



# About me, provenance and adoption of PROV

Nicholas J. Car

[https://en.wikipedia.org/wiki/Provenance#Computer\\_science](https://en.wikipedia.org/wiki/Provenance#Computer_science)



## Computer science [ [edit](#) ]

Within [computer science](#), [informatics](#) uses the term 'provenance' to mean the [lineage of data](#), as per data provenance, with research in the last decade extending the conceptual model of causality and relation to include processes that act on data and agents that are responsible for those processes. See, for example, the proceedings of the International Provenance Annotation Workshop ([IPAW](#)) and Theory and Practice of Provenance ([TaPP](#)). [Semantic web](#) standards bodies, such as the [World Wide Web Consortium](#), have recently (2014) ratified a standard data model for provenance representation known as [PROV](#) which draws from many of the better known provenance representation systems that preceded it, such as the [Proof Markup Language](#) and the [Open Provenance Model](#).

Inter-operability is a design goal of most recent computer science provenance theories and models, for example the Open Provenance Model's (OPM) generation workshop aimed at "establishing inter-operability of systems" through information exchange agreements.<sup>[26]</sup> Data models and serialisation formats for delivering provenance information typically reuse existing metadata models where possible to enable this. Both the OPM Vocabulary<sup>[27]</sup> and the PROV Ontology<sup>[28]</sup> make extensive use of metadata models such as [Dublin Core](#) and [Semantic Web](#) technologies such as the [Web Ontology Language](#) (OWL).

## About Nick

- Electrical engineering degree, became CompSci/IT
- 10 years on-and-off at CSIRO, agricultural Informatics
- Several years as an IT consultant
  - Built normal and specialised IT systems, inc. metadata sys
- Last 5 years in provenance informatics at CSIRO
- Now at GA
  - Data Architect, leading prov & data sys implementation
- Chairing/leading a series of groups
  - RDA prov, Aust. Prov, Aust Vocab SIG, LD WG Sol.

# The PROV provenance model

- The W3C PROV provenance model was developed by the authors of several pre-existing provenance models
- A large W3C initiative
- A good DM technically
  - Useful for modelling techniques, not just provenance content
- Overview: <https://www.w3.org/TR/prov-overview/>

