne, at the signe of the Brazen Serpent in Pauls Church yard (1)
Printed by M. Flesher for R. Dawlman, and L. Fawne, at the signe of the Brazen Serpent in Pauls Church-yard (2)
Printed by M. Flesher for R. Dawlman, dwelling at the signe of the Brazen Serpent in Pauls Church-yard (2)
Printed by M. Flesher for Robert Davlman at the signe of the Brazen-Serpent in Pauls Churchyard (1)
Printed by M. Flesher for Robert Davlman, at the signe of the Brazen-serpent in Pauls Churchyard (1)
Printed by M. Flesher for Robert Dawlman, at the signe of the Brazen Serpent in Pauls Church-yard (1)
Printed by M. Flesher, for Luke Fawne, and S. Gellibrand, at the Brazen Serpent in Pauls Churchyard (1)
Printed by M. Flesher, for Luke Fawne, at the Brazen Serpent, in Pauls Church-yard (1)

<https://historicaltexts.jisc.ac.uk/>

I want to go back to the example of the bookshops in Pauls Cross Yard. If Peter Blayney had been able to link his knowledge about the booksellers and their shops to works - it would then be possible to bring this into library resources. Imagine how this 'publisher browse' could be improved by using the information from Peter Blayney's research

"The catalogue could be an
information source, rather
than just an inventory"

Karen Coyle in "Catalogers + Formats,
the Wider Web - Open Discussion"
<https://youtu.be/OtY3bWhUT9M?t=2371>
(39mins 30 sec)

Anyone can say Anything

The opportunities offered here are large, but there are problems as well. Anyone can say Anything. We know the web is full of mis-information (deliberate or otherwise) and not only that, there can be legitimate disagreement on a topic.

So is this just a free for all? Yes and no.

In general terms it is a free for all - Anyone can say Anything. But in terms of the data we choose to use - that's up to us. As Neil said, there is no 'truth', and we need to care about the provenance of data. We can select those sources we trust, and ignore those that we don't. We can look for contradictions in data we can see and use that to flag issues.

For example - maybe we would trust information about musicians from the BBC

Results - [Filter by use case](#) ▼

▶ MusicRecording (37) All good ✓

▼ MusicGroup (1) All good ✓

MusicGroup

name:	George Frideric Handel
image:	http://ichef.bbci.co.uk/images/ic/976x549/p01bqfdl.jpg
foundingDate:	1685-02-23
sameAs:	https://musicbrainz.org/artist/27870d47-bb98-42d1-bf2b-c7e972e6befc

▶ VideoObject (19) All good ✓

▶ Customised Search Result Filters

Here is linked data (schema.org) extracted from the BBC page about George Frederick Handel. It includes an image of Handel and his exact date of birth, as well as a link out to more information in MusicBrainz. It also contains a list of recordings and video material relating to Handel.

It's a rich resource, and just one of many we could make use of in libraries.

Trust him, he's a Doctor

But as well as looking to these trusted organisations, we should look to the expert individuals. These are people in our institutions and communities, and who have spent years developing expertise in their chosen fields - Linked Data gives us a chance to work with experts across the globe, and harness that expertise to improve what we do

We live in an open world - we can't assume that what libraries, or indeed other cultural heritage institutions have chosen to record, is the end of the story. Peter Blayney has been researching and documenting information on the book trade in 16th and 17th Century London for 20 years - I think we can trust the information he publishes as much as any we create.

Using these slides

These slides were developed by Owen Stephens
(owen@ostephens.com).

Unless otherwise stated, all images, audio or video content are
separate works with their own licence, and should not be
assumed to be CC-BY in their own right

This work is licensed under a Creative Commons Attribution 4.0
International License <http://creativecommons.org/licenses/by/4.0/>.

It is suggested when crediting this work, you include the phrase
“by Owen Stephens”

