Enabling Interoperability of Government Data Catalogues

Fadi Maali, Richard Cyganiak, Vassilios Peristeras

Enabling Interoperability of Government Data Catalogues

Fadi Maali[1], Richard Cyganiak[1], and Vassilios Peristeras[1,2]

[1] DERI, National University of Ireland, Galway.
{fadi.maali,richard.cyganiak,vassilios.peristeras}@deri.org
[2] Greek National Center for Public Administration and Decentralization

Abstract. Opening public sector information has recently become a trend in many countries around the world. Online government data catalogues with national, regional or local scope act as one-stop data portals providing descriptions of available government datasets. These catalogues though remain isolated. Potential benefits from federating geographically overlapping or thematically complementary catalogues are not realized. We propose an RDF Schema vocabulary as an interchange format among data catalogues and as a way of bringing them into the Web of Linked Data, where they can enjoy interoperability among themselves and with other deployed datasets. The vocabulary's design was informed by a survey of seven data catalogues from five different countries, and has been verified by unifying four data catalogues to allow cross-catalogue queries and browsing.

Keywords: Government Catalog, RDF, Vocabulary, Interoperability, Linked Data

1 Motivation

"Open Data" and "Open Government"—these terms describe a recent trend towards more openness and transparency in government that is both demanded by advocates in the public and embraced by some administrations. As a part of this trend, information that was previously inaccessible to the public is increasingly opened up, often using the Web [4, 5, 1]. This development promises social benefits through increased transparency and openness; economic benefits and private sector cost savings through realising the full potential of data that has already been produced as part of the administration's day-to-day operations and paid for by the taxpayer; and to enable the provision of new innovative services that the government cannot or will not provide [10].

Data catalogues such as data.gov in the US[3], data.gov.uk in the UK[4], CA.gov Data in California[5], the New York City Data Mine[6], and the London Datastore[7]

[3] http://www.data.gov/
[4] http://data.gov.uk/
[5] http://www.ca.gov/data/
[6] http://www.nyc.gov/html/datamine/
[7] http://data.london.gov.uk/

have recently appeared as one-stop web portals that facilitate access and increase findability of such data by providing lists of government datasets along with metadata such as name of the publishing agency, file format, geographic coverage, and category of the dataset. Catalogues differ in scope (national, regional, local) and in operator (official or citizen initiatives). The phenomenon of the data catalogue, including its history and public policy environment, is studied in [13].

Public sector data ranges from census data to lists of locations of fire hydrants. On the technical side, it may take the form of office documents (PDF, Excel); geographical data (ESRI shape files, KML files); statistical data (SDMX, PC-Axis); developer-oriented XML files or web service APIs; and web sites with wizards for searching complex databases.

It is common for the data catalogues themselves to be available not just as a web site, but also in some format that is amenable to machine processing, such as CSV, RSS feed, or embedded RDFa markup. This enables bulk processing of datasets, automated checks for updated data in applications, and refined search over the often thousands of catalogue records.

In this paper, we propose a standardised interchange format for such machine-readable representations of government data catalogues. The adoption of such a format—either directly by the catalogue operators, or through wrappers that convert from the currently available machine-readable formats to the proposed standard—has several benefits:

1. Embedding machine-readable metadata in web pages increases findability by next-generation search engines.
2. Decentralised publishing: Individual agencies could publish separate catalogues, which could be aggregated into national or supra-national (e.g., EU-wide) catalogues.
3. It enables federated search over catalogues with overlapping scope, such as the catalogues for San Francisco, California, and the entire US.
4. Application developers can benefit from one-click download and installation of data packages into local databases.
5. Manifest files with accurate dataset metadata are crucial in efforts towards archiving and digital preservation of valuable government datasets.
6. Software tools and applications, such as improved search and data visualisation interfaces, can be built to work with multiple, or even across, catalogues.

An interoperability format for data catalogues becomes particularly interesting when the catalogued datasets are also amenable to machine processing. Our effort is therefore well aligned with the increased exploration and adoption of the Linked Data technology stack in government data publishing [7, 8, 11].