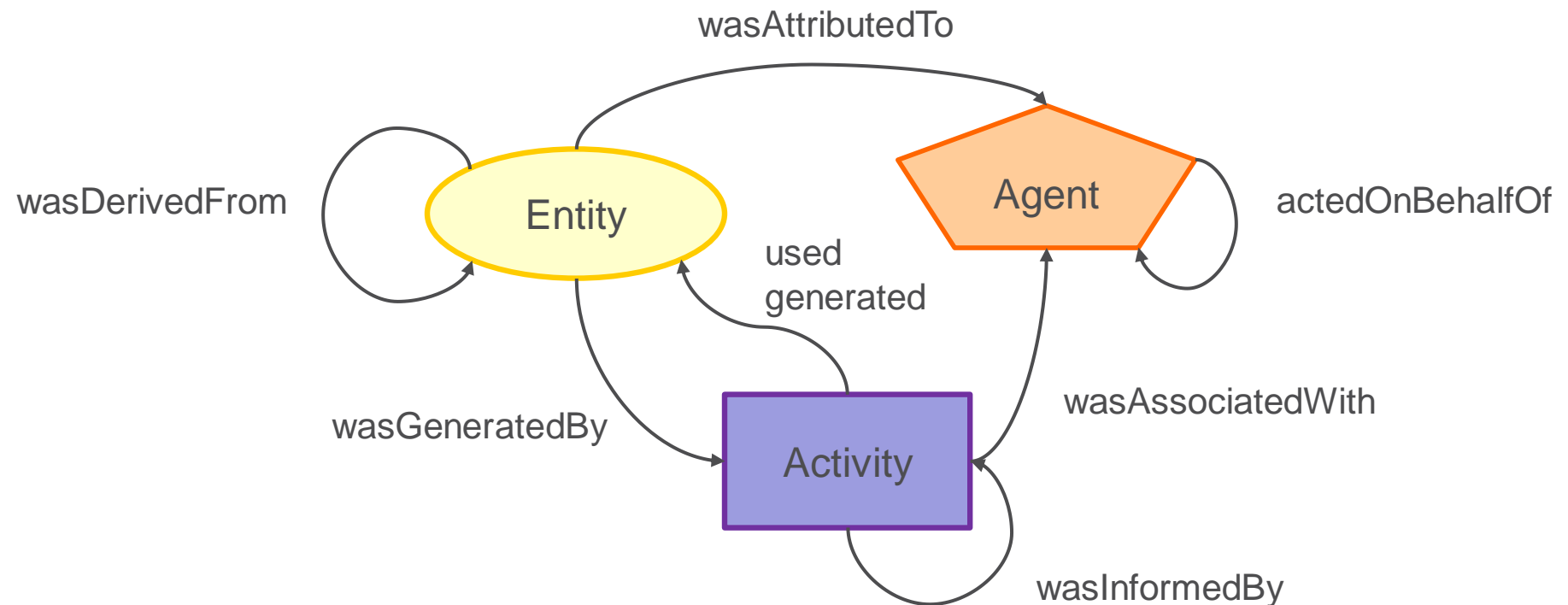
# The PROV provenance model

## The PROV family of documents

- [PROV-OVERVIEW](#) (Note), an overview of the PROV family of documents (this document);
- [PROV-PRIMER](#) (Note), a primer for the PROV data model [[PROV-PRIMER](#)];
- [PROV-O](#) (Recommendation), the PROV ontology, an OWL2 ontology allowing the mapping of the PROV data model to other ontologies [[PROV-O](#)];
- [PROV-DM](#) (Recommendation), the PROV data model for provenance [[PROV-DM](#)];
- [PROV-N](#) (Recommendation), a notation for provenance aimed at human consumption [[PROV-N](#)];
- [PROV-CONSTRAINTS](#) (Recommendation), a set of constraints applying to the PROV data model [[PROV-CONSTRAINTS](#)];
- [PROV-XML](#) (Note), an XML schema for the PROV data model [[PROV-XML](#)];
- [PROV-AQ](#) (Note), mechanisms for accessing and querying provenance [[PROV-AQ](#)];
- [PROV-DICTIONARY](#) (Note) introduces a specific type of collection, consisting of key-entity pairs [[PROV-DICTIONARY](#)];
- [PROV-DC](#) (Note) provides a mapping between PROV-O and Dublin Core Terms [[PROV-DC](#)];
- [PROV-SEM](#) (Note), a declarative specification in terms of first-order logic of the PROV data model [[PROV-SEM](#)];
- [PROV-LINKS](#) (Note) introduces a mechanism to link across bundles [[PROV-LINKS](#)].

# The PROV provenance model – PROV-O

- <https://www.w3.org/TR/prov-o/>



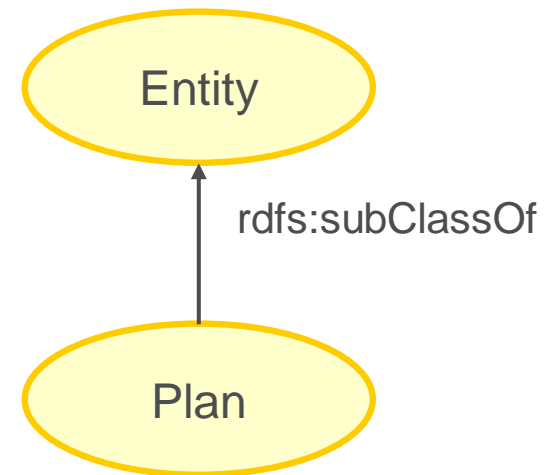
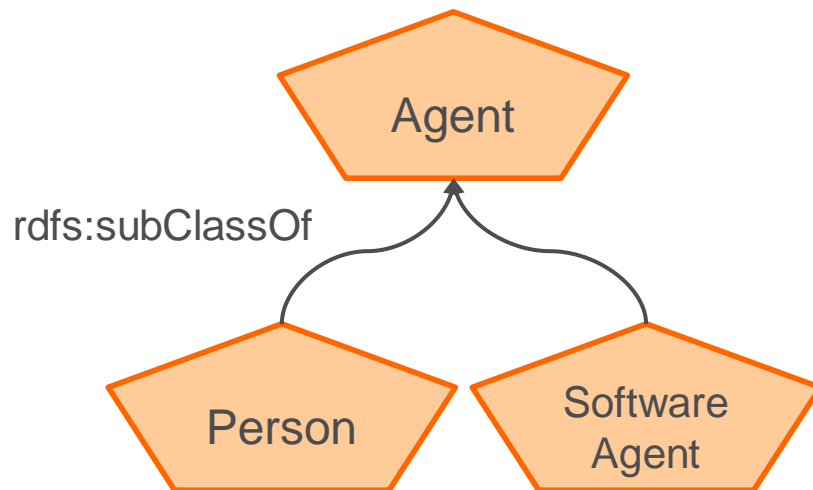
# The PROV provenance model – PROV-O

