



Manaaki Whenua
Landcare Research

Our Land, Our Future

Tō tātou whenua, mō āpōpō



Manaaki Whenua
Landcare Research

Automated Data Provenance Tracking for Ecosystem Services Assessments

Alexander Herzig, Anne-Gaelle E. Ausseil, Raphael
Spiekermann, Ben Jolly, Tom Burleigh, David Medyckyj-Scott



Content

- Why Data Provenance?
- Ecosystem Services Accounting & Data Provenance
- Implementation of Data Provenance for ES modelling
- Visualising Data Provenance
- Conclusions
- References



Why Data Provenance?

- Data-driven decision-making
- Flawed use of science-based knowledge
- Justification for decisions (spending)
- Demand for more transparent science

Data provenance tracks and documents information about entities, activities, and people involved in producing a piece of data (Belhajjame et al. 2013)

> authenticity, trust, transparency, reproducibility



ES Accounting & Data Provenance

- System of Environmental-Economic Accounting (SEEA)
Experimental Ecosystem Services Accounts (EEA)
 - Evidence-based decision making
 - Data-driven decision making
- Data provenance applications
 - Data suitability
 - Data quality
 - Audit trail
 - Replication recipes
 - Informational



What is Data Provenance?

Lineage – steps taken to produce the data (Lanter 1991)

“record that describes entities and processes involved in producing and delivering or otherwise influencing that resource”

(W3C 2010)

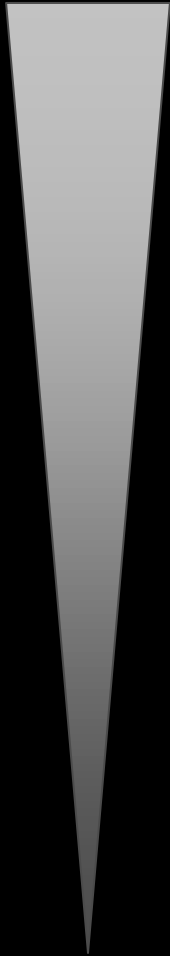
W7 model of data provenance (Ram & Liu 2009):

