SciDataCon 2016

Conference paper

Automatically Calculating the Adherence to License Requirements

Summary

Through the experiences of building several information infrastructures in Australia we have come to wanting to calculate the properties of data licenses in as automated a fashion as possible. To do this, we have trialed decomposing licenses into individual requirements, building on the Creative Commons Rights Expression Language (CC REL) information model[1]. Such a decomposition, along with the separation of rights management from licenses as per the Open Data Rights Statement vocabulary (ODRS) model[2], allows us to model very many different licenses as simple collections of common requirements. In **Car & Stenson (2015)** we posited requirement resolution actions that systems or people can undertake that satisfy individual requirements and show how systems can be made to automatically satisfy certain classes of requirements but certainly not all.

We are not currently able to automate every aspect of license requirement adherence and we believe that full automation is impossible, however every bit of automation improves the efficiency of data delivery.

In this presentation we report on some aspects of our license model, including the modelling of requirements and their resolutions. We also show how such modelling can enhances data access by comparing the *status quo* and possible future delivery of data via a large multi-agency, Australian data generation project; the Bioregional Assessments Programme.

Background

The Bioregional Assessments programme[3] collected several thousand datasets from several dozen government agencies, community groups and private companies. Much effort was devoted to catering for all their licensing needs which resulted in 66 distinct licenses[4] which, while ably expressing very many particular licensing needs of particular organisations, proves difficult to deal with in aggregator systems, such as data stores, which benefit from licensing consistency. While it was not possible to reduce all of the distinct licenses down to a few standard licenses, it is possible to model the individual requirements that the licenses impose on data users and this results in a just a small group of about 7 distinct requirements which can be more easily dealt with. Some example requirements from the Bioregional Assessments programme are given at http://data.bioregionalassessments.gov.au/id/requirement/.

An enabling pre-step to this decomposition of licenses into individual requirements was the separation of requirements relating to copyright attribution from the license. This is the approach promoted by the ODRS vocabulary's license & rights model[2].

Our license model showing this ODRS license/rights separation, decomposition of licenses into CC REL-inspired requirements[1] objects and license-to-license relations via provenance ontology relationships[5] is show in Figure 1.

