ALGIM

ASSOCIATION OF LOCAL GOVERNMENT INFORMATION MANAGEMENT

LINKED DATA

FOR INTERLINKED LOCAL GOVERNMENT

Association of Local Government Information Management White Paper

Contents

EXECUTIVE SUMMARY

There is an increasing demand for local authorities to open and share data that complies with standards for meeting open and transparency accords by national governments. Traditionally, this has been achieved through proprietary data management and systems, which has been fraught with difficulties.

Linked Data is an approach of connecting and publishing structured data from different data sources by using standard Web technologies. Data that was not previously connected or shared before can be interlinked and queried across multiple data sources. Also included are semantics of the data by applying vocabularies of terms to define the data. This white paper describes a local government approach to the supply and demand of Linked Data and the supporting architecture that underlies this technology.

The uptake of Linked Data has been slow by organisations, as generally they are unaware of this simple, yet effective semantic technology and number of successes that this approach provides. The organisations that supply Linked "Open" Data enjoy greater interoperability, re-use, accessibility, and discoverability of their information asset, where other potential opportunities can be realised.

Based on analysis by Gartner, Linked Data and semantic technology has been identified as a strong participant in the future of the Web. Local authorities adopting this approach can be content with the knowledge that their data asset will be a solid information service provided to communities for many years to come.

INTRODUCTION

This white paper describes a local government Linked Data approach, supported by standards and best practice, to enable local authorities to become Linked Open Data[1] enterprises.

It describes the challenges local authorities face when endeavouring to share or exchange data across boundaries. The Linked Data approach of interlinking data and the benefits gained are introduced.

The white paper also describes future information trends, links to other projects, finishing with a call to action. This white paper has been written by the ALGIM working party.

PROBLEM DEFINITION

There is an increasing demand for the local government sector to share data and regulatory information frameworks at national and international levels. With existent heterogeneous data sources from disparate silo systems lacking common data schemas, combining data sources is a big challenge using traditional methods.

Business Challenges

The common problems that local authorities find challenging when sharing information are:

- Communities cannot agree on common terms within a domain
- Data hugging (e.g. questioning the accuracy of the data therefore not releasing it)
- Limited resources (i.e. traditional approaches to sharing data incurs high costs)
- Open data can be published, but what are the sharing capabilities
- Proprietary data formats that are not standards compliant

[1] Linked Open Data refers specifically to data that is open and publically available and appropriately licensed.

- Traditional legacy heterogeneous systems and methods.

There is also increasing pressure from central government for local government to increase:

- Transparency and accountability [19]
- Democratic engagement and participation [19]
- Access to services
- Digital services and data sharing.

A Unified Approach Is Needed

The Evolving Web

In the early days of Web 1.0, users viewed content in the form of interlinked web pages. The majority of existing sites are now Web 2.0 that cater for user interaction and collaboration particularly with the explosion of mobile and social media technologies. The next phase of the Web evolution is Web 3.0, which accommodates the Web of data and semantic searches, thereby connecting data, information, concepts and applications. The Web is continuing to evolve with no set endpoint in sight.

By 2017, Web-scale IT will be an architectural approach found operating in 50 percent of global enterprises. Gartner (2014)

Open Government

As there is more pressure to open data and make it more accessible, national governments are forming partnerships with international groups, for example, the Open Government Partnership (OGP). The OGP is a group of governments that aim to promote transparency, empower citizens, fight corruption and harness new technologies to strengthen governance [1]. With regard to data access, they intend to increase the availability of information about governmental activities that facilitates open standards, interoperability, reuse and public feedback.

The Open Data Barometer,[2] created by the World Wide Web Foundation, monitors open data initiatives around the world, particularly open data that is accessible, machine-readable and re-usable. Their findings show, even though in 2013 there was a global movement to make government data 'open by default', the uptake has been slow. The countries that have achieved high capacity open data have open data policies and strong political backing. They also have extended open data practices across government departments, and increasingly at local government levels. Open data was achieved with greater success where governments took open data activities to the local or city-levels [2]. An example of this is in the UK, where a "Release of Data Fund" has been set up to support the wider release of public open data with a strategy for local government open data [3]. Local authorities in the UK have opened up datasets that are accessible, queryable and available for reuse on the Web by using the Linked Data method.[3]

LINKED DATA

Linked Data is an approach of connecting and publishing structured data from different data sources by using Web technology. This means data that was not previously connected or shared before can

[2] http://barometer.opendataresearch.org/
[3] http://opendatacommunities.org/

be interlinked and queried across multiple data sources. Linking data removes barriers bounded within traditional methods of data exchange and opens up data 'silos'. This section describes Linked Data and the technologies that support it.

Linked Data Components

Linked Data consists of three main components that are grounded on standard Web technology. These are: HTTP, URIs, and RDF.

- *Hypertext Transfer Protocol* (**HTTP**) is a mechanism for transferring files by applying rules for formatting and transmitting messages on the Web [4].
- *Uniform Resource Identifiers* (**URIs**) identifies the name of a resource on the Web. By entering a URI in a browser, HTTP fetches and transmits the requested resource in a machine-readable or human-readable form. For example, a human readable form may be represented as a hypertext document (i.e. HTML page), and the machine-readable version at the same URI is encoded information about the data (e.g. RDF/XML) [5].
- *Resource Description Framework* (**RDF**) is a standard model for data interchange on the Web, where the datasets are published as *triples* [8].

