Stories for machines, data for humans

Author : Tim Sherratt

Tagged as : Linked Open Data, narratives

Date : April 10, 2015

Presented at the New Factual Storytelling symposium, 10 April 2015, University of Canberra

I feel like the nerdy kid at the cool kids' party.

There are lots of interesting and creative projects on show today and I… well I want to talk to you about metadata.

Data. According to some pundits it's the new oil, or the new electricity. Fuel for economic development — a raw material ready to be 'mined' for insights, innovation and our purchasing preferences.

In the cultural heritage sector the data metaphors are more likely to be framed around liberation than exploitation. Our data wants to be 'open'. But there's still a tendency to think on an industrial scale — it's about pumping out large datasets for potential re-use.

What can be lost in metaphors of extraction and scale is an appreciation of the human origins of data. We are not buoys bobbing in the ocean reporting on the heights of passing waves. Big data is made up of many small acts of living.

So today I want to talk about small-scale, free-range, artisanal data. I want to talk about data, alongside storytelling, as the product of creativity, imagination, frustration and fury.

Let's think for a moment about the work of a historian — identifying actors, defining relationships, documenting the complex networks that bring together people, places and events over time. It's painstaking, exhilirating and potentially soul-destroying work. It's also an exercise in data modelling. Whether the results are preserved in a triplestore, a spreadsheet, or on a drawer full of index cards — it's nodes and edges, it's entities and relationships, it's data.

And that's ok. Making data doesn't condemn you to a rigidly empirical, deterministic framework. There's always room for nuance, interpretation and doubt. there's always room for stories.

But what happens when historians undertake the oddly-named process of 'writing up'. The complex data models are flattened down to a series of sentences neatly arranged in linear sequence — our things become strings. The data is squeezed out and discarded, glimpsed only as fragile echoes hiding in footnotes.

This is of course part of the skill of historical writing — the ability to represent complex relationships through narrative. But why can't we have our stories and data too?

This is a question I've returned to a number of times over the last few years.

It's come up because I get excited about Linked Open Data's potential to deliver structured, machine-readable information via the web. But then I wonder, whose stories will we be telling to the machines. How can we explore the expressive possibilities of Linked Open Data and not be constrained by instrumentalist assumptions about the models we make.

It's come up because I get excited about embedding cultural heritage collections within the passion and practice of everyday life. Why squeeze out the data from historical publications when every article could be an online exhibition, every book could be a digital portal, every footnote could be a link for exploration and aggregation?

I don't really have answers, but I do make stuff. I've a <u>few goes</u> at trying to <u>create narratives</u> that <u>embed</u> Linked Open Data.

They're not very exciting from a design point of view, but I keep coming back to them because there still seems to be a lack of alternatives. There's lots of talk about publishing Linked Open Data, but much less about how the use and consumption of Linked Open Data can be built into creative practice.

So here I am again.

This exercise has a number of constraints built in. The main one is NO PLATFORMS — a historian using a series of simple tools should be able to create and publish a data-driven web page without any dependencies. It should be as simple as uploading an html page to a server.

