The Metadata Model for Directgov

Abstract: Peter Jordan and Helen Lippell describe the role of metadata, including taxonomies, in re-designing the flagship government website, Directgov.

Keywords: websites; government; taxonomies; metadata

The landscape

Directgov is the UK government's vehicle for online public service delivery to individuals, consisting of the website www.direct.gov.uk, plus interactive TV and mobile services. It aims to offer users an easy-to-use and coherent experience, enabling people to find information and transact with government by bringing together content and services from across the whole delivery spectrum. The direct.gov.uk site currently receives around 20 million visits a month and is regularly the most visited government website. Traffic is expected to increase as the range of available services grows and public awareness of the brand rises.

Directgov also has a valuable role to play within government. It helps enable departments to reduce the amount of money they spend creating and maintaining websites and it helps promote a shift away from more expensive telephone and face-to-face contacts. Because of its significant reach to the public, it is a powerful digital channel for departments to promote key messages.

Directgov, along with the other two key government sites, http://www.businesslink.gov.uk and NHS Choices, is at the heart of a convergence programme which will see hundreds of public sector websites closed down. Their content will be re-purposed and moved onto these three main sites. Directgov will increase its scope to include many new pages of content, and new tools, across a wide range of public service areas.

This rapid ingestion of new content presents an interesting challenge in terms of findability and site structure. There is a small risk that important content may not be as easy to find as at present and that key user journeys will be lengthened or made more complex. Therefore, work has been underway for some time to design the next iteration of Directgov which will offer a flexible, scaleable infrastructure that protects and enhances the user experience and continues to provide departments with an effective medium for communication. The use of metadata, in conjunction with business rules, underpins this work.

This article will cover the user experience and product developments that metadata will play a role in delivering, and detail the work that is being done to make this happen. An expanded metadata model and new taxonomies will provide the infrastructure to tag content and, more importantly, to use that tagging to meet various business needs.

How metadata helps

Flexible site structure

Currently Directgov is organised hierarchically with 16 top-level sections, with a mix of topics and audiences. This navigation "tree" is hard-wired into the Content Management Scheme (CMS), meaning that moving content or implementing new nodes requires technical intervention. In future, the site hierarchy should be stored and managed outside the CMS, and simply uploaded when changed.

The site structure is, somewhat understandably for a website that aggregates content from across government, subject to internal debate and can be contentious. Managing it as metadata and business rules offers an opportunity to assemble the site more flexibly, so that changes can easily be made to move nodes around, or to add and delete whole new sections when content changes and user needs demand it. It also provides greater flexibility for creating content aggregations that are only temporary, such as budget day or severe weather events. Directgov has faced criticism from various sources about the findability of its content through browsing, although its findability ratings do compare favourably with other websites in the financial sector, which often have complex and legal content. The metadata-based approach, subject to the right business processes and technical implementation, means that content can easily be found in different places.

Allied to the more flexible site structure will be the ability to cross-sell related content more intuitively. For example, related links can be generated automatically based on relationships in the taxonomies, or related tags can be used as a means of exposing contextually-relevant content which users might not otherwise have seen by searching or browsing through the navigation.

7

Peter Jordan and Helen Lippell

Richer search dialogue

With increasing volumes of content coming onto Directgov, there is a need for site search to provide more tools to help users refine queries. Many users search with only one or two keywords, for example "tax", which can mean pages of results that are not relevant enough. Metadata-based facets and filters will help users funnel that initial query to get to the information they need quickly. Faceted search interfaces are common in e-commerce and major web search engines such as Google and Bing are making more use of facets to improve the search experience. Directgov Search will make use of metadata tags for facets such as topic, audience and life event.

Synonyms in the new taxonomies will complement Directgov's search engine optimisation (SEO) and search technology by ensuring that commonly-used user language maps to the "official" terminology and thus delivers the right search results. For example, there are still a sizeable number of searches for "family allowance" despite the fact this has been known as "child benefit" for some years. Using the taxonomy to tag means that the commonest synonyms are automatically mapped to the content without editors having to think of them every time they write about the subject. Acronyms and variants are other valuable synonyms that the taxonomy includes – National Insurance number has a number of variants including "NI number", "N.I. number", "NINO", "N.I.N.O.". The taxonomy will not capture every possible variant and synonym, of course, but rather, will map the commonest ones as identified by search log and customer insight analysis.

Increasing trust in content

A lot of useful information can be locked away in metadata. New designs for articles will include space to show extra information about the content, in order to increase quality and provide better signposting to users. For example, information about devolved administration (DA) applicability, which is currently tucked away in the HTML header would now be clearly presented on the page, along with links to content on respective DA sites if the Directgov information does not apply to the whole of the UK.

Another example is the idea of publishing to the page the name(s) of the departments or agencies who own the policy the content relates to. This provides reassurance to users who are used to interacting with certain brands. A behind-the-scenes benefit of tagging with policy holder(s) is that it will be possible to report on content usage by department. As content organisation moves away from a rigid model, departments will still be able to see how much their content is being used regardless of where and how often it is actually surfaced.

