



INFORMATION MANAGEMENT: The Family Information Directory Case

Summary

Since August 2008, we have been responsible for the design, development, implementation and governance of the controlled vocabularies which underpin the Department for Education's Family Information Directory (FIDy). All Local Authorities in England upload records of information about services for children and families such as Childcare, child protection, leisure and recreation, education and so forth. This constitutes 10s of 1000s of individual records which are stored centrally in a single database known as the Aggregator. In turn, the Aggregator is accessed from multiple different portals including DirectGov and portals operated by third sector organisations. This means that people are able to easily search for information across Local Authority boundaries. This innovative system works because all Local Authorities use the same controlled subject vocabulary to classify ("tag") their information records. The vocabulary itself was developed using a multi-iteration approach, with parents and practitioners, with the aim of developing a thesaurus that would be more reflective of the language used by parents than the jargon used by the policy makers. And it works.

Introduction

The Family Information Directory ("FIDy") was originally called the Information System for Parents and Providers ("ISPP"), constituting a core part of UK's Department for Education's Parent Know How Programme ("PKHD"), now known as Family Information Direct ("FID").

The FIDy was established to assist Local Authorities throughout England in meeting their statutory obligations under Section 12 of the 2006 Childcare Act. This sets out the requirements for information that must be made available to parents and prospective parents by Local Authorities in England. More specifically, such information, advice and assistance must be made available through Local Authority provided services which are *maintained*. The aim is to make (online) directories of local and national information accessible to parents through a number of channels, including DirectGov, Third Sector Organisations and others.

Rationale

FID identified a need for a controlled subject vocabulary which could be used by all Local Authorities, thus representing a common language of data classification and management. Together with other ancillary vocabularies ("Complementary Controlled Lists"), this would deliver a consistent and meaningful "searchability" across multiple directory platforms. The objective was to improve the reach and quality of parenting information and support services, and to ensure that professionals are able to access the right information – locally and nationally – when they need it.

Controlled vocabularies of terms are used to describe and classify discrete units of digital information within the context of Metadata (“data about data”). Metadata is used to label information items – rather like barcodes - for storage within directories, repositories and so forth. The controlled vocabulary is the encoding scheme which gives a finite list of all terms that are available to describe information items within a given field of the Metadata. The introduction of a common controlled subject vocabulary means that the people who are doing the tagging are choosing terms from the same list. This becomes particularly important and indeed essential where, as in this case, we have more than 150 different organisations tagging information items which will be stored in a single national directory. For the information searcher, the use of a common vocabulary reduces learning time and leads to a closer connection between the language used by the information provider and that used by the information seeker.

The approach

The development of the FIDy controlled subject vocabulary was split into a number of project phases. The original aim was to create a vocabulary that would represent an extension to the Integrated Public Sector Vocabulary (“IPSV Version 2.0”, published in 2006). IPSV is a controlled vocabulary used with subject metadata. The e-Government Metadata Standard (e-GMS) version 3.1 mandates the use of at least one IPSV term for all UK public sector information resources that are to be shared with people, businesses or other public sector bodies. Consequently, it was assumed that a number of Local Authorities in England would already be using IPSV for classifying information items. A significant drawback of a close affiliation with IPSV is that the vocabulary remains, to this day, unmaintained. The meaning of this is that the vocabulary contains out of date terms. However, by establishing the FIDy vocabulary, gaps in the existing IPSV could be addressed.

Two principle methods were used to develop the initial FIDy vocabulary. First, a wide range of relevant existing vocabularies (e.g., IPSV), websites and documents were consulted. Terms which were seen to be prominent, used in the context of overall information architecture of a site, terms used in publication titles or terms used as categories, and which were relevant to the scope of FIDy, were harvested. Secondly, top level headings for the vocabulary and keywords were further developed in workshops and one-to-one sessions with groups of parents and practitioners from around the country. Local Authority Family Information Service Managers were also consulted. The primary aim was to develop a vocabulary that was reflective of parents’ natural language.