$p(D) = \{\text{what, when, where, who, how, which, why}\}$

Granularity of Data Provenance



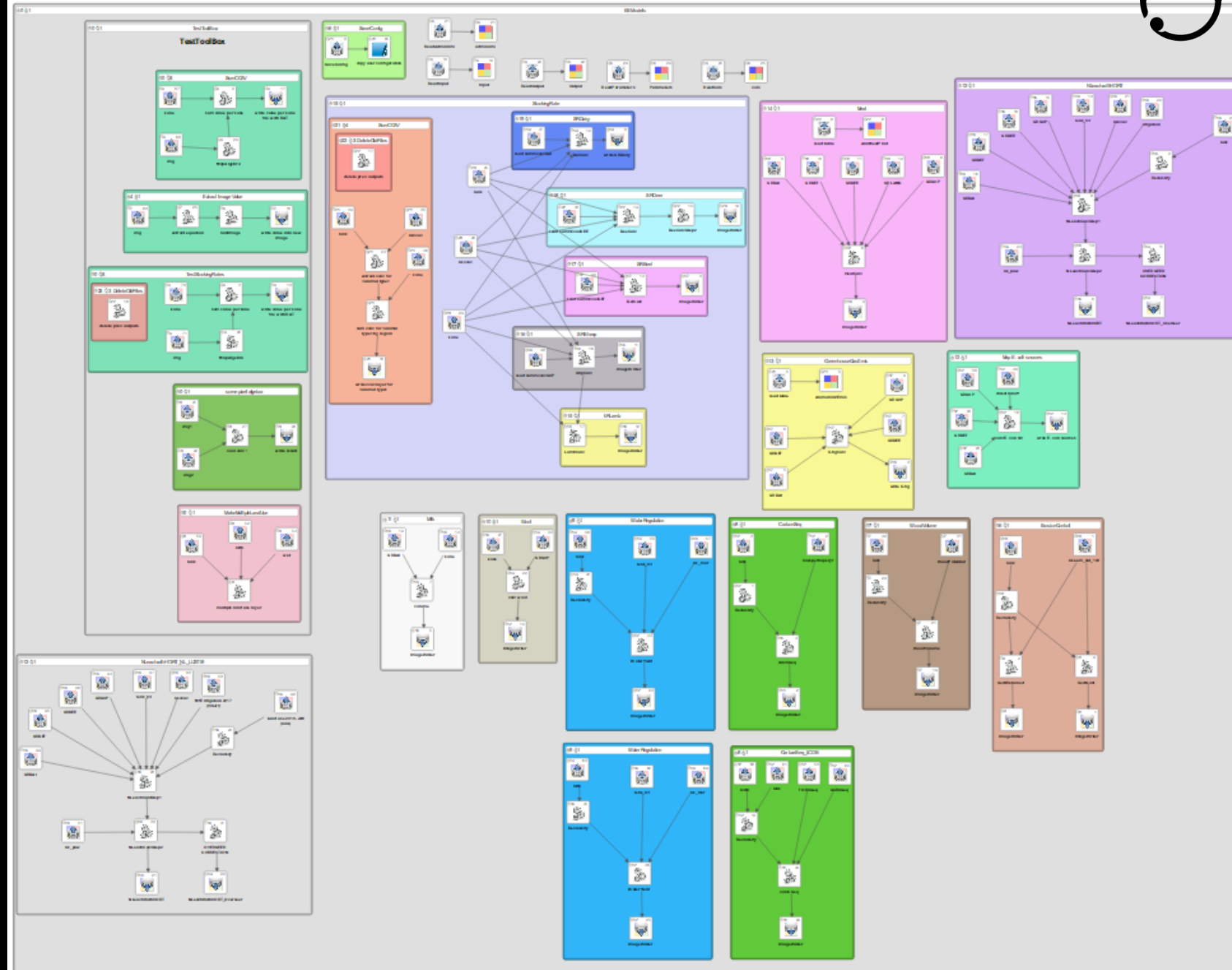
Transparency



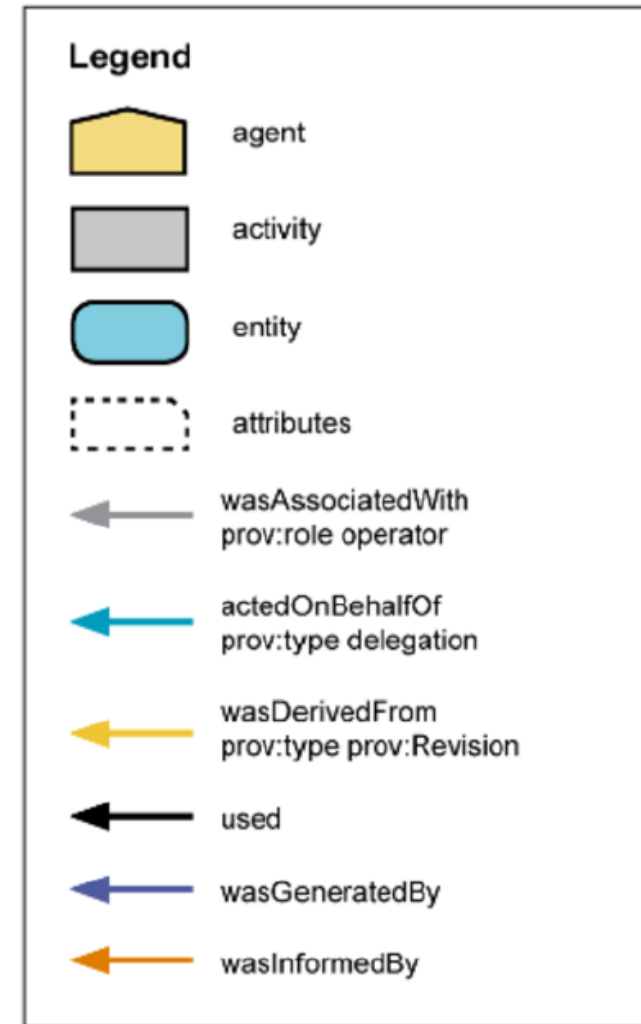
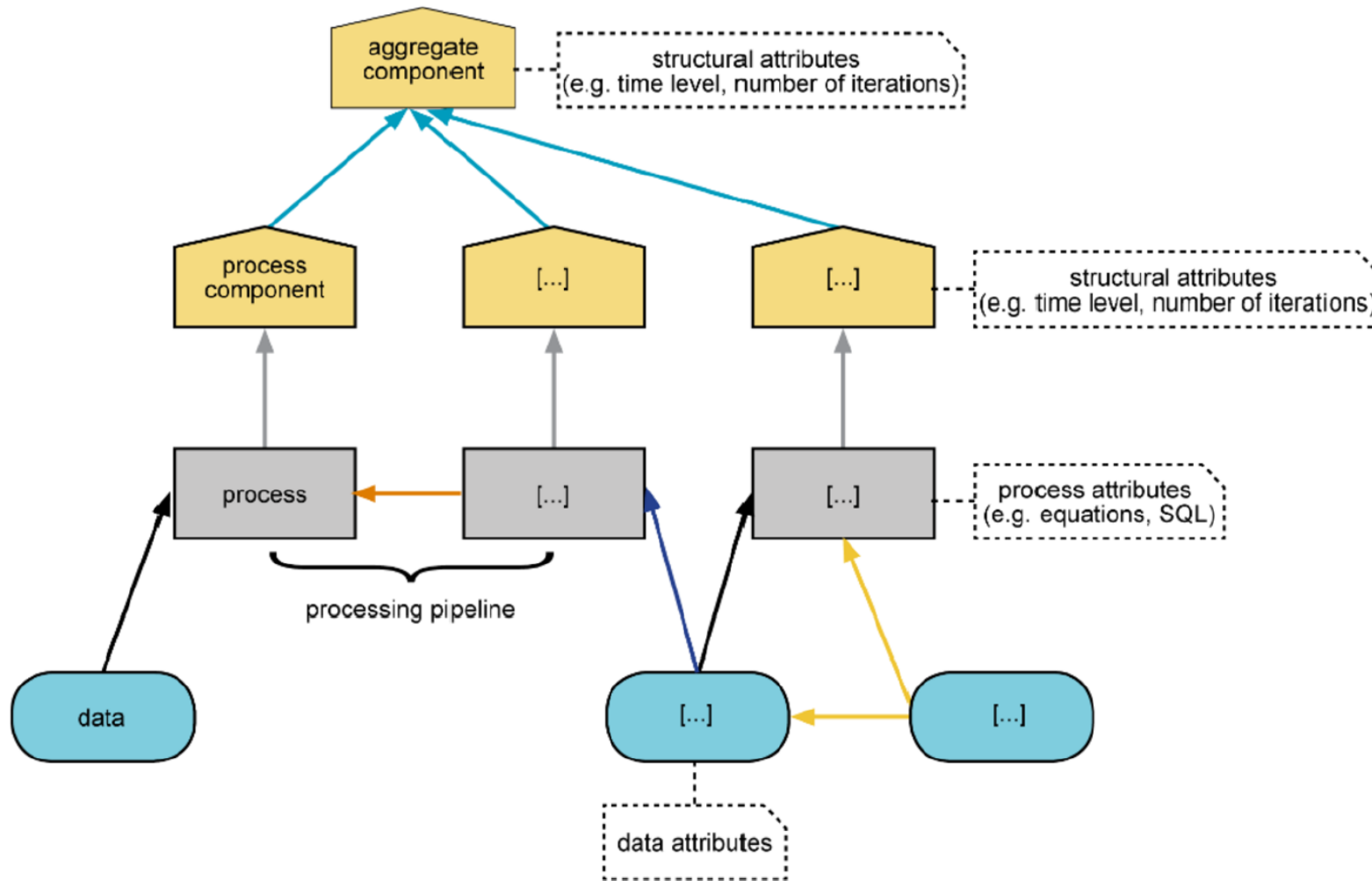
- + Model provenance – autom. fine-grained dynamic model execution sequence and parameter tracking
- Workflow provenance – autom. model and inputs execution sequence tracking
- Open data and open source software – implementation + data
- Model Application paper – described scenarios and parameters
- Model paper – published concept as algorithms & equations