- PROV has specialised subclasses, allows for more



# The PROV provenance model – PROV-O

- PROV has specialised subclasses, allows for more



# The PROV provenance model – PROV-O

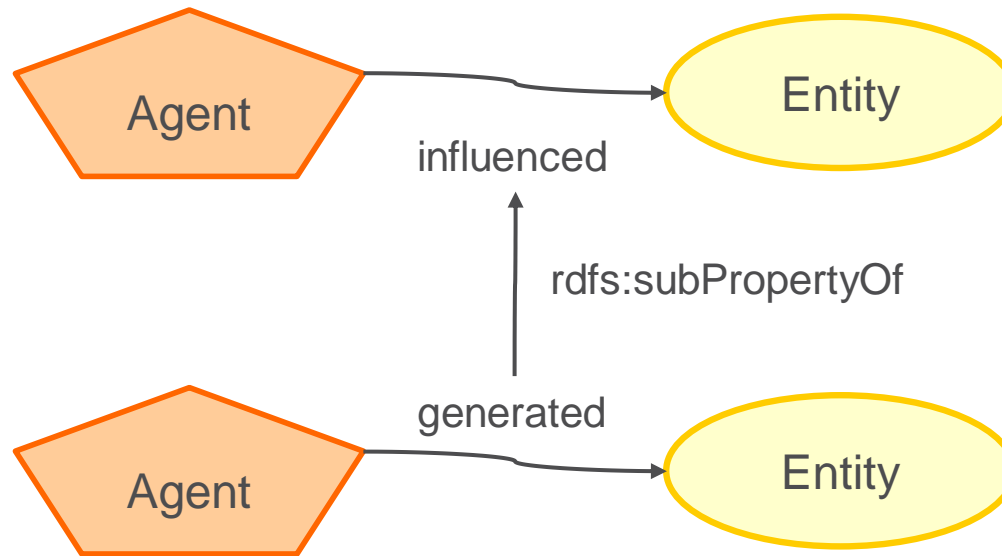
- PROV has specialised sub*properties*, allows for more