Defining an interoperability format for data catalogues is challenging. Catalogues differ widely in their scope, terminology, provided metadata fields, and the quality and amount of structure in the collected dataset descriptions. To clarify the requirements and guide the design of the interoperability format, we undertook a survey of seven existing data catalogues. We report on the results in Sect. 2. Section 3 presents our proposed interoperability format, the *dcat* RDF vocabulary. Section 4 reports on a feasibility study that unifies the contents of

four data catalogues. Section 5 discusses related work, and Sect. 6 concludes and reports on ongoing work within the W3C towards broader adoption and standardisation of *dcat*.

2 Survey of Data Catalogues

Most government data catalogues are in beta version and under constant development, so rather than conducting a comparative study, the goal of our analysis was to identify commonalities and overlap in the structure, and to document challenges and practices in this new and rapidly evolving area. We believe that this analysis is timely and that it can guide future initiatives towards setting up new data catalogues.

The goal of the survey was to ensure that our model reflects the reality of current data catalogues. The model must cover what's available in the catalogues, without requiring investments in new data acquisition or manual data cleanup. This encourages quick uptake and lowers the cost of adoption. After listing the surveyed catalogs (Sect. 2.1) and reporting on general characteristics of the catalogues (Sect. 2.2), we thus examine their structure to identify common metadata fields, to determine which ones should be treated as required, recommended and optional (Sect. 2.3). As several of the use cases listed in Sect. 1 require accessing and processing of the actual data files, we examine download links in Sect. 2.4.

The survey was done by first importing the studied catalogs into a relational database, which required development of a custom importer or screen-scraper for each catalog. SQL queries were executed against the database in order to study the completeness and consistency of values within each metadata field. This was combined with study of the catalogue websites and additional documentation on the catalogues' metadata schema where available. Manual inspection of all datasets was used to determine the availability of direct download links.

2.1 Catalogue Selection

For our analysis we select seven catalogues from five different countries:

1. *data.gov*: A catalogue of machine readable datasets generated by the Executive Branch of the US Federal Government.
2. *data.gov.uk*: A catalogue of UK governmental data.
3. *data.govt.nz*: A directory of publicly-available New Zealand government datasets.
4. *data.australia.gov.au*: The home of datasets created by different Australian government agencies.
5. *datasf.org*: A clearinghouse of datasets from the City of San Francisco.
6. *data.london.gov.uk*: An initiative by Greater London Authority (GLA) to release as much of the data that it holds as possible.
7. *statcentral.ie*: Provides information about official statistics produced by Ireland's government departments and state organisations.

This selection was chosen to include a range of different kinds of catalogues. All the major national catalogues available at the time of writing were included. Two local catalogues (SF, London) are included as representatives of smaller-scale catalogues. We made sure that some catalogues overlap in their geographical coverage (US and SF; UK and London). Finally, statcentral.ie was included as the "home catalogue" of the authors, and because of its focus on statistical data, a mature discipline with well-established metadata management practices.

2.2 General Characteristics

Size. The size of a catalogue here refers to the number of datasets it includes. This is an indicator of limited value which we present just to enable broad comparison, as there exists no consensus on the definition of dataset. For example, "2005 Toxics Release Inventory data for Texas" and "2006 Toxics Release Inventory data for Texas" may or may not be considered a single dataset. Most of the catalogues are constantly updated, so the numbers are only valid at the time the data was collected (January 2010).

Machine-readability. The data catalogue itself is considered "data" and should be published as structured data, so that third parties can extract information about the datasets [7]. This is achieved in one of the following ways:

- *RDFa*, a syntax for embedding structured RDF data in HTML pages.
- As described in [14], *feeds* can serve not just as a notification mechanism but also as persistent access points.
- A machine-readable version of the catalogue (usually CSV or XML) is listed as a dataset within the catalogue, e.g., *data.gov* dataset #92.

Table 1 summarizes size and machine-readability of the studied catalogues.