Figure 1 shows an example of a simple sentence formed as a triple that contains a *subject*, *predicate* and an *object*.

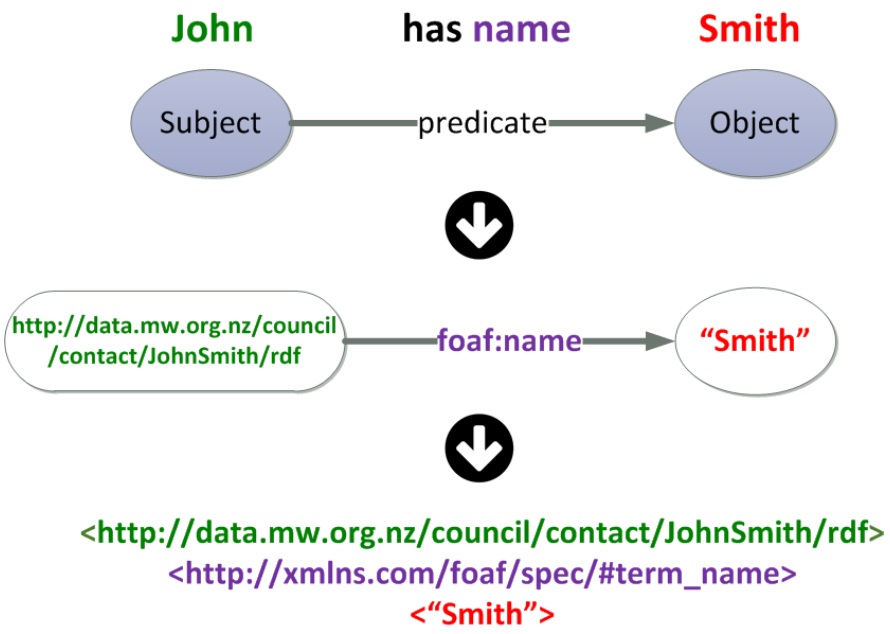

Figure 1: Example of a simple triple

The *subject* is the URI representing a resource and the *object* can be a value or a URI. The subject and object can be regarded as the two ends of the link. In the example in Figure 1, the *subject* is a URI and the *object* is the value "Smith". The *predicate* is also a URI that represents the relationship between the two ends of the link. The predicate in Figure 1, is the URI representing the term 'name' that has already been defined in the *Friend Of A Friend*[4] (**FOAF**) vocabulary.

[4] http://xmlns.com/foaf/spec/

Vocabularies

Vocabularies not only provide valuable meaning to Linked Data, but also describe the relationships between data sources. The adoption of well-known vocabularies give the advantage of not having to redefine terms that already exist and can be applied as common terms across many datasets. For example, the Friend of a Friend (FOAF) vocabulary defines basic terms for a 'Person' (e.g. name, title, age). Another example, is the Dublin Core[5] vocabulary that defines standardised terms for metadata (e.g. publisher, license, description). If terms cannot be found in well-known vocabularies, then vocabularies can be custom-built. For example, well developed thesauri and controlled vocabularies in use by records managers and archivists, link into their functional classification schema. Subsequently, all aspects of the data sources can be described, particularly with data integration by optimising semantics.

There are many methods for modelling vocabularies visually; an example of a method is described in Appendix B.

Table Data to Linked Data

Traditional methods for storing data are either in a hierarchy (e.g. XML) or in a relational database (e.g. SQL). For Linked Data, RDF describes the data structure as a type of database called a graph [6]. The RDF graph is a collection of triples, where the nodes of the graph are its subjects and objects.

Most organisations store data using traditional database methods. Access to relational databases as read-only RDF graphs can be accomplished through many semantic Web tools. Some tools (e.g. D2RQ platform[6]) enable browsers to navigate and query the content of the database by mapping the database schemas to vocabularies and transforming the database to an RDF graph [7]. Figure 2 depicts an example of how data components from a table are mapped to a triple (i.e. the subject, predicate and object).