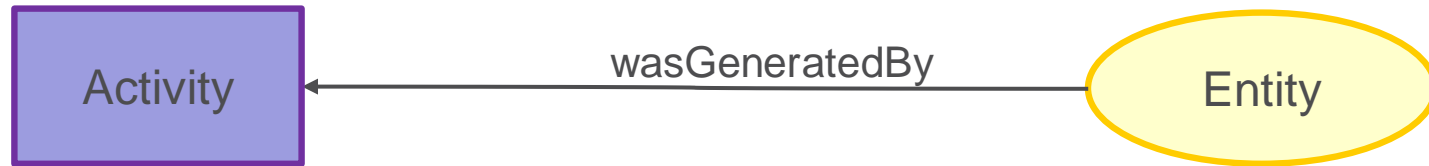
# The PROV provenance model – PROV-O

- PROV has specialised subproperties, allows for more



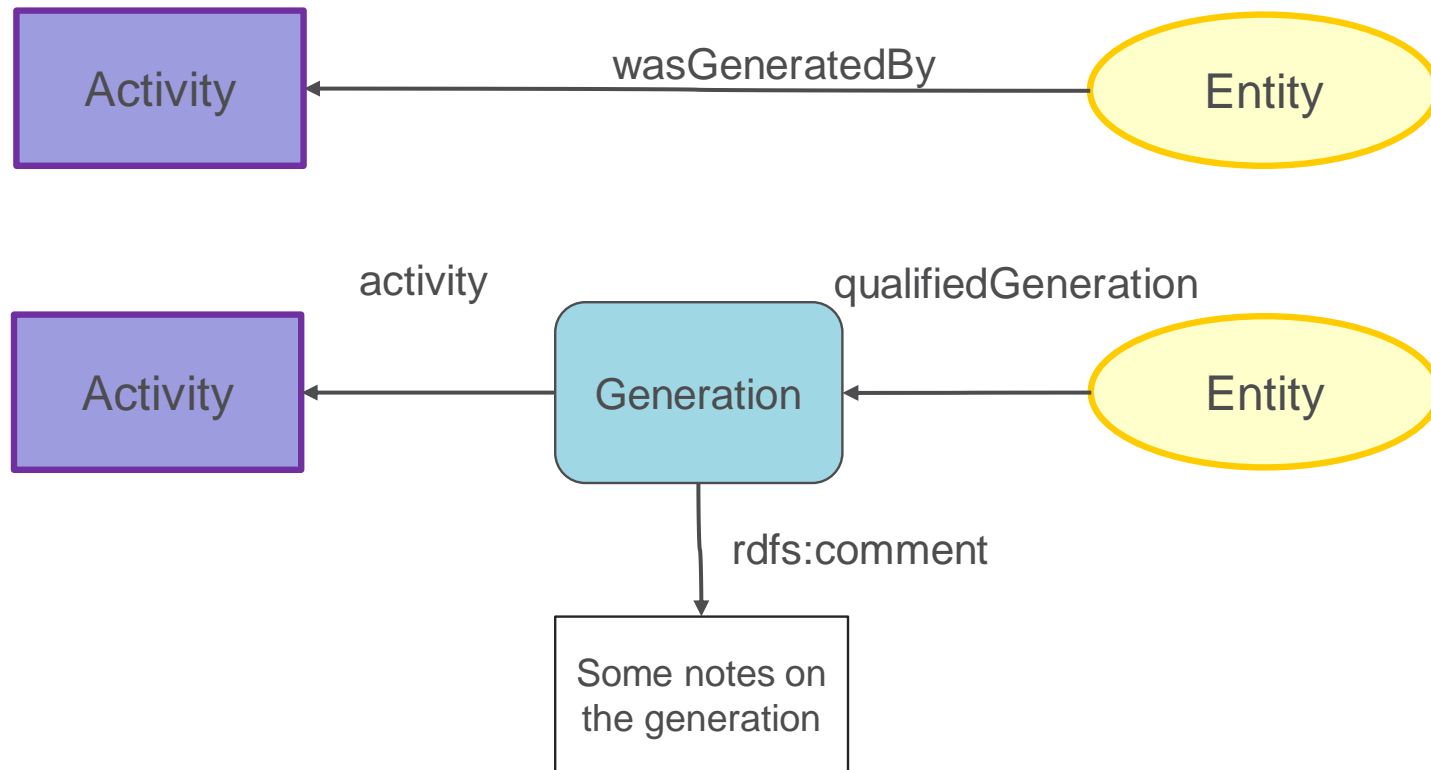
# The PROV provenance model – PROV-O

- PROV likes to qualify things



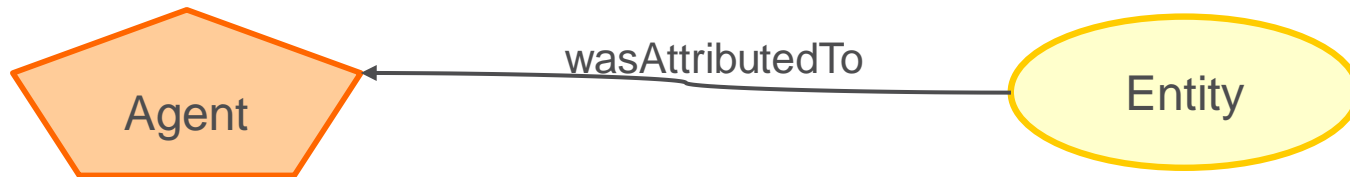
# The PROV provenance model – PROV-O

- PROV likes to qualify things



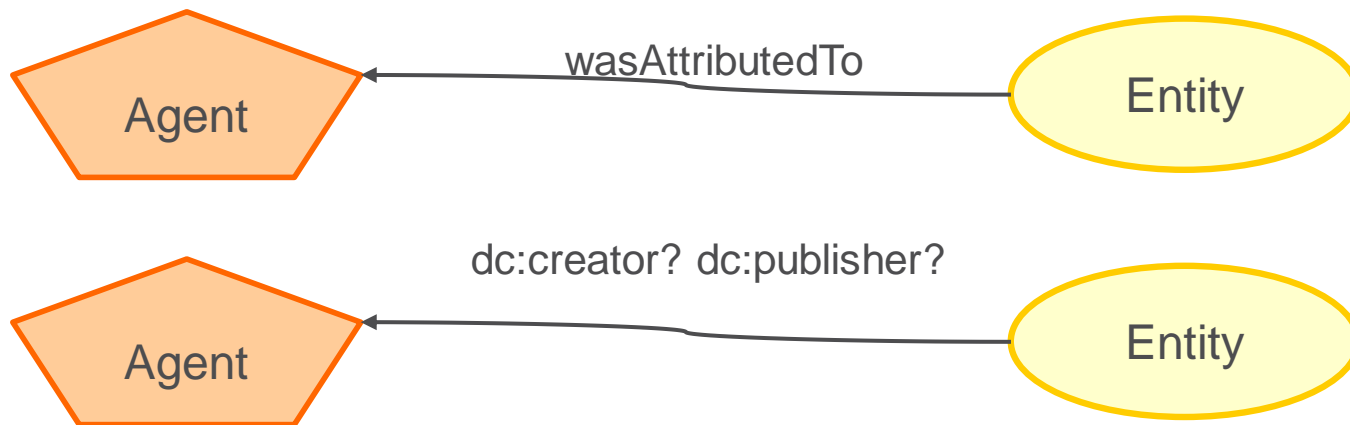
# The PROV provenance model – PROV-O

- PROV likes to qualify things



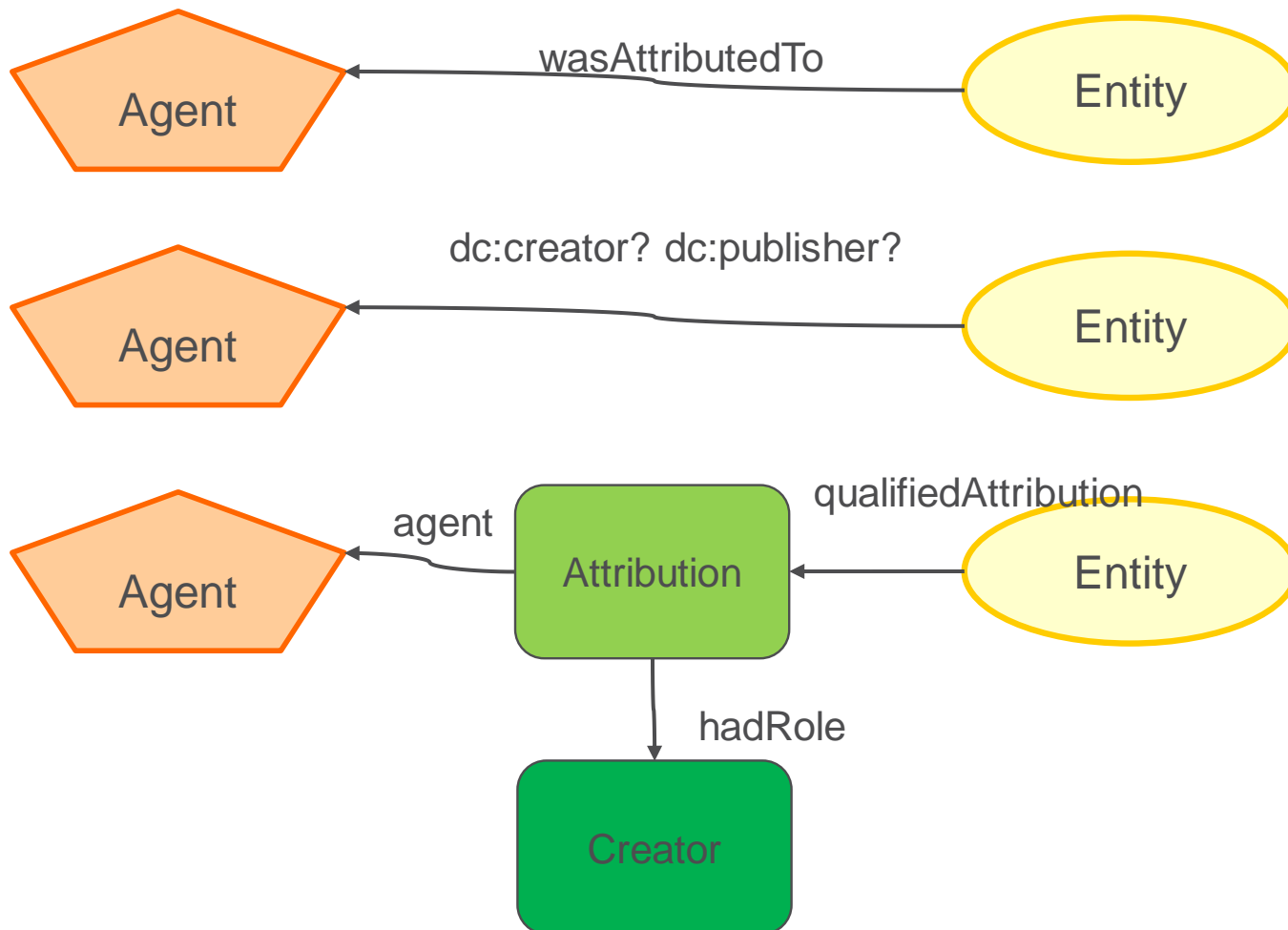
# The PROV provenance model – PROV-O

- PROV likes to qualify things



# The PROV provenance model – PROV-O

- PROV likes to qualify things



# The PROV provenance model

## My work on PROV

- Technical
  - Built a PROV-compliant provenance store, PROMS
  - Working on PROV-AQ theory mechanisms