In my idealised workflow, the historian would manage their data about people, places, events and resources in a simple database capable of exporting a flavour of Linked Open Data known as JSON-LD.
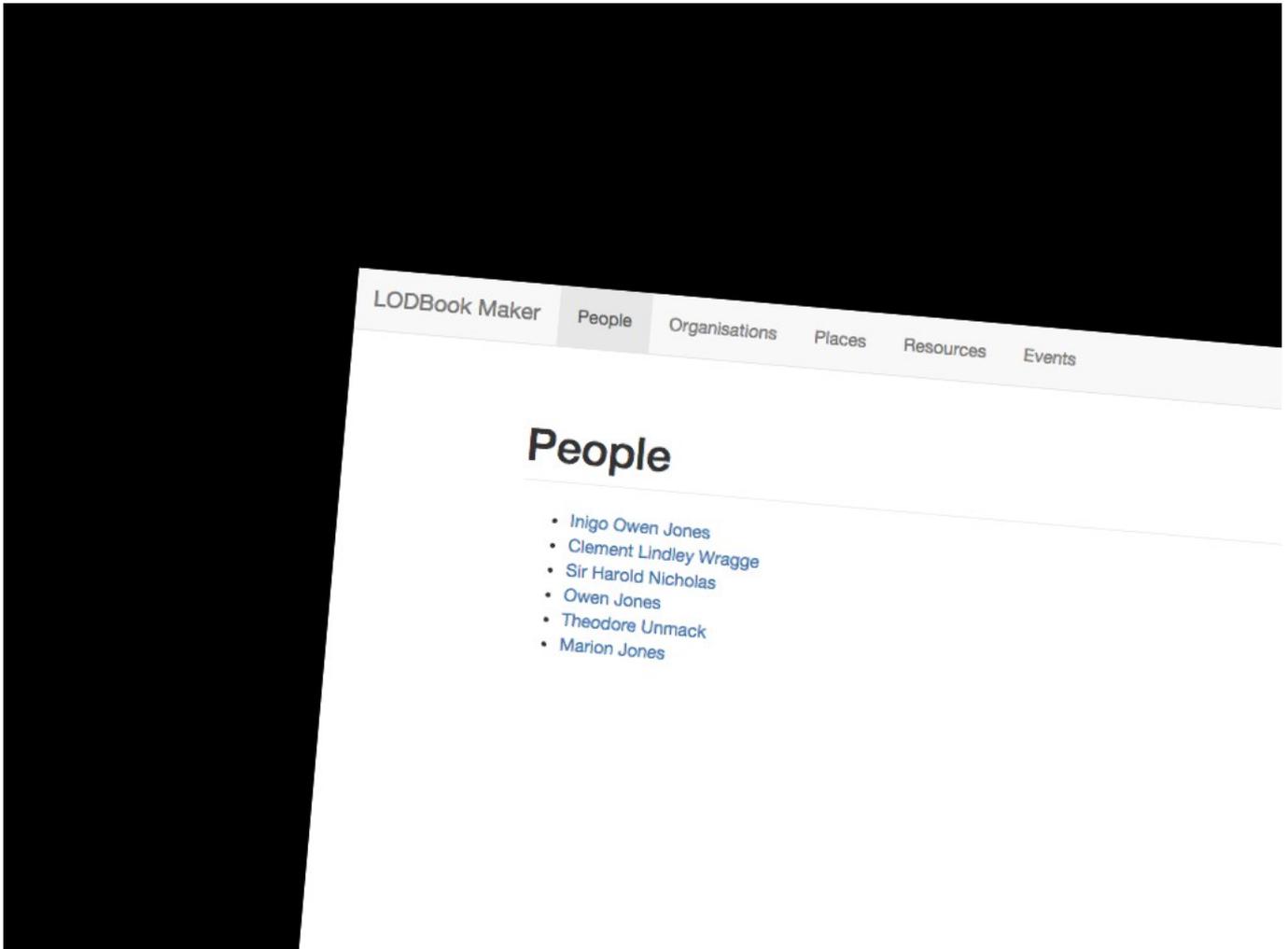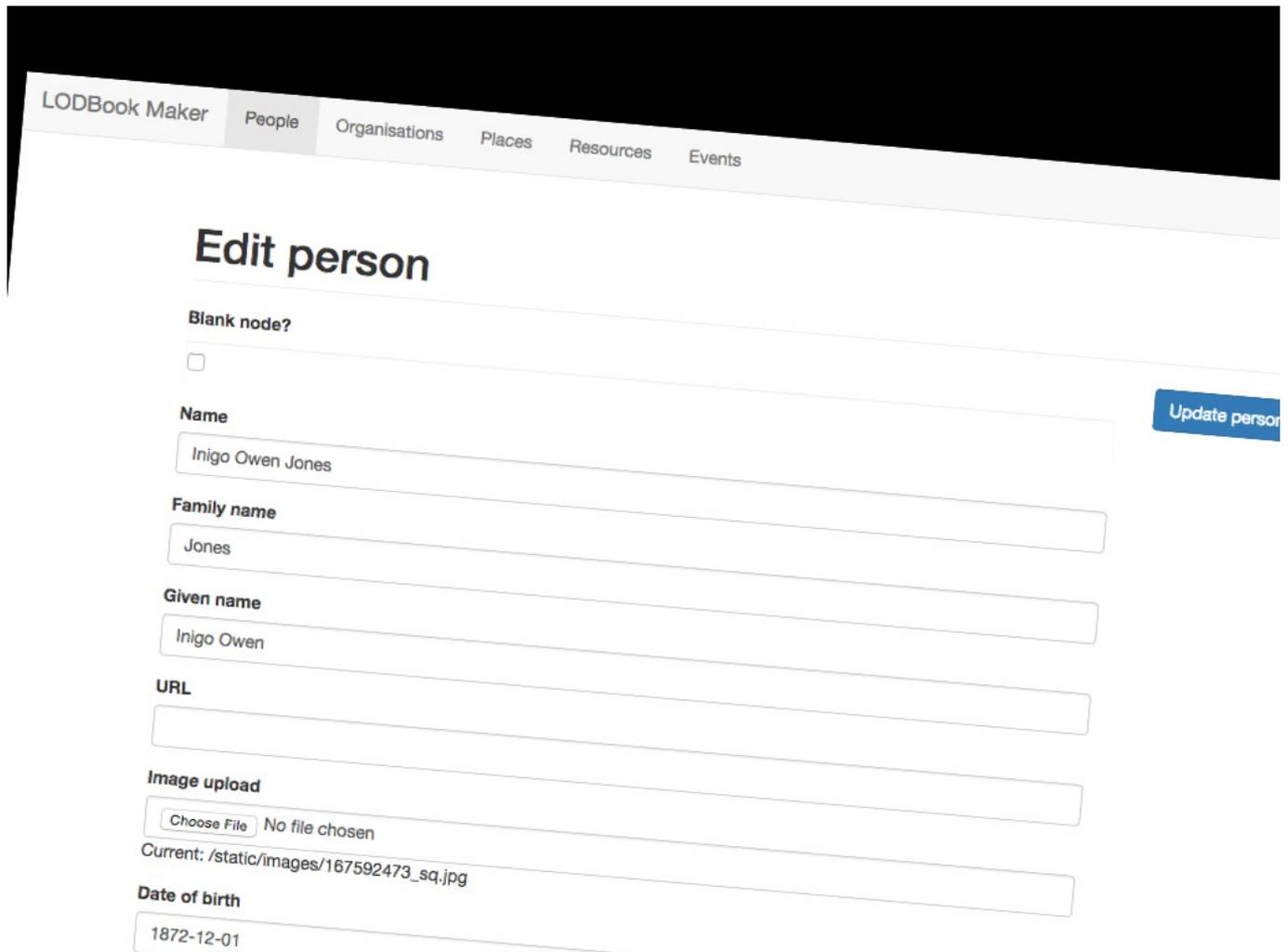
Then, having created their narrative, they'd mark it up in the tool of their choice to relate specific names or phrases in the text to the entities in their database.