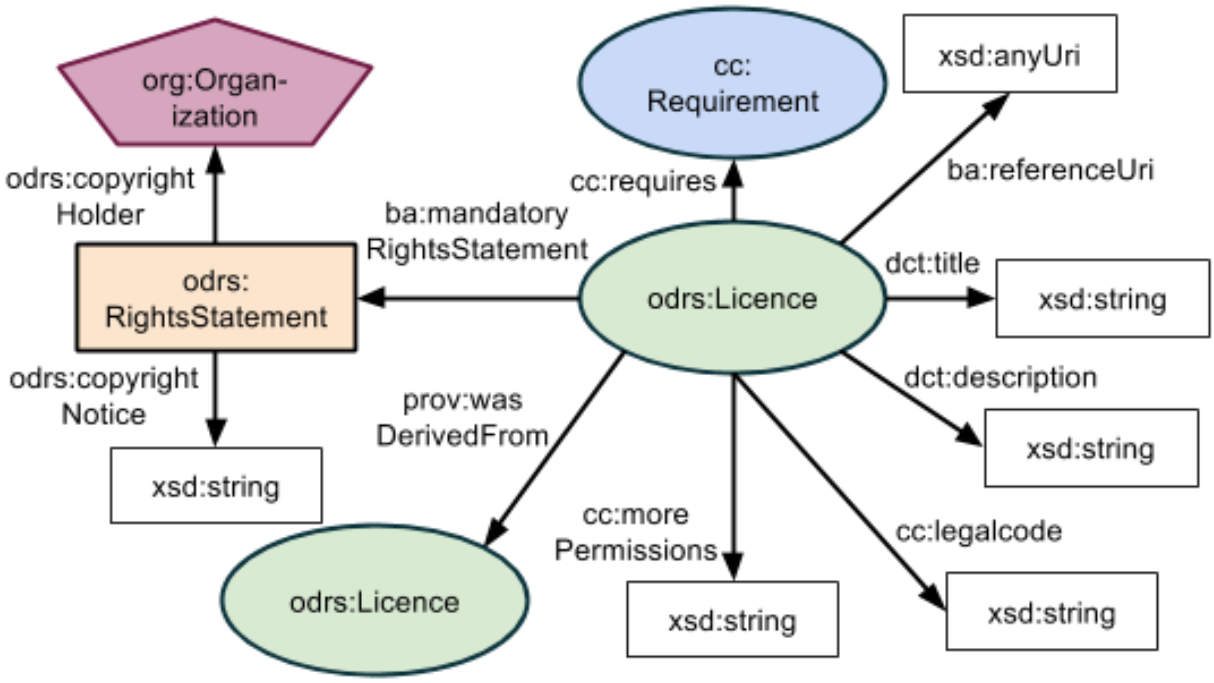


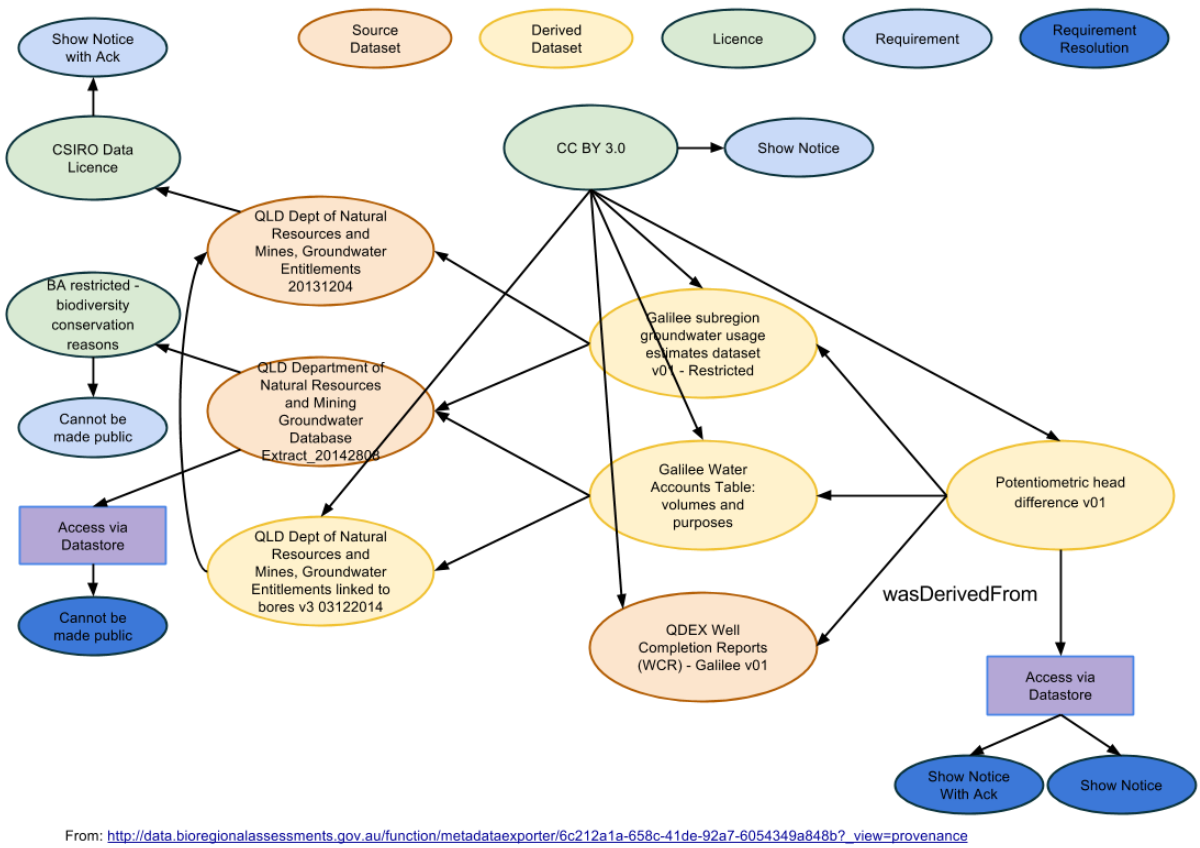
Figure 1: The license information model as used in **Car & Stenson (2015)** presented as an OWL[6] diagram. The namespace prefixes shown are referenced in Table 1.