Implementation of Data Provenance

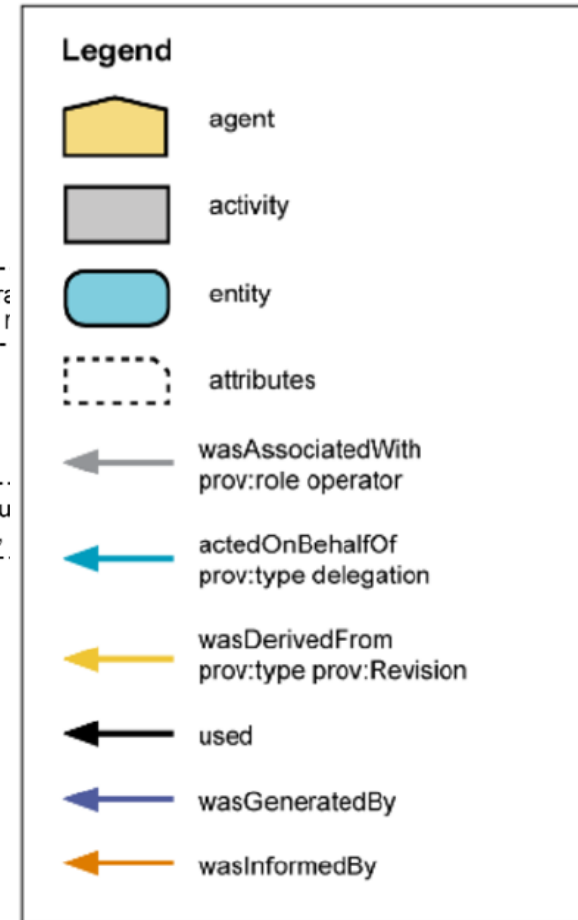
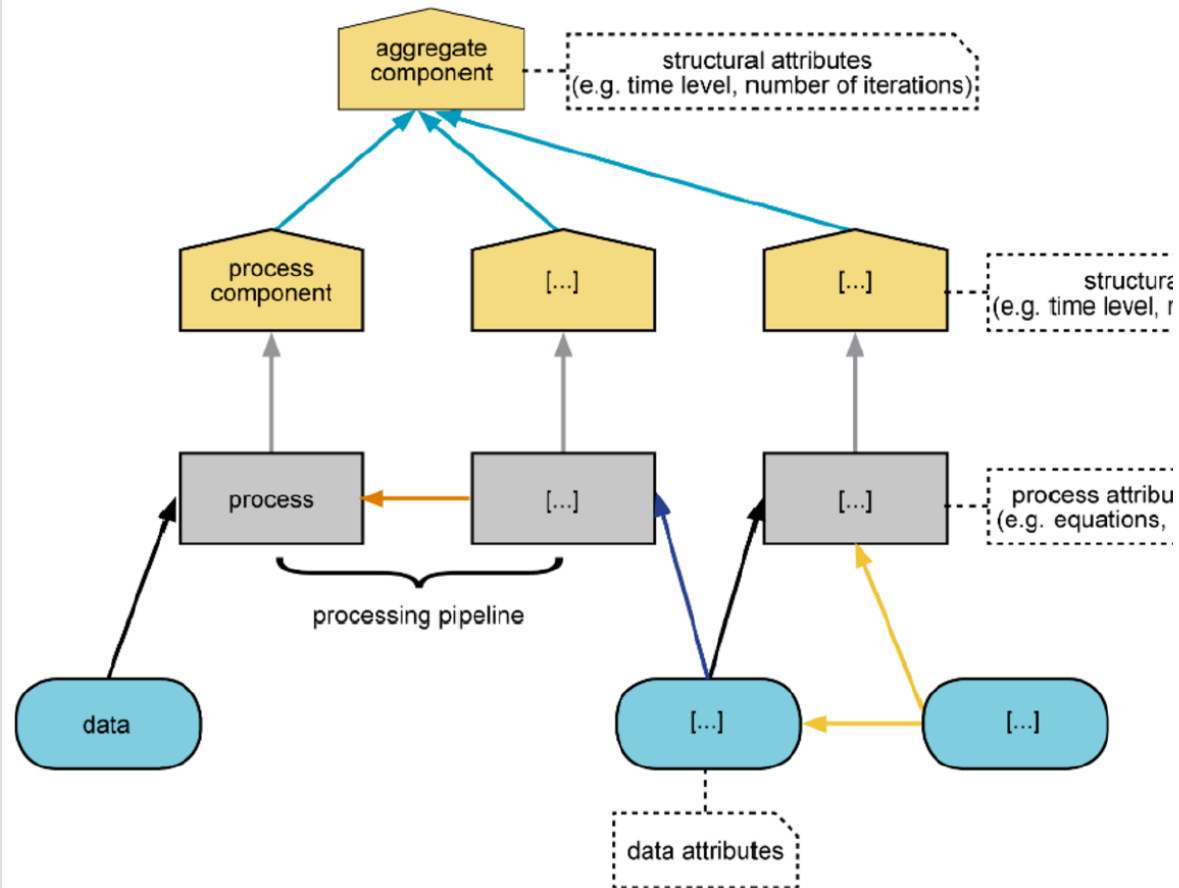