Table 1. General characteristics of catalogues

Catalogue	Size	Machine readability
data.gov	1320	CSV
data.gov.uk	2879	RDFa, CSV
data.govt.nz	251	Feeds
data.australia.gov.au	69	RDFa
datasf.org	132	–
data.london.gov.uk	189	–
statcentral.ie	227	–

2.3 Dataset Metadata

Next, we examine structure, consistency and availability of the metadata that makes up the catalogues. A detailed analysis of metadata quality in government catalogues is out of the scope of this paper. We focused on identifying metadata

properties that are used consistently across different catalogues, and on understanding the level of control applied to the values of various metadata fields.

Metadata structure. We looked at what properties are used to describe each dataset. Table 2 summarizes common properties used across the studied catalogues. We find that, although widely different terminology is used to label metadata fields, close examination reveals large overlap in the used fields.

Table 2. Metadata structure of catalogues

	General												Categorization		Access			Other		
	title	description	publisher	frequency	release date	update date	temporal coverage	geographic coverage	license	data dictionary	granularity	metadata update	theme	tags/keywords	dataset URL	format	size	references and citation	quality characteristics	data collection charact.
data.gov	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y		Y	Y	Y	Y	Y	Y	Y	Y
data.gov.uk	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y		Y	Y	Y	Y		Y	Y	
data.govt.nz	Y	Y	Y		Y				Y				Y	Y	Y	Y				
data.australia.gov.au	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y
datasf.org	Y	Y	Y	Y	Y		Y			Y			Y	Y	Y	Y				
data.london.gov.uk	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y		Y		
statcentral.ie	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y		Y	Y	Y

Metadata consistency. In Table 3, we summarize the findings on metadata consistency on a qualitative scale. Date fields are considered consistent if they follow consistent syntactical format within a catalogue. Other properties are considered consistent if their values are drawn from a fixed set of options (controlled vocabulary). Absence of such control is evident when multiple values refer to the same entity, e.g., "U.S." and "United States". Many catalogs do not even maintain syntactical consistency of date values.

Metadata availability. Providing a sparse set of property values, where the value is often missing or has an uninformative value like "not specified", adversely affects the usefulness of metadata. Table 3 shows the percentage of availability of values for a set of common properties across catalogues.

Dataset categorization. All catalogues use both themes (broad catagories, usually functional domains like *Education* or *Health*), and tags or keywords to categorize datasets. Table 3 shows that while themes are always chosen from a controlled vocabulary, tags are not. Themes enable intuitive browsing of datasets

Table 3. metadata consistency and availability. +,= and – represent high, medium and low consistency. Numbers represent the percentage of datasets in a catalogue for which the metadata attribute is specified.

	Metadata consistency							Metadata availability				
	geographic coverage	temporal coverage	frequency	release date	update date	theme	tags	geographic coverage	release date	license	frequency	tags
data.gov	–	–	–	–	–	+	–	95	79		99	100
data.gov.uk	+	+	–	=	=	+	–	99	52	100	52	94
data.govt.nz				+		+	–		100	98		100
data.australia.gov.au	–	=	=	+	+	+	=	81	70	93	8	68
datasf.org		–	+	–		+	–		100		38	100
data.london.gov.uk	+		+	–	+	+	–	93	95	91	94	62
statcentral.ie	+	–	–	+	+	+	–		100		100	100

and give an instant overview of the available data in a catalogue. A sufficient description of a catalogue should clearly distinguish themes from keywords.

2.4 Dataset Accessibility

Catalogues do not always provide direct download links for datasets. The data might be available only after accepting a click-through license, or there might be a splash page that lists the parts of a multi-file download, or data access might require use of a web service[8]. While direct download links are available for virtually all datasets in *data.london.gov.uk* and for 95% in *data.gov*, they are provided only for about 10% in *datasf.org* and 7%[9] in *data.gov.uk*. A vocabulary should support a distinction between direct and indirect download links as this is required for scenarios that involve bulk processing of datasets.

Catalogues provide data in different formats. While some of them are machine-readable, others are not (e.g., PDF, HTML wizards). The format of datasets is very important to mashup developers and for bulk processing of the data and should be explicitly expressed.

3 The *dcat* Vocabulary

Based on the survey described in the previous section, we have developed an RDF Schema vocabulary that allows the expression of data catalogues in the

[8] E.g. NextMuni XML data at datasf.org is available as RESTful web service: `http://www.datasf.org/story.php?title=nextmuni-xml-data`

[9] Because of the large size of the catalog, we examined only a random sample of 75 data.gov.uk datasets.