  - Written several PROV-related ontologies, e.g. Agreements
- Institutional
  - PROV and provenance generally adoption in Orgs
  - PROV adoption in Aust, gov-wide
  - PROV IG/WG in the RDA

# PROV-O and other Os

- PROV-O requires:
  - RDF & OWL things: RDF, RDFS, OWL, XSD
- PROV-O commonly uses:
  - FOAF, DC
- PROV-O can use:
  - Anything!
- I use these with PROV-O:
  - SKOS, DCAT, PAV
- My extensions:
  - PROMS, AGR, LIC



# PROV and the Provenance Community

- Provenance researchers
  - No more single focus
  - Large projects like TRITIUM involve many
  - Many ontology researchers doing related things
    - Spatial data on the web WG
    - SHAPES WG
- PROV adoption underway in:
  - Aust gov
  - RDA
  - CODATA
  - Surely many others

# Pathways for Schema X & PROV

- Lite adoption
  - Schema X adopts some PROV classes & properties
    - Can profile PROV, Can emulate PROV
- Heavy adoption
  - Schema X wholly adopts PROV classes & properties
    - Full profiling
- Handover
  - Schema X morphs to use PROV as it's primary forms
- AND
  - Schema X joins with PROV
    - Allows PROV constructs in addition to own
    - Needs to define linkage points

# Pathways for Schema X & PROV

- Technical commensurability
  - How does Schema X relate *semantically* to PROV?
  - How does Schema X record PROV values or PROV links?
  - How does Schema X export its values + PROV?



Australian Government  
Geoscience Australia



# Thanks!

**Nicholas J. Car**

Data Architect, Geoscience Australia  
Canberra, Australia

[nicholas.car@ga.gov.au](mailto:nicholas.car@ga.gov.au) | [www.ga.gov.au](http://www.ga.gov.au)