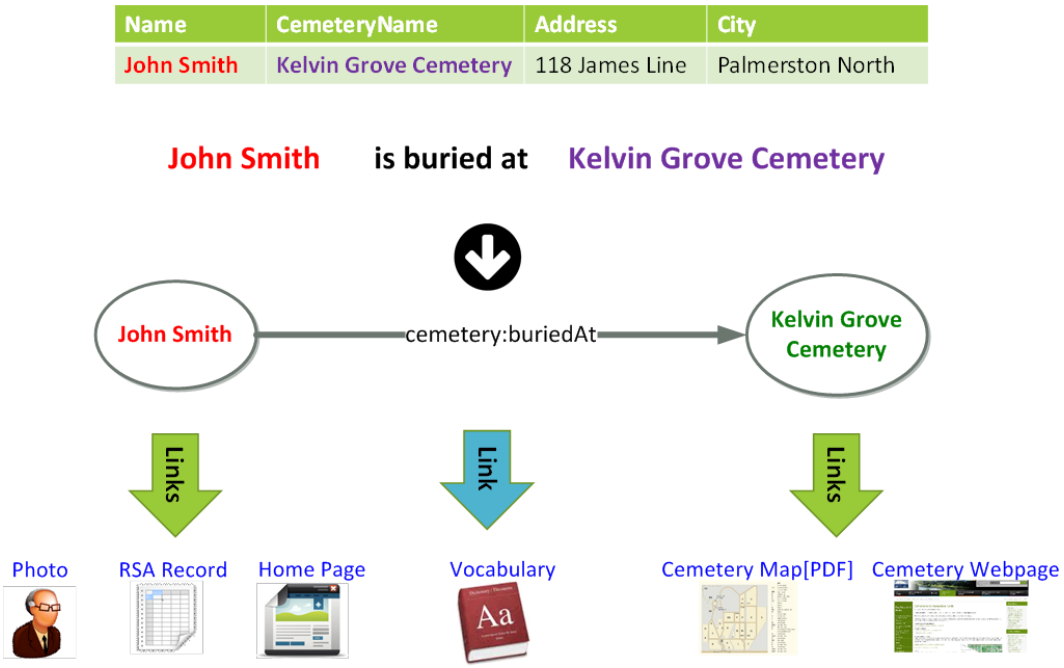

Figure 2: Example of an RDF triple within a database table

[5] http://dublincore.org/documents/dcmi-terms/
[6] http://d2rq.org/

This basic example from Figure 2 highlights the potential data combinations within the table and linkages to other data sources resulting in the creation of a collection of triples or RDF graph database.

> *Using this simple triple model (i.e. subject, predicate, object), allows structured and semi-structured data to be mixed, exposed, and shared across different applications.*
>
> *RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution* of *schemas over time without requiring all the data consumers to be changed.* W3C[8]

The range of Local Government services can appear complex and diverse. In the UK, the Local Government Service List[7] lists over 700 separate types of services that a council can provide. This could lead to hundreds of disconnected data definitions. However, the underlying data tends to follow a similar pattern. The British Standards Institute have developed a high level concept model for Smart Cities (PAS182 Smart City Concept Model[8]) from which data sharing vocabularies can be derived, leading to improved discoverability.

What Can Linked Data Be Used For?

The Linked Data approach can be applied across many contexts within an organisation's business. It can be applied within or used as a mechanism to publish open data. Figure 3 identifies these organisational contexts and within each the degree of openness that is generally applicable to the data (i.e. public or protected).

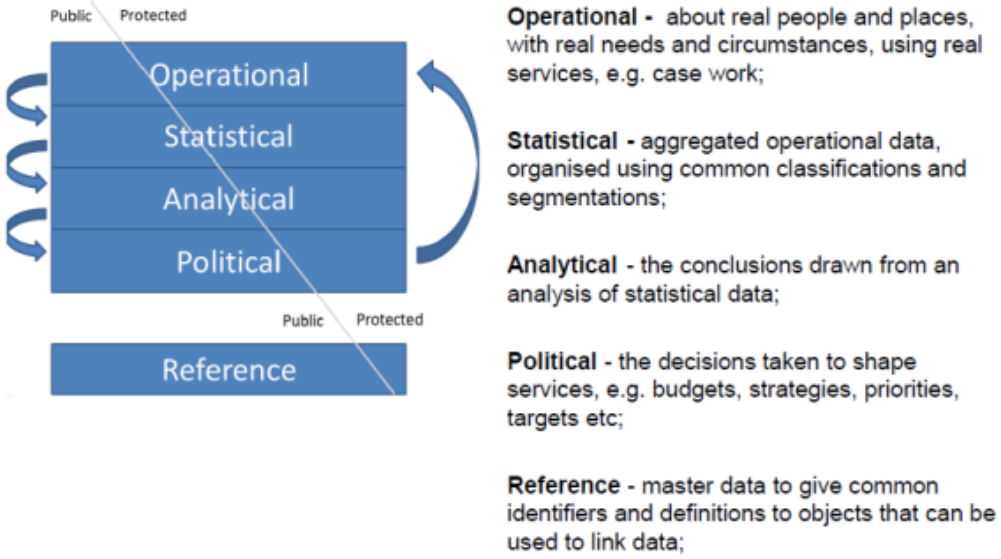

Figure 3: Context Model for Linked Data in an Organisation

[7] http://standards.esd.org.uk/?uri=list%2FenglishAndWelshServices
[8] http://www.bsigroup.com/en-GB/smart-cities/Smart-Cities-Standards-and-Publication/PAS-182-smart-cities-data-concept-model/

Problems Linked Data Cannot Solve

In some cases, there are better ways to solve business problems not using Linked Data. For example, sharing private operational data in real-time identified in Figure 3. Linked Data works well in an environment that is open for everyone to consume for their own intentions, but sharing private operational data in real-time between trusted organisations may be best achieved with a defined API and schema using an authenticated service over a secure network.

Alternative Technologies

Another method of publishing open data is using the Open Data Protocol or OData[9] initiated by Microsoft in 2007. OData is a RESTful service that exposes data sources to be queried and sharable across disparate systems to be consumed by RESTful APIs. OData is based on OASIS[10] standards and utilises HTTP, Atom, JSON and URIs to address and access data sources [9]. OData creates data silos (i.e. a single source) where discoverability is not easy [10].

TECHNICAL DETAILS

This section describes a Linked Data lifecycle for local authorities to follow when implementing the semantic Web technology. This section is divided into two basic parts: 1) the lifecycle processes for producing Linked Data and making it available and 2) the Linked Data architecture that supports the production of Linked Data.

Linked Data Lifecycle

There are several *government* Linked Data lifecycles available that are based on experience of professionals who exposed existing government datasets using Linked Data[11]. The lifecycle model shown in Figure 4 is an adaptation of these approaches, but focuses on local government.