RDF data model. We have chosen RDF because (i) most of the use cases considered in Sect. 1 involve querying of aggregated data, which is well-supported in RDF; (ii) re-use and extension of existing metadata standards such as Dublin Core is straightforward in RDF; and (iii) for compatibility with Linked Data [8]. The use of more expressive formalisms such as OWL ontologies was considered unnecessary because the goal is not domain modelling or reasoning, but interoperable data exchange. Classes and properties from existing vocabularies, especially Dublin Core, were re-used whenever possible[10]. Here we will briefly describe the vocabulary's main classes, as shown in Fig. 1. Full documentation is available online[11].

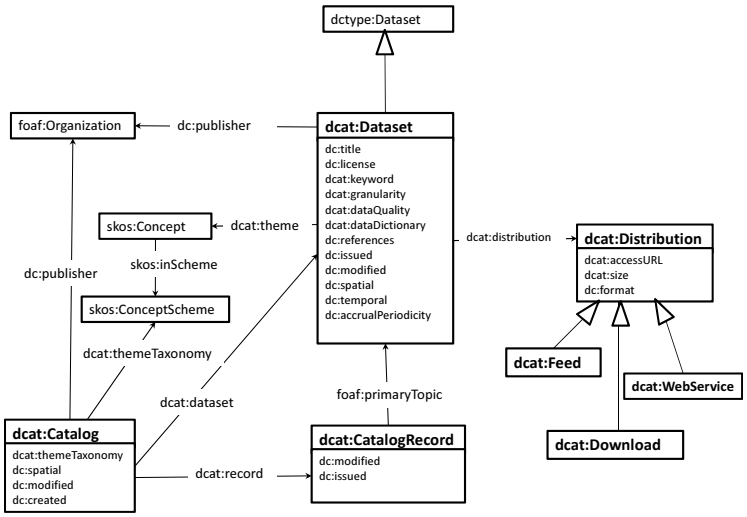

Fig. 1. Overview of the *dcat* vocabulary

dcat:Catalog A Catalog represents a collection of dataset descriptions. A Catalog does not own or provide the actual datasets, but provides a structured description of them. Properties used to describe Catalog include `dcat:themeTaxonomy`, `dc:modified`, `dc:issued`, `dc:publisher` and `dc:spatial`.

dcat:Dataset A Dataset represents a collection of data which is published or will be published. Properties used to describe Dataset include `dcat:theme`, `dcat:keyword`, `dcat:granularity`, `dcat:dataDictionary`, `dcat:dataQuality`, `dc:modified`, `dc:issued`, `dc:license`, `dc:publisher` and `dc:references`.

[10] In accordance with RDF conventions, we use *QNames* to identify terms. Terms beginning with *dc*: are part of Dublin Core, terms beginning with *dcat*: are defined in our vocabulary. Other prefixes have their conventional meaning.

[11] `http://vocab.deri.ie/dcat`

dcat:CatalogRecord One source of ambiguity in catalogues results from the absence of clear distinction between a dataset and its corresponding description in the catalogue. For example, does the "last update" refer to the actual data of the dataset or to its description in the corresponding catalogue? Having a stand-alone entity for a catalogue record resolves this ambiguity and allows adding further description about the metadata provided by the catalogue for a specific dataset. Properties used to describe `CatalogRecord` include `foaf:primaryTopic`, `dc:modified` and `dc:issued`.

dcat:Distribution A `Distribution` represents the availability of a dataset in a particular format. `accessURL` property refers to the dataset location. `Download`, `Feed` and `WebService` refine `Distribution` to indicate availability in those forms. Properties used to describe `Distribution` include `dcat:accessURL`, `dcat:size`, `dc:format`.

Dataset Categorization The `theme` and `keyword` properties describe categorization of datasets. It is recommended to use a controlled vocabulary or taxonomy described using SKOS for the theme. Cross-catalogue browsing by theme can be enabled by mapping the local scheme to standardized schemes such as the SDMX List of Subject-matter Domains[12] or the Integrated Public Service Vocabulary (IPSV)[13].