Simultaneously, current practice in vocabulary mapping, and vocabulary governance were investigated with the results used to inform future decisions and actions.

- A key decision taken was to map the FIDy vocabulary to IPSV version 2.0, using a one-to-one direct mapping method. Mapping between vocabularies is generally seen as aiding the persistence and good curation of information items. In this case, it would also reduce the “tagging” workload on those Local Authorities which had already tagged content using IPSV, as mapping would negate the need to re-tag.
- A further core decision was that the FIDy vocabulary should be proposed to the (then) DCSF Information Standards Board for adoption as a new national standard. This would have implications for the vocabulary’s governance.

- As a direct consequence of the standards-related decision, it was determined that the vocabulary should be developed in Zthes as a BS 8723 compliant polyhierarchical vocabulary.

Refining the product

The second phase of the vocabulary development saw the vocabulary undergo a number of iterations with Local Authorities, parents and practitioners to validate its content. It was subsequently developed in machine-readable format (Zthes) using the Lexaurus vocabulary development tool. The vocabulary was then mapped to IPSV using the following methodology:

- Mappings were based on term semantic equivalence;
- Mappings held internally to the FIDy vocabulary, with the original FIDy mapped term retained within the vocabulary but deprecated in favour of its IPSV equivalent;
- An unmapped version of the FIDy vocabulary was produced as a by-product of this process.

During this phase of the project, considerable effort was focused on determining the most efficient and effective method of rolling out the vocabularies (principally the subject vocabulary), requirements for training and support, and means of ensuring 100% take up of the vocabularies. To this end, a consultation was carried out with Local Authorities and vendors of their Family Information Systems (FIS). Guidelines were also prepared to assist and support Local Authority personnel in their approach to and implementation of the vocabulary in record classification.

Feedback and comments on Version 3 of the subject vocabulary suggested that the “Disability” category could be improved on. Based on consultations with various organisations with experience and expertise in this area, a new Disability category vocabulary was developed.

Version 4.0 of the vocabulary sees one other major change and improvement. In previous versions, the IPSV mapping had been held internally to the vocabulary itself. In version 4.0, the mapping is held externally as a separate vocabulary formulated as an A to Z list of FIDy terms, with the mapped IPSV term related to each as a Non-Preferred Term. This means that the FIDy vocabulary can be easily mapped to as many vocabularies as needed without compromising the core vocabulary itself.

What the vocabulary consists of

Currently the vocabulary comprises 19 Top Level Headings (Level 1), Level 2 and Level 3 terms and non-preferred terms, around 3,000 individual terms in all. All Level 1 and 2 terms have full Scope notes contained in the metadata. The subject vocabulary exists in the following formats:

- Zthes
- Excel spreadsheet
- A-Z word document (Alpha list)
- Esd-Toolkit version

Supporting documents include:

- Tagging guidelines

- Guidelines for Channel Partners
- Change notes

“TagFests” : the strategy for vocabulary adoption

One of the greatest challenges for this vocabulary project would be its adoption by all 150+ Local Authorities in England.

During the end of June and early July 2009, we staged a number of “TagFest” events – workshop/seminars – for Local Authorities (LAs) around England. These focused on helping LAs to prepare for and use the Subject Vocabulary within their FSDs. Events were held in Birmingham, Liverpool, Bradford, Sheffield, Carlisle, Norwich, Exeter, Portsmouth, and London. The events were extremely well attended, with some being over-subscribed. Most agreed that the TagFests were very useful and effective in addressing both generic and individual needs and issues. For us, these events were also useful in gathering opinion, feedback and commentary.

Outcome

All Local Authorities in England are using the FIDy Subject Vocabulary to tag their children’s and families’ services records, prior to uploading them to the Aggregator. It contains more than 100,000 individual records providing information on a wide range of services that are used and needed by families. Without this, people would need to trawl through countless different portals and websites, each using its own classification schema (or none at all). This is harnessing the power of words.