Figure 4: Local Government Linked Data Lifecycle [14]

[9] http://www.odata.org/
[10] https://www.oasis-open.org/
[11] http://www.w3.org/2011/gld/wiki/GLD_Life_cycle

The key functions in Figure 4 can be classified into two groups: supply (blue) and demand (green) [13]. The impetus on supply and demand within the lifecycle means there will be a strong emphasis on integrating data and applying user focused applications for consumption with mechanisms for the community to provide feedback.

Table 1 describes the lifecycle function for the supply and demand of Linked Data.

Group	Process	Description
Supply	Specify [11]	• Prepare and educate stakeholders to engage them in the lifecycle processes [14]. • Identification and analysis of the raw data. • URI design and construction in accordance with best practice guidelines for persistent naming strategies and implementation plans (refer to government guidelines, if available). *Refer to Appendix C.* • Data versioning strategies (i.e. data to be published for different time periods). • Specify an appropriate open license in accordance to government recommendations. *Refer to Appendix C.*
	Model [11]	• Search for suitable standard vocabularies/ ontologies for reuse. • Search for non-ontological resources (e.g. catalogues, domain related sites, authoritative websites). • Model and build a new vocabulary (only if required).

	Publish [12, 14]	• Transform raw data to RDF.
		• Cleanse data. For example, clean errors that occur for:
		  ○ HTTP.
		  ○ Incorrect namespaces for ontologies.
		  ○ Incompatible datatypes.
		• Link:
		  1. Identify dataset as linking targets.
		  2. Discover relationships between data items.
		  3. Validate the relationships discovered.

Demand	Discover [14]	• Dataset publication: tools for storing RDF, SPARQL endpoints and Linked Data frontends. • Metadata publication: metadata about Linked Data. Describe the data resources to a minimal standard. • Data discovery: register Linked Data with data registries and catalogues (e.g. create sitemaps and submit to Google) [12].
	Integrate [13]	Integrate various Linked Data resources using SPARQL specifications.
	Consume [13]	• Frontends for using the Linked Data in an application. • Applications developed from Use Cases for providing information and functions to consumer. • Linked Data search engines allow the search on Web of Data [20].

Table 1: Descriptions of Lifecycle components

Linked Data Architecture

There are other important elements that support the Linked Data Lifecycle and need to be established within an organisation to maximise the benefits of Linked Data output. These are:

- ***Governance***
  Governance is critical to effectively managing the Linked Data resource. The governance function manages the assets and the performance of the Linked Data functions. Also, informs and educates user of benefits of Linked Data.

- ***Stewardship***
  Data Stewardship is the role that ensures information and metadata meet high levels of quality, accuracy, format and value criteria; ensuring that information is properly defined and understood (i.e. standardised).

- ***Best Practice and Standards***
  Best practice and standards for Linked Data are published documents that establish standardised methods for design, development and publication of Linked Data. These are universally understood and adopted and can be identified as consistent and optimises interoperability between heterogeneous data sources. These cover metadata, data licensing and digital preservation.

- ***Infrastructure***
  The infrastructure refers to the fundamental services, facilities and systems that must be implemented before raw data can be published and accessed as Linked Data. This also includes implementing and administering security procedures. There are many software tools available (including many open source tools) for generating RDF and ontologies.

- ***Communities***
  Communities are groups that explore, query interlinked data and develop applications (i.e. APIs). Communities cover a broad range of areas, e.g. developers, data suppliers and public. Communities will define existing and future requirements for Linked Data based solutions to meet their needs. Engaged communities have the ability to supply feedback (e.g. data issues) and provide other data sources (e.g. social media) to add value the data sources.

CASE STUDIES

This section identifies Linked Data case studies from the local government sector.

1: Manawatu-Wanganui Region (NZ)

Councils within the Manawatu-Wanganui region collaborated in a spatial shared services project that focused on the exchange of cemetery data from disparate platforms. After identifying Use Cases and the lack of common entities within data schemas, it was concluded the use of Linked Open Data technology using the semantic web was the best means of making all cemetery information held by each council available to the public with minimal impact to each Council's resources.

To achieve this, the D2RQ platform was connected to the various relational databases and data schema that each individual council operates with. This resulted in no change to existing council staff processes, and provided another means for data to be shared openly.

Ontologies were developed in collaboration with the University of Brighton and applied to each member council SQL to RDF mapping. The resulting RDF output included semantics, licencing, and provenance information for every view of the data, whether at the individual record level or whole data set.

2: Sedgemoor District Council (UK)

The project has taken published documents about the impacts, and mitigating projects, of the site preparation and build of the proposed Hinkley Point Power Station, and republished them as 'Linked Data' using an open source publishing platform.

Web application: http://www.sedgemoor.gov.uk/cim

Data model and Linked Data: http://data.sedgemoor.gov.uk

3: Auckland Museum

Auckland Museum (NZ) has released its records as open data, free to the user. In doing so, it is among the first museums in the world to exploit the advantages of 'Linked Open Data', sharing knowledge and presenting unprecedented opportunities for learning, research, scholarship and general interest. 'Linked Data' ensures that collections are transparent, automatically interconnected and easier and faster to search.