4 Feasibility Study

To verify our claim that different catalogues can be rendered in the *dcat* vocabulary, we applied the vocabulary to represent the *data.gov*, *data.australia.gov.au*, *data.london.gov.uk* and *datasf.org* catalogues in RDF[14]. After importing the catalogues into a relational database, we used D2R Server[15] for generating RDF data. D2R Server also provides a SPARQL endpoint for querying the data and an HTML interface for browsing. D2R Server's basic out-of-the-box Linked Data enabled web interface, without any customisation, already provides functionalities that are not available on many of the catalogue websites, such as browsing by category/keyword and browsing by agency.

Figure 2 shows an RDF snippet describing a *data.gov* dataset and its availability as a downloadable XML file. We use the Turtle RDF syntax.

Figure 3 shows a SPARQL query to retrieve all datasets about health which have XML distribution, and the results of the query[16]. Results come from both *data.london.gov.uk* and *data.gov*. Such cross-catalogue queries are enabled by the common representation in RDF.

The data can be enriched by linking it to other available Linked Datasets, to enable further useful queries and navigation vectors. Figure 4 shows a SPARQL

[12] `http://sdmx.org/wp-content/uploads/2009/01/03_sdmx_cog_annex_3_smd_2009.pdf`

[13] `http://www.esd.org.uk/standards/ipsv/`

[14] `http://lab.linkeddata.deri.ie/govcat/`

[15] http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/

[16] Queries can be tested at `http://lab.linkeddata.deri.ie/govcat/snorql/`

```
:data.gov/dataset/1263 a dcat:Dataset ;
     dc:title           "FinancialStability.gov TARP..."  ;
     dc:accrualPeriodicity "approximately twice weekly" ;
     dc:modified          "2009-12-04"^^xsd:date ;
     dc:publisher         :data.gov/agency/Department_of_the_Treasury ;
     dc:temporal          "October 2008 - present" ;
     dcat:dataDictionary <http://www.financialstability.gov/impact/...> ;
     dcat:granularity     "Financial transactions" ;
     dcat:distribution    :data.gov/1263/distribution/2566 ;
     dcat:keyword          "tarp", "cbli" ;
     dcat:theme           :data.gov/category/banking_and_insurance ;
     foaf:homepage        <http://www.data.gov/details/1263> .

:data.gov/1263/distribution/2566 a dcat:Download ;
     rdfs:label           "text/xml distribution of FinancialStability" ;
     dc:format            "text/xml" ;
     dcat:accessURL    <http://www.financialstability.gov/impact/cbli.xml> ;
     dcat:size            [ dcat:bytes 4 ] .
```

Fig. 2. Sample RDF description of a dataset

query that retrieves all *data.gov* datasets published by an agency having a budget of more than 50 billion. Budget information is obtained from DBpedia[17].

5 Related Work

We identify related work in the areas of catalogue aggregation, metadata standards for documents, and from the linked data field. We briefly discuss these efforts below.

Following the need for horizontal access to federation of catalogues, efforts to aggregate catalogues started to emerge recently, most notably Guardian's *World Government Data site*[18] and Sunlight Labs' *National Data Catalog*[19]. However, lack of a standardised model obliges these solutions to rely on coding a custom importer per catalogue. Imported catalogues are then translated to some proprietary unified model defined for the federated catalogue. This limits flexibility, reusability and extensibility as it substantially increases the required effort for each new catalogue addition.

Many metadata standards for document description already exist, like e-GMS [2] and AGLS [3]. These standards were motivated by the concerns of document management, so they closely resemble the widely-used Dublin Core standard. Such standards provide rich properties for resource description but