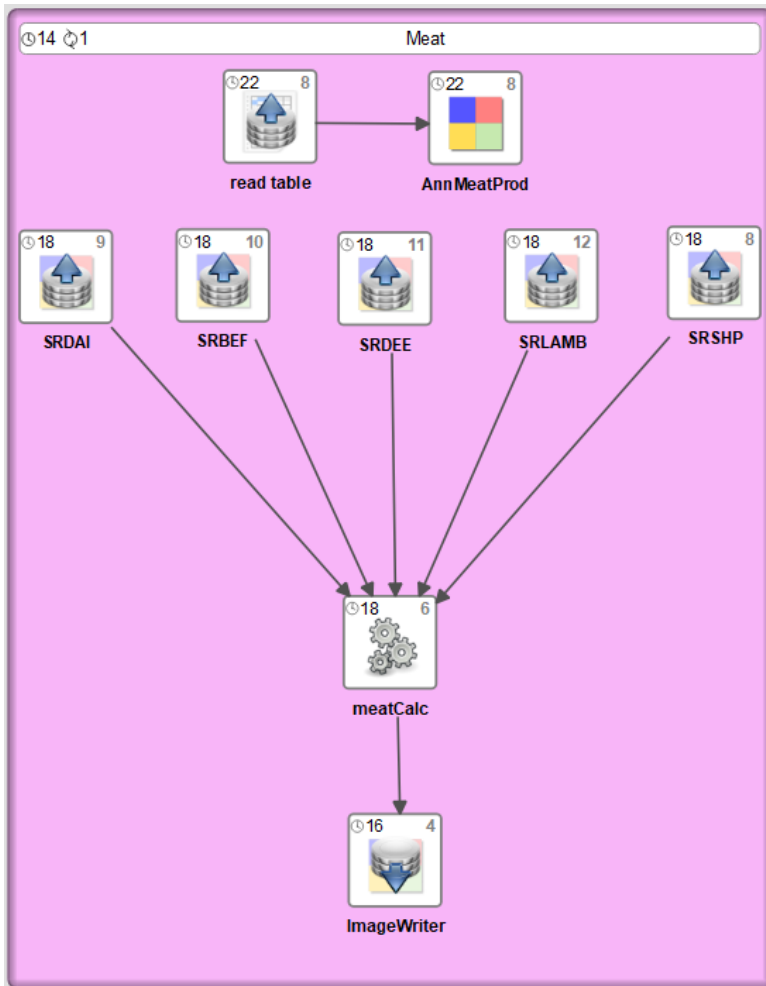
Ecosystem Services Models in the LUMASS modelling and optimisation framework



