

Building a legislation tracker for the UK Parliament: mapping Linked Data ontologies and taxonomies into a parliamentary knowledge graph

Anya Somerville¹ and Michael Smethurst² and Silver Oliver³

¹ House of Commons Library, UK Parliament, UK

²Parliamentary Computational Section, UK Parliament, UK

³Data Language, UK

Abstract. We present a case study of building a site to track secondary legislation through Parliament. Describing how librarians and technologists in the UK Parliament are moving away from an approach centred around annotating documents to a taxonomy and instead building upon a knowledge graph for parliamentary business. This drives the need to look to new techniques and technologies including domain-driven design and Linked Data to build a semantic model for parliamentary procedure.

Keywords: Legislation, Linked Data, Domain-driven design, Knowledge Graph.

1. Problem statement

The UK Parliament is comprised of the House of Commons and House of Lords which are generally autonomous, as are the assorted offices within them. Many of our software systems have evolved separately and in parallel: digitising analogue processes in isolation. Because the systems are disjointed, the content and data produced is disjointed, which leads to the website being disjointed - and to both internal and external understanding of Parliament being disjointed.

An impact of this disjointedness is the difficulty in following the passage of secondary legislation through the two Houses. Users require knowledge of parliamentary procedure and understanding of where various offices publish relevant information. This has made it difficult to follow a piece of legislation through Parliament or to find items of legislation at the same stage within procedures.

The volume of secondary legislation expected to pass through Parliament as a consequence of Brexit brought this problem into focus and led to the development of an integrated secondary legislation tracker. Prior to the launch of this service, a Word document published online was the primary centralised resource describing what stage of procedure a particular item of secondary legislation had reached.

2. Our Approach

The secondary legislation tracker [1] was built to align with the principles adopted by Parliament to build a new web presence on a strong KOS foundation [2]. This takes the form of a knowledge graph built using Linked Data principles[3].

The capability of Linked Data allows us to be more expressive of semantics than the broader or narrower relationships of a typical taxonomy. For example: capturing the stages of a procedure, how they interrelate and the various possible paths a piece of legislation might take. To provide a working model we adopted novel approaches to knowledge representation.

2.1 Applying Domain-Driven Design to parliamentary business

To get to a model of legislative procedure we took a step back from the internal processes of individual offices and software systems. We used domain-driven design: a methodology for exploring, unpacking and disambiguating the language and processes of a community of practice. This was of benefit because an overarching model for the process of legislation was not available and we wished to ensure the complexities of Parliament were captured.

Parliamentary procedure is documented in multiple places, including Standing Orders and Erskine May. Individual offices might capture the elements of procedure pertinent

to their remit as a hand-drawn flowchart, but these fragments had not been assembled into a depiction of procedure relating to secondary legislation.

As a tool, domain-driven design provided a way to work with parliamentary clerks, subject matter experts in the Library and external experts to explore their understanding of parliamentary processes. This allowed for collaboration between technical and domain experts to inform design decisions and allows design and user needs to be emergent over time.

The output of this analysis was a model to support the secondary legislation tracker (see Fig. 1). This is a generic process model which we overlay with data, reflecting individual parliamentary procedures.