[17] http://dbpedia.org/About
[18] http://www.guardian.co.uk/world-government-data
[19] http://nationaldatacatalog.com/

```
SELECT DISTINCT ?title ?url
WHERE {
     ?dataset a dcat:Dataset;
     dct:title ?title;
     dcat:theme ?theme;
     dcat:distribution ?distribution.
     ?distribution dcat:accessURL ?url;
     dct:format ?format.
     ?theme skos:prefLabel ?themeLabel.
     FILTER regex(?themeLabel, "health", "i").
     FILTER regex(?format,"text/xml")
}
```

title	url
"Census 2001 Key Statistics 08: Health"	<http://data.london.gov.uk/datafiles/demographics/census-2001-ks08-borough.xml>
"Census 2001 Key Statistics 21: Long Term Illness"	<http://data.london.gov.uk/datafiles/demographics/census-2001-ks21-borough.xml>
"Legal Abortion Rates"	<http://data.london.gov.uk/datafiles/health/abortion-legal-rates-pct.xml>
"Alcohol Related Hospital Admissions"	<http://data.london.gov.uk/datafiles/health/alcohol-admissions-borough.xml>
"Crime Attributable to Alcohol"	<http://data.london.gov.uk/datafiles/health/alcohol-crime-borough.xml>
"Land Transport Deaths due to Alcohol"	<http://data.london.gov.uk/datafiles/health/alcohol-deaths-transport-borough.xml>
"Hazardous, Harmful and Binge Drinking Rates"	<http://data.london.gov.uk/datafiles/health/alcohol-drinking-rates-borough.xml>
"Alcohol Related Hospital Admissions Indicators"	<http://data.london.gov.uk/datafiles/health/alcohol-indicators-borough.xml>
"Alcohol Related Mortality"	<http://data.london.gov.uk/datafiles/health/alcohol-mortality-borough.xml>
"Births and Fertility Rates"	<http://data.london.gov.uk/datafiles/health/births-fertility-rates-borough.xml>
"Births with Low Birthweight"	<http://data.london.gov.uk/datafiles/health/births-low-weight-borough.xml>
"Births by Birthplace of Mother"	<http://data.london.gov.uk/datafiles/health/births-mother-birthplace-borough.xml>
"Benefit Claimants due to Alcoholism"	<http://data.london.gov.uk/datafiles/health/claimants-alcoholism-borough.xml>
"Working Age Disability"	<http://data.london.gov.uk/datafiles/health/disability-working-age-borough.xml>
"MyPyramid Food Raw Data"	<http://www.cnpp.usda.gov/Innovations/DataSource/MyFoodapediaData.zip>

Fig. 3. SPARQL query across catalogues, with results

would require subsetting or other additional guidelines before they can be used to describe data catalogues.

The Linked Data community has developed a number of vocabularies for the description of datasets. VoiD [6], SCOVO [12] and SDMX-RDF [9] are RDF vocabularies for the description of RDF datasets and statistical datasets, respectively. They are not intended for describing other kinds or formats of datasets and hence are not applicable to data catalogues, but can be used in conjunction with *dcat* to provide more detailed descriptions of applicable datasets. A simple RDF vocabulary for the description of data catalogues has been defined at CTIC[20], but unlike *dcat* it just defines 2 basic classes and no metadata fields. Sunlight Labs also suggests a list of common properties to use for dataset description[21]. This list is not defined in any standard way, limited to datasets description and not comprehensive.

Koumenides et al. [13] also approach the problem of integrating data catalogues using RDF, and contribute an in-depth review of the literature and public policy setting that surrounds the data catalogue phenomenon. They convert a

[20] `http://data.fundacionctic.org/vocab/catalog/datasets.html`
[21] http://sunlightlabs.com/blog/2010/drafting-guidelines-government-data-catalogs/

```
SELECT ?title
WHERE {
     :data.gov dcat:dataset ?dataset.
     ?dataset dc:title ?title;
     dc:publisher ?agency.
     ?agency dbpedia:budget ?budget.
     FILTER (?budget>50000000000)
}
```

Fig. 4. SPARQL query integrating an external dataset (DBpedia)

number of data catalogues to RDF and compare them both quantitatively (e.g., growth over time) and qualitatively (e.g., by visualisation via tag clouds), but do not continue their work to tackle the problem of developing a unified RDF vocabulary.

As already discussed, our proposed solution goes beyond the state of the art as a) it takes into account and reuses many terms of the vocabularies mentioned above, b) extends their catalogue representation by a richer description of not only datasets but also of catalogues and data files, c) presents a formal description of a data catalogue in RDF, d) enables a single importer to be used to import all catalogues that support the format, e) makes the federation process loose and easy to participate as the individual catalogue owners need only to map their metadata fields to *dcat*.