Foundations for semantic web

With the use of a comprehensive topic taxonomy to tag content, it will be possible to tap into the growing ecosystem of linked data and its potential to join and expose data from across the wider web. This could be particularly beneficial as mechanisms for inbound and outbound syndication are built on, helping Directgov content to move beyond its own domain. The taxonomies that link together topics, audiences and life events through relationships may in future need to evolve into more structured ontologies.

Recognising the government-wide effort to open up data, Directgov will be reviewing how content is structured and marked up for re-use, e.g. as being for particular channels, of relevance to certain audiences, or containing structured information such as contact details or tax rates.

Improving editorial productivity

An important principle for the project is that it delivers efficiency and productivity savings. This means that even though editors will be asked to apply or review more metadata for each piece of content, there are changes that can be implemented to speed up the process of content creation. One of these will be the implementation of an automatic classification solution to suggest tags when editors are creating content. This will avoid the need for them to pick terms from the taxonomy itself (unless they really want to) and ensure a level of consistency that manual tagging just couldn't provide.

Metadata will also support the automation of tasks that are currently manual, such as adding related links or aggregating content relating to a single theme. The level of actual automation and editorial oversight will be designed through a wider business change programme, but metadata is very much the enabler for the functionality. It supplies the tags and the relationships between tags and the basic framework for business rules and content to work off, to create an optimal user experience.

Metadata model

There was already a mature metadata model in place, developed over the lifetime of Directgov and its predecessor, UK Online. It supports a wide range of standard content management needs, including enabling editors to find content in the CMS (almost every metadata field can be searched on). It ensures Directgov's compliance with the eGovernment Metadata Standard (eGMS), including using the Integrated Public Sector Vocabulary (IPSV) for subject tagging.

An important piece of work was to review what kind of metadata would be needed in the next version of Directgov. This was distilled into four main "purposes" that helped to make metadata a bit less abstract and

better understood by the wider business. The four purposes were:

- Describe content – the means to describe the "aboutness" of content and to describe using different facets of the information such as audience or topic

- Support Search – the key fields such as Title and Description that are used by Directgov Search and external engines, to ensure good relevancy and provide accurate information in long lists of results

- Manage content – e.g. policy approval date, editor comments

- Describe ownership – this means being able to define content owners at various levels, from an individual author, the business unit and the government department that owns the policy. These fields help in reporting and management information.

The taxonomies

To support the population of the new descriptive metadata fields, new controlled vocabularies are needed. The largest of these will include topics from across the full public services spectrum that Directgov covers. Its development will need to be cognisant of both existing and future content coming from convergence.

The taxonomy will include terms for the work of government such as benefits, initiatives and campaigns and consumer-focused topics of relevance to public services such as mobile phones. The taxonomy will map together official terminology and user language primarily to support search, as described in the "Richer search dialogue" section.

Taxonomies for audiences and life events are being developed alongside the topics and mapped to them where appropriate. Audiences are groups of users with similar information needs or circumstances, such as lone parents, part-time workers or mature students. They might even be a group with immediate but finite needs, such as Cumbrian flooding victims. Similarly, life events are situations when people might need to interact with different parts of government and there is value in aggregating content to provide the full picture. The terms in the audience and life event taxonomies will be flexible and logical, and a useful means of bringing together disparate content.

Metadata governance

The changes being developed for the website have implications far beyond what users see. To implement the new model will require a large amount of business and cultural change. Critical to this will be the management of metadata and associated business rules. Work is being done to define the skills, roles and processes needed to ensure high quality metadata supports the aims of the project.

Taxonomy maintenance

Any controlled vocabulary runs the risk of ossifying and becoming less relevant over time. Terms will change, especially in the government arena as priorities and politics change. New initiatives and technology will emerge. Key services like benefits will change their name or be superseded altogether e.g. Incapacity Benefit has been replaced for new claimants by Employment and Support Allowance. The relationships between concepts could change too – for example, in IPSV, "cycling" is a child of "outdoor pursuits," which is itself a child of "sports and recreation," whereas since IPSV was last updated in 2006, the government has been promoting and investing in cycling as transport for commuting, school etc. A robust but flexible governance regime will be needed to ensure that the taxonomies continue to support business needs.

User language will change too – users are unpredictable and will not always refer to things as government expects them to. Ongoing effort to understand user needs and expectations is required to ensure that the metadata stays relevant.

Conclusion

Ultimately, the enhanced metadata model and associated taxonomies must work in harmony with the design, the content, the technology and business processes to create the flexible website that will support both the needs of departments and users. It is an exciting time for Directgov, and a vindication of the power of metadata that it is central to the future success of the nation's official website.

Biography

Helen Lippell

Helen Lippell has worked in online publishing since 1998, and after starting out classifying news articles for FT.com, worked as an information architect for the BBC on projects such as redesigning the homepage and rebuilding the A-Z index. She is now the Metadata and Taxonomy Analyst at Directgov. HelenLippell@directgov.gis.gov.uk.

Peter Jordan

After an early career in university administration and adult education, Peter Jordan has spent over 10 years working in digital media, increasingly focusing on addressing findability through search and metadata. He is Strategic Search Manager at Directgov. PeterJordan@directgov.gsi.gov.uk.