Table 1: Namespace prefixes

Prefix	Namespace	Scheme
ba	http://data.bioregionalassessments.gov.au/def/ba#	The Bioregional Assessments Ontology[9]
cc	http://creativecommons.org/ns#	Creative Commons Rights Expression Language[1]
dct	http://purl.org/dc/terms/	Dublin Core Terms[10]
odrs	http://schema.theodi.org/odrs#	Open Data Rights Statement vocabulary[2]
org	http://www.w3.org/ns/org#	The Organization Ontology[8]
prov	http://www.w3.org/ns/prov#	PROV-O: The PROV Ontology[5]
xsd	http://www.w3.org/2001/XMLSchema#	XML Schema Part 2: Datatypes Second Edition[7]

Calculating Requirements Resolutions

In order to automatically calculate the resolution of licenses' requirements, we model requirement resolution objects which are the result of actions undertaken by people or systems. Figure 2 shows a data derivation hierarchy from the Bioregional Assessments programme that, through license conditions that encumber descendent datasets with certain conditions. Some of the license requirements are shown and the corresponding requirement resolution objects and their generating activities too. The generation of requirement information for Figure 2 can be undertaken by license decomposition which is mostly a manual task. The generation of requirement resolution information can undertaken by repository programming against the requirement information and the dataset-dataset relationships recorded when datasets are registered. Once all the information is collected, it can be automatically traversed in order to ensure that each requirement is met, at the appropriate juncture, by a requirement resolution.



From: http://data.bioregionalassessments.gov.au/function/metadataexporter/6c212a1a-658c-41de-92a7-6054349a848b?_view=provenance

Figure 2: a derived data scenario from the Bioregional Assessments programme showing some licenses, requirements and requirement resolution objects. The relationships between dataset objects are `prov:wasDerivedFrom`, those between datasets and licenses, `dct:license`, those between licenses and requirements, `cc:requires`, those between datasets and activities (rectangles) `prov:used` and activities and requirements resolution `prov:generated`.

Generalization beyond licenses

In a related SciDataCon2016 paper, *Agreeing about agreements: modelling social contracts, people and data* by Car & Box, an ontology for modelling agreements is given. Licenses are one form of agreement, as are standards, MoUs and so on. The decomposition of licenses into requirements is as applicable to these other forms of agreements as much as licenses.

Conclusions

This decomposition methodology goes some way to enabling the clear and ready determinability of use conditions datasets. It also allows for some automation of usage assessments within smart repositories able to respond to individual requirements. It also addresses accumulated conditions for use in some circumstances.

More work is required to codify general methodologies for license decomposition into requirements and for describing likely methods of generating requirement registration objects.

Competing Interests

The authors declare that they have no competing interests.

Notes

1. http://creativecommons.org/ns
2. http://schema.theodi.org/odrs/
3. http://www.bioregionalassessments.gov.au
4. http://data.bioregionalassessments.gov.au/id/licence/
5. https://www.w3.org/TR/prov-o/
6. https://www.w3.org/TR/owl2-syntax/
7. https://www.w3.org/TR/xmlschema-2/
8. https://www.w3.org/TR/vocab-org/
9. http://data.bioregionalassessments.gov.au/def/ba
10. http://dublincore.org/schemas/rdfs/

References

Car, N.J. and **Stenson, M.P.** 2015. Complex licence requirements for the Bioregional Assessments Programme managed by provenance. Proceedings of MODSIM2015, 21st International Congress on Modelling and Simulation. MSSANZ, December 2015, pp. 613–619. ISBN: 978-0-9872143-5-5. http://mssanz.org.au/modsim2015/C4/car.pdf.