6 Future Work and Conclusion

We proposed the *dcat* RDF vocabulary as an interchange format to enable standardised description of government data catalogues as part of the nascent Web of Data. To identify the most relevant concepts to include in such vocabulary, we analysed a number of existing data catalogues. We have also demonstrated how our vocabulary can be used to allow cross-catalogue querying and browsing over four major data catalogues.

Besides refining the *dcat* vocabulary and validating it against more catalogues, particularly interesting areas for future work include the exploration of dcat's potential for improved user interfaces over integrated catalogues in situations where structural information about the datasets is available, e.g., for datasets in RDF format, or highly-structured statistical and geographical datasets; and as a driver for discovering cross-links between datasets within a catalogue and to linked datasets elsewhere on the Web.

The operators of data catalogues tend to aggressively pursue an agenda of openness and are technologically sophisticated. Many more data catalogues are likely to appear in the near future. We believe that the cost for implementing *dcat* is low, especially when deployed as embedded RDFa. We therefore expect that *dcat* can play an important role in facilitating wider re-use of government

data. To achieve this goal, the W3C's eGovernment Interest Group has set up a task force to support the further development and deployment of *dcat*[22]. A first successful outcome is the adoption of *dcat* on the `data.gov.uk` site[23].

Acknowledgements

The work presented in this paper has been funded in part by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2) and the European Union under Grant No. 238900 (Rural Inclusion).

References

1. C. European Parliament, "Directive 2003/98/ec on the re-use of public sector information" (2003)
2. e-Government Metadata Standard, version 3.1 (2006), Cabinet Office, e-Government Unit, UK.
3. AGLS Metadata Standard (2008), `http://www.agls.gov.au/documents/terminology/`, National Archives of Australia
4. Open Government Directive. Memorandum for the heads of executive departments and agencies (2009), office of Management and Budget, Washington, D.C.
5. Putting the frontline first: smarter government. UK Government (2009), `http://www.hmg.gov.uk/media/52788/smarter-government-final.pdf`
6. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Describing linked datasets. In: WWW 2009 Workshop: Linked Data on the Web (LDOW2009) (2009)
7. Bennett, D., Harvey, A.: Publishing open government data. W3C working draft, World Wide Web Consortium (2009), `http://www.w3.org/TR/gov-data/`
8. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data: The story so far. International Journal on Semantic Web and Information Systems (IJSWIS) 5 (2009)
9. Cyganiak, R., Field, S., Gregory, A., Halb, W., Tennison, J.: Semantic statistics: Bringing together SDMX and SCOVO. In: Proceedings of the Linked Data on the Web Workshop (LDOW2010). Raleigh, NC (April 2010)
10. Dekkers, M., Polman, F., te Velde, R., de Vries, M.: MEPSIR: Measuring european public sector information resources. final report of study on exploitation of public sector information. Tech. rep. (2006), `http://ec.europa.eu/information_society/policy/psi/docs/pdfs/mepsir/final_report.pdf`
11. Ding, L., DiFranzo, D., Graves, A., Michaelis, J.R., Li, X., McGuinness, D.L., Hendler, J.: Data-gov Wiki: Towards Linking Government Data. In: The 2010 AAAI Spring Symposium on Linked Data Meets Artificial Intelligence (2010)
12. Hausenblas, M., Halb, W., Raimond, Y., Feigenbaum, L., Ayers, D.: SCOVO: Using statistics on the web of data. In: 6th European Semantic Web Conference (ESWC2009) (2009)
13. Koumenides, C., Alani, H., Shadbolt, N.: Global integration of public sector information. In: Proc. of WebSci10 (2010), `http://journal.webscience.org/303/`
14. Wilde, E., Kansa, E., Yee, R.: Web Services for recovery.gov. Tech. Rep. 2009-035, UC Berkeley School of Information (October 2009)

[22] `http://www.w3.org/egov/wiki/Data_Catalog_Vocabulary`
[23] `http://data.gov.uk/blog/project-update`