Web application: http://api.aucklandmuseum.com/v1/doc

BENEFITS AND RISKS

This section identifies the benefits (Table 2) and risks (Table 3) with the deployment of Linked Data within an organisation**.**

BENEFITS	Description
Agile	• Data published as Linked Data can be extended to address specific needs without URIs being affected. • Applications that consume Linked Data can consume all or parts of the datasets. • Vocabularies can evolve into standards or best practice and be applied to all or part of datasets.
Non-disruptive	Linked Data does not disrupt current information systems and / or data collection practices. Minimal impact to business as usual.
Standards & best practice	• Complies with web standards. • Applicable to other authoritative well-known vocabularies. Customised vocabularies may evolve into standardised terms by governing communities.
Discoverable	• Data and web content are interlinked and can be easily navigated by users. • Registries and catalogue assist Linked Data harvesting.
Describable	• Linked Data is self-describing (i.e. data relationships, terms and metadata are described using vocabularies and link to other authoritative sources giving context). • Provides data interpretation, transparency and accountability. • Value added.

Reusable	• Reuse of well-known vocabularies and Linked Data can be consumed by multiple applications. • Published Linked Data can be consumed by multiple applications.
Interoperable	• Links are created between existing data and other web content that is human and computer readable. • Queries integrate data from disparate sources. • No specific application schemas are required for publishing purposes, vocabularies may be applied for common terms shared within applications.
Open	• Complies with 5 star open data deployment scheme.[12] • Satisfies various government and governance groups (e.g. OGP) requirements for open data.
Exposable	Exposes existing data held in silos and federating data content that is accessible and usable.
Opportunities	The creation of new data linkages, where data previously did not link before, provides greater potential for new innovative applications.
Scalable	Implement small-scale or large-scale data sources with the ability to extend and add to in the future.
Economical	Minimal expenditure required with a variety of software tools that are open source with supported communities.
Future proofing	Uses the full capability and potential of the Web (e.g. Web 3.0). That is, enabling communities to engage with councils and their decision making by using social media technology.
Governance	• Greater governance of information on a national scale, e.g. protected records, and standardised application of retention schedules. • Standardised terms (i.e. taxonomies and thesauri) for sharing, including standardised encoding schemes (e.g. date formats and data quality codes) developed into national vocabularies. • To build a high level concept model for local government data would be a useful template to build Linked Data solutions to maximise the opportunities to make useful joins.

Table 2: Linked Data Benefits

The risks with associated solutions are listed in Table 3.

RISK	Solution / Mitigation
Uptake of staff and consumers	• Education by Linked Data governance groups. • Exploiting discoverability. • Engaging developer's communities (e.g. hackathons).
Too many tools to choose from	Need for a local government toolkit recommending tools suitable for implementation.
Lack of guidance for developing persistent URIs	Establish Government Governance Groups for guiding URIs design and construction based on best practice and building URI registries.
Skills capability	• The semantic web and Linked Data technologies are maturing and freely available.

[12] http://5stardata.info/

	• Learn from, and work with the Linked Data and open data communities.

Table 3: Linked Data Risks

Linked Data techniques allow organisations to publish more,
reuse more and combine more data
for a fraction of the cost of older methods.

Bernadette Hyland, CEO, 3 Round Stones

OTHER PROJECTS

This section describes two projects (Born Digital and LAWA) that a Linked Data approach could underpin.

1: Born-Digital Project

The Born Digital project [21] has identified Councils ability of accessing born digital information now, and into the future, is at risk. This is due to the original created digital information quickly becoming fragile and unusable over time. The answer to this problem is for councils to curate, archive and preserve data, but this is also problematic. The main issues identified for preserving data by local authorities are:

- Some information is lost or nearly gone.
- Councils keep everything, but can't find what they need.
- Not all information is coming across when data is moved. Information is lost.

Consequently, the project team identified a roadmap for councils to implement to aid the preservation of data for ensuring their data assets now can be used in the future.

- Discovery: identifying assets that need preservation.
- Strategy: standardising methods for current information creation and maintenance now and into the future.
- Development: ongoing management of digital information.
- Monitoring: quality assurance processes for ensuring digital continuity.
- Governance: management of the data preservation processes.

Read more: http://www.algim.org.nz/globalassets/whitepapers/algim-born-digital-whitepaper-final.pdf

2: LAWA

Land, Air, Water Aotearoa (LAWA) is a collaboration of organisations with a common aim: to tell the story of our environment.

Initially a partnership between New Zealand's 16 regional and unitary councils, LAWA has grown to include the Cawthron Institute, Ministry for the Environment and Massey University with support from the private Tindall Foundation.

Established to help local communities find the balance between using natural resources and maintaining their quality and availability, LAWA allows data collected at individual councils to be shared and compared.

"LAWA is the first time the people can learn about the state of New Zealand's rivers and beaches in one place."

Read more: https://www.ict.govt.nz/guidance-and-resources/case-studies/open-data/lawa/

FUTURE

This section describes information trends for organisations to work towards the future and also where the Web is progressing to. Linked Data will aid and assist organisations to achieve these goals.

Data Management Trends

Gartner released a series of reports, 'Predicts 2015', where each report identified a specific area and key findings for how organisations will deploy information assets and what these information sources will consist of. Table 4 describes these areas and key findings.

Category	Prediction
Information Innovation [15]	• By 2020, 30% of data will be prescribed provenance, business, security and value metadata at the time of its creation. • By 2017, 15% of global organisations will use Linked Data methods to create open data applications to generate alternative revenue streams.
Enterprise Content	By year-end 2017, more than 50% of the world's business websites will be based on open-source Web content management (WCM) software.
Government & Local Government [16]	• By 2018, more than 30% of local government agencies will depend on data supplied by the Internet of Things (IoT) to support at least 50% of their mission-critical programs. • By 2018, over 30% of digital government projects will treat any data as open data.
Integration [17]	• By 2018, the number of new connections for IoT devices will exceed all other new connections for interoperability and integration combined. • Through 2018, integration will be required for more than 70% of public APIs.