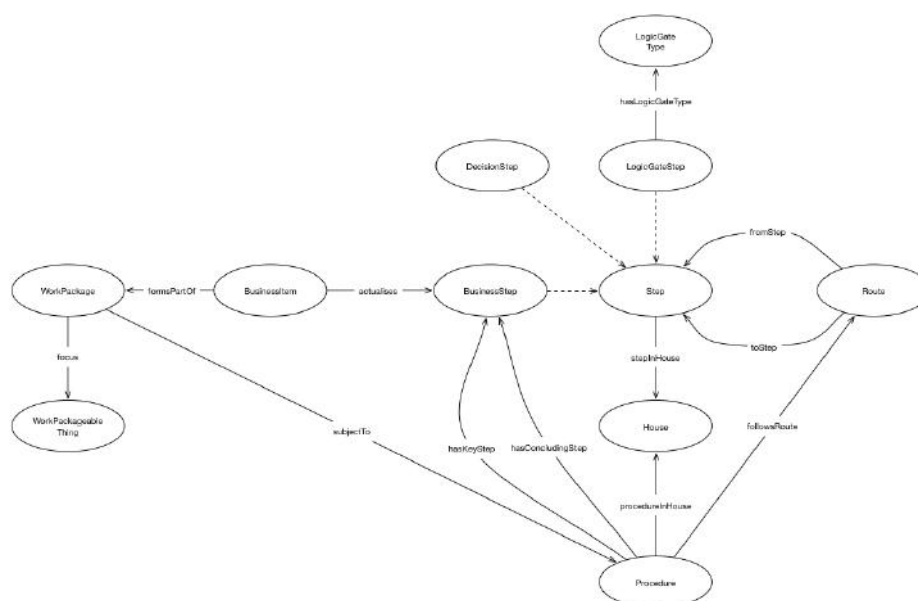


Fig. 1. The procedure conceptual model.

2.2 Ontologies, taxonomies and mappings

The next step in building a secondary legislation tracker is to take the procedure model and document it, with definitions for use by machines and people [4].

For core parliamentary procedure we document and publish ontologies. A key part of this process is the recording of definitions for both the classes and relationships of the model. These capture the thinking that has gone into the modelling process and the decisions made. For those concepts outside our core domain we maintain a SKOS based taxonomy of concepts [5] for the purpose of subject indexing.

Once the models have been defined they need to be populated with data. Some of the data used by the legislation tracker has not previously been modelled and does not exist in any line of business system. For example: procedures and possible routes through business steps.

New tools have been developed and the management of this data has been taken on by the Indexing and Data Management team in the House of Commons Library. This is more an evolution than a step change to the work this team already undertakes. In addition to taxonomy management, subject indexing and linking together parliamentary business outputs, the team also now manage maps of parliamentary procedure. It points to a future where librarians manage data that is explicitly more knowledge graph like in its representation.

From the outset the model and identifiers in the tracker have been designed with interoperability in mind. SKOS concepts are mapped to identifiers in the knowledge graph where required. We also map our entities to descriptions elsewhere on the web - often through Wikidata[6].

We work with colleagues in the UK Government Digital Service to model their processes and where they touch on those of Parliament. Together we intend to use Wikidata identifiers as a triangulation point between Parliament, government and beyond.

3. Outcome

The main output of the work to date has been the development of a tracker website [1] for the parliamentary scrutiny of secondary legislation and treaties. By understanding and mapping the underlying procedure we have provided a system allowing internal and external users to track business across both Houses and assorted offices, where no such system was previously available.

3.1 Unanswered questions

We're making progress improving the experience of consuming parliamentary information from our own systems, we don't yet have an answer as to how this impacts across the web.

We make our information available to be consumed and processed by news organisations, but how attribution flows across the web and how processing by third parties happens is outside our control. In an era of partisan journalism and fake news this is a problem we need to solve as a wider community.

References

1. Statutory instruments tracker, <https://beta.parliament.uk/find-a-statutory-instrument>.
2. Tetlow, J.: UK Parliament website public beta, <https://pds.blog.parliament.uk/2017/03/27/uk-parliament-website-public-beta/>.
3. Lang, S.: api.parliament.uk, <https://medium.com/@langsamu/api-parliament-uk-7b87597019a4>.
4. UK Parliament Ontologies, <https://ukparliament.github.io/ontologies/>.
5. UK Parliament controlled vocabulary, <https://api.parliament.uk/vocabulary/browser>.
6. Wikidata: WikiProject British Politicians, https://www.wikidata.org/wiki/Wikidata:WikiProject_British_Politicians.