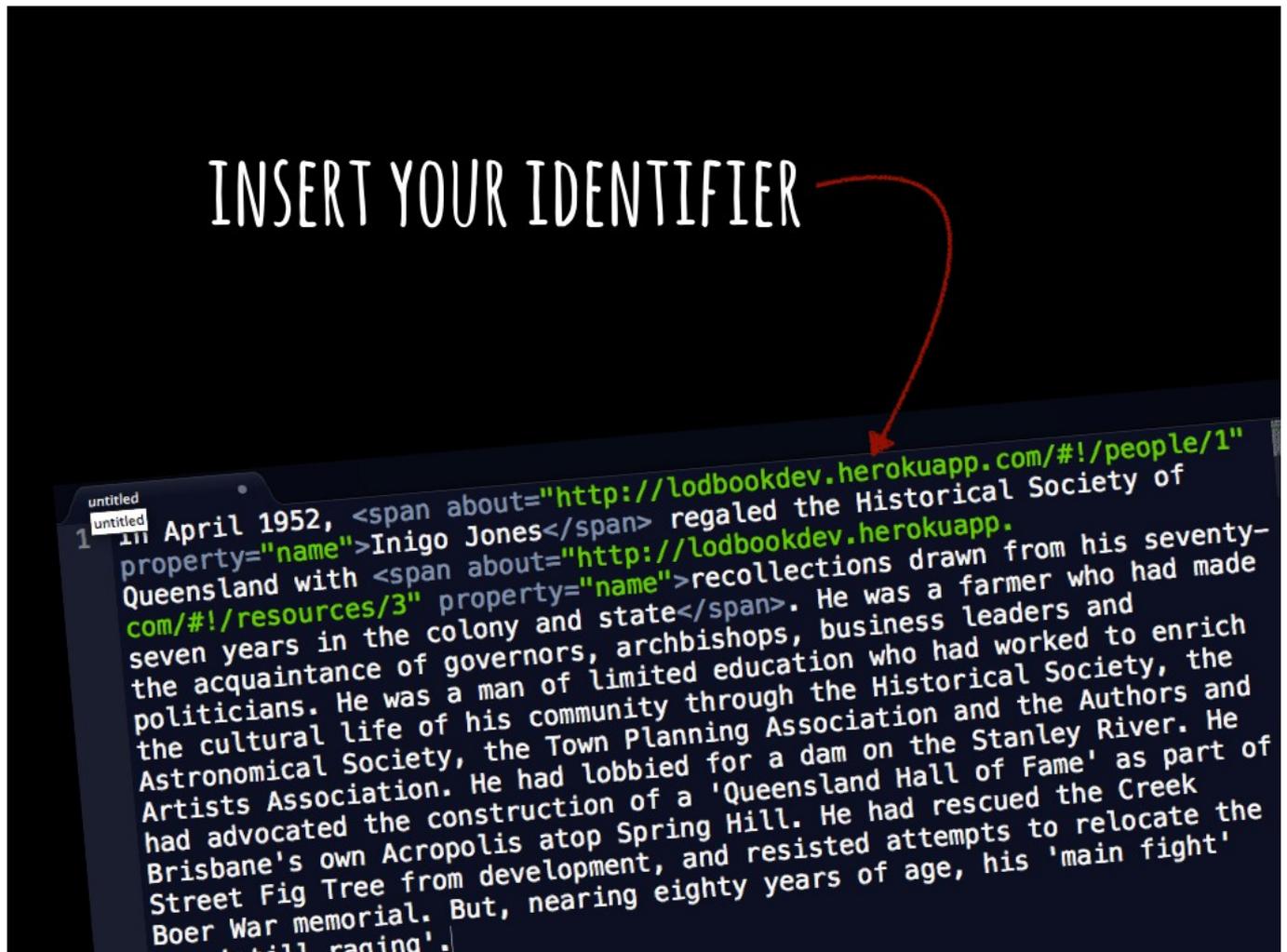
Then they'd just drop the text and the data into a html page and whack it on the web. With a bit of javascript magic to activate the data, you'd have [something like this](#)...