Table 4: Gartner's Information Predictions

Over the next decade, Gartner found an expectation that central and local governments will increase their efforts to leverage exogenous data sources and will be more innovative around methods for capturing it [16].

Web Trends

The Web is gaining more momentum, particularly with the development of the Open Web Platform (OWP). The OWP is a set of technologies for developing distributed applications that delivers the most interoperability. The OWP focuses on the developers needs through W3C activities by enabling them to develop powerful applications that reach most people, on any device. Consequently, the W3C propose application foundations that identifies services and capabilities that should be available to all applications. One of the eight application foundations proposed is the *Service* foundation that includes payments, annotations, Social Web and Web of Data. The Web of Data, (i.e. the semantic web and Linked Data) is a future inclusion for developers to be able to have further access to the data and vocabularies [24].

Linked Data in Local Government

When local government consistently embraces Linked Data, we can envisage a future in which

- Real-time information can be found about a subject, such as a property, in a single query, without having to discover and visit each agency's website;
- National applications are developed to access local information and local services;
- Councils no longer rely on local copies of other agency's information;
- Communities debate the future of their locality using reliable evidence and analysis.

CONCLUSION

This white paper has described how Linked Data can deliver significant benefits not only for the local government sector, but across all sectors. For an organisation to expose and federate data using the Linked Data approach will realise the full benefits and potential of Web 3.0.

Using the Linked Data approach instead of traditional methods will break down data silos, provide new ways of linking information that is accessible for communities to engage in council services. Councils that adopt the Linked Data approach will establish a future-proofed data asset that is standards compliant and flexible to grow. Consequently, building a global information economy that is more open and transparent.

CALL TO ACTION

We call on local authorities to:

- Adopt the Linked Data approach;
- Supply cemetery schema information to build a national / international vocabulary for describing cemetery data.

We call on national governments to:

- Adopt the Linked Data approach;
- Implement guidance for URI design that follows best practice;
- Publish high quality national reference data to link to, that also encourages local authorities to publish to similar standards;
- Set up national registers for datasets and vocabularies;
- Engage and encourage with local authorities to open data;
- Work with local authorities to develop a high level concept model for local government data as a useful template from which to build Linked Data solutions to maximise the opportunities to make useful joins.

We call on national and international communities to:

- Contribute best practices for building a local government Linked Data toolkit.

The Semantic Web is not a separate Web, but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. Tim Berners-Lee

---

ALGIM WORKING PARTY

ALGIM would like to thank the working party made up of:

- Adrienne Bonnington – Horizons Regional Council,
- Jacqueline Davidson – Auckland Council,
- Anthony Woodside – Rangitikei District Council,
- Ian Tate – Horowhenua District Council,
- Paul Davidson – Sedgemoor District Council,
- Byron Cochrane – Land Information New Zealand (LINZ)
- Amanda Cockburn for their assistance in the writing of this White Paper.

The Association of Local Government Information Management (ALGIM) represents the national and international interests of the information, communication and technology (ICT) sector within New Zealand's city, district and regional councils.

ALGIM provides best practice in the local government ICT sector by enhancing professional development through scholarships, training, events, awards and networking, and offer leadership through toolkits, advocacy, research and shared services.

ALGIM are facilitating a Linked Data project with the main goal to develop and promote the extensive use of Linked Data within New Zealand Councils. To achieve this goal, ALGIM has formed a Local Government Linked Data Working Group in collaboration with ALGIM's international sister organisations. One of the first tasks is to develop a Linked Data Toolkit.

FOR MORE INFORMATION CONTACT ALGIM
Email: info@algim.org.nz
Phone: 06 351 6330
Web: www.algim.org.nz

REFERENCES

[1] Open Government Partnership (2011). *What is the Open Government Partnership?* Retrieved 2015-08-11 from: http://www.opengovpartnership.org/about

[2] World Wide Web Foundation (2015). *Open Data Barometer.* Retrieved 2015-08-11 from: http://www.opendatabarometer.org/report/summary/the_barometer.html

[3] Department of Communities and Local Government (2013). *The Second Department of Communities and Local Government Open Data Strategy.* Retrieved 2015-08-11 from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254495/131031_2013_DCLG_Open_Data_Strategy.pdf

[4] Bizer, C., Cyganiak,, R. & Heath, T. (2007). *Introduction to Linked Data.* Retrieved 2015-07-02 from: http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/#intro

[5] Berners-Lee, T. (2002). *What to HTTP URIs Identify?* Retrieved 2015-07-03 from: http://www.w3.org/DesignIssues/HTTP-URI.html

[6] LinkedDataTools (2015). *Tutorial 1: Introducing Graph.* Retrieved 2015-07-03 from: http://www.linkeddatatools.com/introducing-rdf]

[7] D2RQ Software. *D2RQ: Accessing Relational Databases as Virtual RDF Graphs.* Retrieved 2015-07-03 from: http://d2rq.org/

[8] World Wide Web Consortium (2014). *Resource Description Framework.* Retrieved 2015-08-03 from: www.w3c.org/rdf

[9] Hausenblas, M. (2011). *Oh – it is data on the Web*. Retrieved 2015-08-15 from: https://webofdata.wordpress.com/2010/04/14/oh-it-is-data-on-the-web/

[10] Hausenblas, M, Kerrin, M, Pizzo, M, Viega, E. & Wilson, N (2012). *Linking Structured Data: Using OData to Access an Triple Store through a semi-structured conceptual model*. White paper. Retrieved 2015-08-12 from: http://research.microsoft.com/pubs/193076/Whitepaper%20on%20Linking%20Structured%20Data.pdf

[11] De Keyzer, M., Loutas, N., & Goedetier, S. (2014). *TM2.1. The Linked Open Government Data & Metadata Lifecycle.* European Commission. Retrieved 2015-08-07 from, https://joinup.ec.europa.eu/community/ods/document/tm21-linked-open-government-data-metadata-lifecycle-en

[12] Villazón-Terrazas, B & Corcho (2011). *Methodological Guidelines for Publishing Linked Data.* Retrieved 2015-08-10 from, http://delicias.dia.fi.upm.es/wiki/images/7/7a/07_MGLD.pdf

[13] Hausenblas, M. (2011). *Linked Data life cycles*. Retrieved 2015-08-20 from: http://www.slideshare.net/mediasemanticweb/linked-data-life-cycles

[14] W3C (2014). *Best Practices for Publishing Linked Data.* Retrieved 2015-08-05 from: http://www.w3.org/TR/ld-bp/

[15] Laney, D., Natis, Y.V., Dayley, A., Bugaiski, J., Scheibenreif, D., Newmand, D., & Schlegel, K. (2014). *Predicts 2015: The Intersection of Information Innovation and Business Digitalization.* Gartner Report. Retrieved 2015-08-19 from: https://www.gartner.com/doc/2928317?ref=ddisp

[16] Howard, R. Archer, G., Cannon, N., & Vining, J. (2014). *Predicts 2015: Government Adapts to the Digital Era.* Gartner Report. Retrieved 2015-08-19 from: https://www.gartner.com/doc/2924021?ref=ddisp

[17] Lheureux, B.J., Thoo, E., Guttridge, K., Thompson, J., Pezzini, M., & Schulman, J. (2014*). Predicts 2015: Digital Business and Internet of Things Add Formidable Integration Challenges.* Gartner Report. Retrieved 2015-08-19 from: https://www.gartner.com/doc/2906917?ref=ddisp

[18] Open Data Institute and Thomson Reuters (2014). *Creating Value with Identifiers in an Open Data World*. Retrieved 2015-08-19 from http://innovation.thomsonreuters.com/content/dam/openweb/documents/pdf/corporate/Reports/creating-value-with-identifiers-in-an-open-data-world.pdf

[19] Schmid, G., (2015). *Implementing an Innovative Data Strategy To Improve Local Service Delivery.* Retrieved 2015-08-20 from: http://www.slideshare.net/GescheSchmid/inside-government-2015-0319-v4

[20] Tim Davis (2011). *The Elements of the Linked Open Data Stack.* Retrieved 2015-07-01 from: http://www.timdavies.org.uk/wp-content/uploads/Diagrams-Elements-of-the-Linked-Open-Data-Puzzle-Draft-3.png

[21] Dewson, N., McIntosh, A., Davidson, J., Cockburn, A. (2015). *Management of Born Digital Records.* ALGIM White Paper. Retrieved 2015-08-19 from: http://www.algim.org.nz/globalassets/whitepapers/algim-born-digital-whitepaper-final.pdf

[22] World Wide Web Consortium (2012). *OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax (*2nd Ed.) Retrieved 2015-08-25 from: http://www.w3.org/TR/owl2-syntax/

[23] Stapleton, G., Howse, J., Bonnington, A. & Burton, J. (2014). *A Vision for Diagrammatic Ontology Engineering.* Paper presented at: VISUAL 2014: Visualizations and User Interfaces for Knowledge Engineering and Linked Data Analytics, Sweden. Retrieved 2015-08-25 from: http://ceur-ws.org/Vol-1299/

[24] World Wide Web Consortium (2015). *Application Foundations for the Open Web Platform.* Retrieved 2015-08-26 from: http://www.w3.org/blog/2014/10/application-foundations-for-the-open-web-platform/

Other Links

Hyland, B. (2011) *Taking Inspiration from Open Government Initiatives.* Retrieved 2015-08-12 from, http://www.dataversity.net/hyland-interview-title-tk/

APPENDIX

Appendix A: Glossary of Terms

Internet of Things (IoT)

The Internet of Things (IoT) is a scenario in which objects, animals or people are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.

Linked Data

Linked Data describes a method of publishing structured data so that it can be interlinked and become more useful through semantic queries. In the context of this white paper, Linked Data can also refer to Linked Open Data.

Linked Open Data

Linked Data that is publically accessible with an appropriate data license.

Ontologies

Ontologies are formalised vocabularies of terms, often covering a specific domain and shared by a community of users. They specify the definitions of terms by describing their relationships with other terms in the ontology. There is no clear division between what is referred to as "vocabularies" and "ontologies". The trend is to use the word "ontology" for more complex, and possibly quite formal collection of terms, whereas "vocabulary" is used when such strict formalism is not necessarily used or only in a very loose sense (W3C).

Open Data

For data to be considered truly open, it must be published in bulk, machine-readable formats, and under an open license.

Open Source Software

Open-source software (OSS) is computer software with its source code made available with a license in which the copyright holder provides the rights to study, change, and distribute the software to anyone and for any purpose.

Resource Description Framework (RDF)

RDF is a standard model for data interchange on the Web.