The demo is live (though still under construction), so have a play.

- Scroll the text to see those carefully inserted identifiers create pop ups in the sidebar.
- The text has itself become data, each paragraph is an object — try filtering the text or linking to an individual paragraph.
- Browse all the people, or resources. Explore all the relationships for Inigo Jones.
- Mapping to existing identifiers from sources like Trove and Wikipedia help put the 'linked' into Linked Open Data.
- There's a rather boring map, a timeline and a wall view. New views could be easily added by dropping in some extra javascript.
- It's all data, so other visualisations and analyses might be created on the fly.

That's what humans see, but what about machines? All the carefully curated data is exposed in a machine-readable form. Lots of triples…
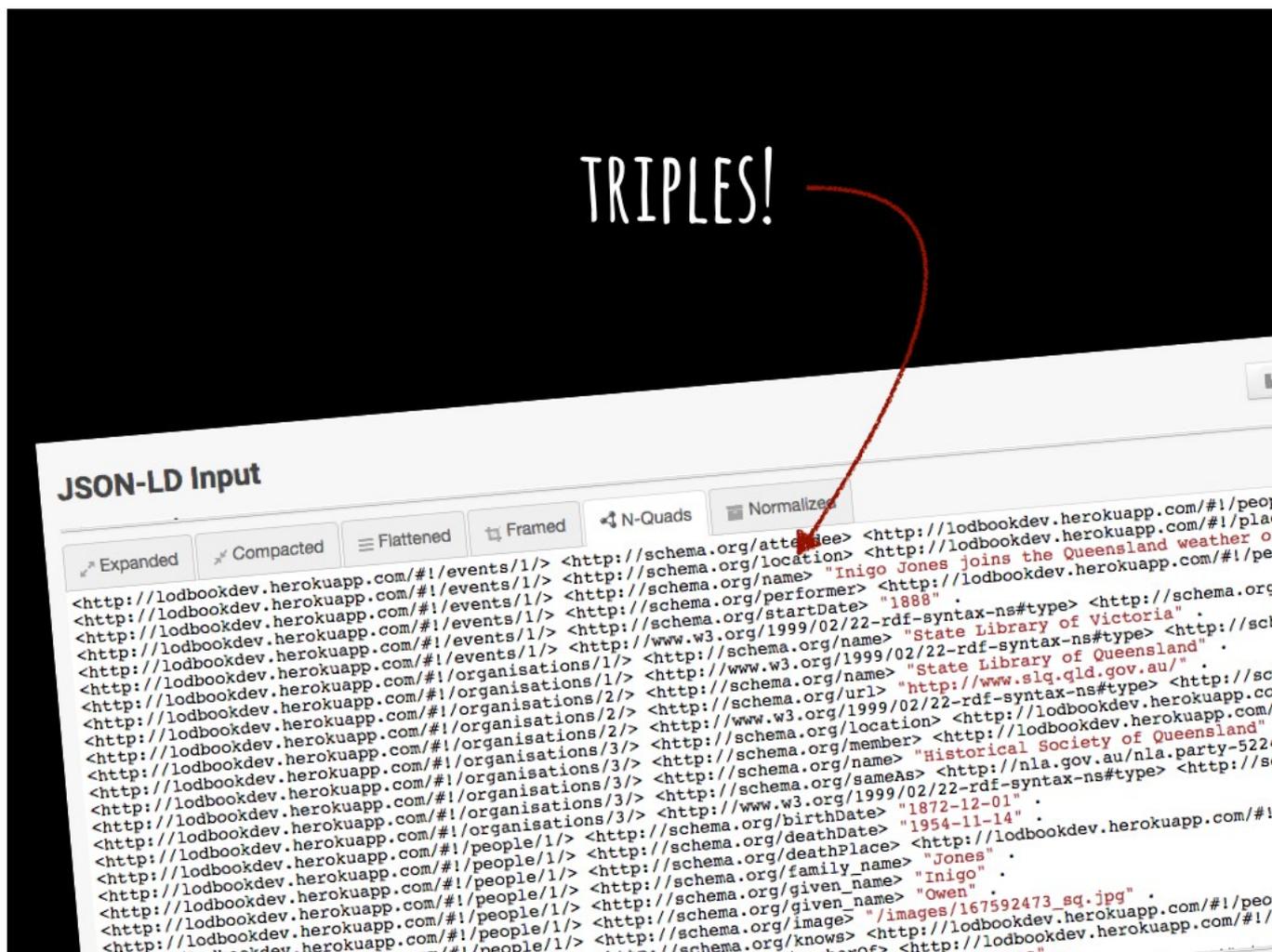
The code for the viewer and maker is all available if anyone wants to play with it, and I'm intending this year to develop two substantial monographs using these tools. Both have many links into cultural collections.

My aim here is not to develop a fully-operational publishing system. I just want to get a better idea of what's useful, what's interesting, what's possible. To think beyond the current limits of scholarly publishing into a world where data and narrative can live together, where interpretative work is represented in all data-inflected glory.

Share this:

- Click to email this to a friend (Opens in new window)
- Click to print (Opens in new window)
- Click to share on Twitter (Opens in new window)
- Share on Facebook (Opens in new window)
- Click to share on Google+ (Opens in new window)
-

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).