Semantic web

The Semantic Web is an extension of the Web through standards by the World Wide Web Consortium (W3C). The standards promote common data formats and exchange protocols on the Web, most fundamentally the Resource Description Framework (RDF).

Uniform Resource Indicator (URI)

A string of characters used to identify a name or resource on the Web.

Uniform Resource Location (URL)

A type of URI that refers to objects on the World Wide Web.

Uniform Resource Name (URN)

A unique name for a resource, does not provide a location.

Vocabularies

Vocabularies define the concepts and relationships (also referred to as "terms") used to describe and represent an area of concern. There is no clear division between what is referred to as "vocabularies" and "ontologies". The trend is to use the word "ontology" for more complex, and possibly quite formal collection of terms, whereas "vocabulary" is used when such strict formalism is not necessarily used or only in a very loose sense (W3C).

Web Ontology Language (OWL)

OWL is a language used to for vocabularies/ontologies.

World Wide Web (WWW)

The World Wide Web (www, W3) is an information space where documents and other web resources are identified by URIs, interlinked by hypertext links, and can be accessed via the Internet.

World Wide Web Consortium (W3C)

The World Wide Web Consortium (W3C) is an international community where Member organisations, a full-time staff, and the public work together to develop Web standards.

5-Star Open Data Model

The World Wide Web Consortium (W3C) has developed a five star model[13] to describe different characteristics of open data, and its usefulness for people wishing to reuse it.

[13] https://www.ict.govt.nz/guidance-and-resources/open-government/toolkit-agencies/applying-5-star-open-data-model-your-high-value-pu/

Appendix B: Vocabulary Modelling

Vocabularies and ontologies give data semantics and contain the following entities as defined by the W3C [22]:

- ***Classes***: sets of individuals
- ***Properties***:
  - Object properties: connect pairs of individuals
  - Data Properties: connect individual pairs with literals
- ***Individuals***: represent actual objects from a domain.

The University of Brighton (UK) has developed an approach for visualising these essential entities using the Euler method. Figure 5 shows an extract from a cemetery ontology that has applied these concepts [23].

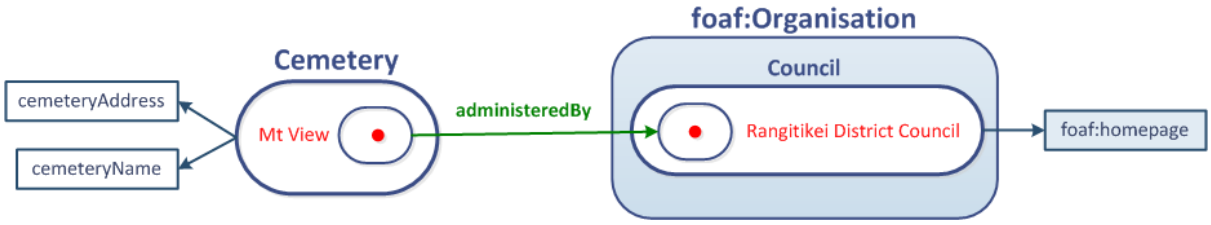

Figure 5 : A Concept Diagram [23]

Figure 5 identifies classes using the closed curve notation, these are Cemetery, Council and Organisation. The placement of Council inside Organisation expresses that Council is subsumed by Organisation. Organisation has adopted the term that pre-exists in the well-known vocabulary FOAF (Friend of a Friend) vocabulary. Since Cemetery and Organisation do not overlap, the diagram expresses that these two concepts are disjoint (have no individuals in common). The arrow labelled administeredBy, sourced on Cemetery and targeting the labelled curve inside Council asserts that Mt View is *administeredBy* by Rangitikei District Council. There are circumstances where there are unlabelled curves within the classes that represent an anonymous concept where individuals cannot be precisely identified (e.g. Person may have too many individuals to identify).

The arrows extending from the classes to the rectangular boxes represent the data properties for the individuals. The Cemetery class shown in Figure 5 contains two data properties, cemeteryAddress and cemeteryName. These represent the field names that relate to the data within the database table. Consequently, the data property foaf:homepage for Rangitikei District Council links to the Council's cemetery homepage. Once again, the homepage data property has been adopted from FOAF.

Appendix C: Government Linked Data Resources by Country

Table 5 identifies the resources recommended for licensing open data and URI Design guidelines available by country.

Country	URI Design Guidelines	Data Licensing
Australia	URI Guidelines for publishing Linked Datasets on data.gov.au v0.1 https://github.com/AGLDWG/TR/wiki/URI-Guidelines-for-publishing-linked-datasets-on-data.gov.au-v0.1	AusGoal http://www.ausgoal.gov.au/
New Zealand	*Nil*	NZ GOAL Framework 2.0 https://www.ict.govt.nz/guidance-and-resources/open-government/new-zealand-government-open-access-and-licensing-nzgoal-framework/
United Kingdom	Design URI Sets for the UK Public Sector (Team led by Paul Davidson, 2009) https://www.gov.uk/government/publications/designing-uri-sets-for-the-uk-public-sector	Open Government License http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

Table 5: Linked Data Resources by Country

The practices of other stakeholders (including EU Agencies and Services, EU Member States, Standardisation bodies and initiatives, and others) are summarised in a study for the European Commission on Persistent URIs conducted in 2012.[14]

[14] https://joinup.ec.europa.eu/sites/default/files/D7.1.3%20-%20Study%20on%20persistent%20URIs_0.pdf