Implementation of Data Provenance



Implementation of Data Provenance



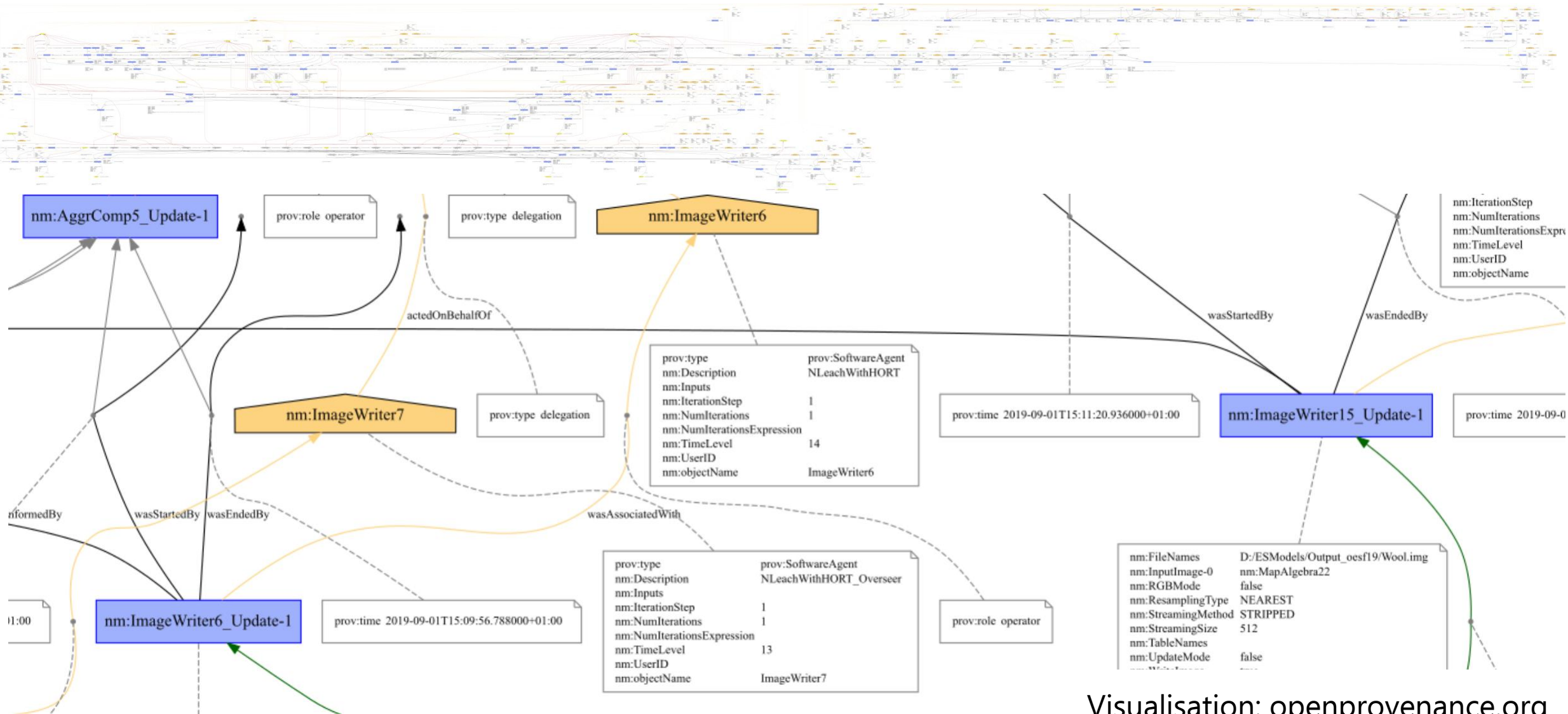
Data Provenance – PROV-N Notation



document

```
prefix nm <https://bitbucket.org/landcareresearch/lumass/nm>
prefix img <https://bitbucket.org/landcareresearch/lumass/img>
prefix db <https://bitbucket.org/landcareresearch/lumass/db>
activity(nm:AggrComp3_Update-1,-,-)
wasAssociatedWith(nm:AggrComp3_Update-1,nm:AggrComp3,-,[prov:role="operator"])
wasStartedBy(nm:DataBuffer21_Update-1,-,nm:AggrComp15_Update-1,2019-09-02T04:30:41.094,[nm:objectName="DataBu
wasStartedBy(nm:TableReader21_Update-1,-,nm:AggrComp15_Update-1,2019-09-02T04:30:41.094)
entity(db:admininfo,[db:DbFileName="D:/lumass_models/ESModels/EcoServices.db",db:TableName="admininfo"])
used(nm:TableReader21_Update-1,db:admininfo,2019-09-02T04:30:41.094)
wasEndedBy(nm:TableReader21_Update-1,-,nm:AggrComp15_Update-1,2019-09-02T04:30:41.094)
wasEndedBy(nm:DataBuffer21_Update-1,-,nm:AggrComp15_Update-1,2019-09-02T04:30:41.094,[nm:objectName="DataBuff
agent(nm:TableReader8,[prov:type='prov:SoftwareAgent',nm:objectName="TableReader8",nm:UserID="tbreader",nm:De
actedOnBehalfOf(nm:TableReader8,nm:AggrComp3,-,[prov:type="delegation"])
activity(nm:TableReader8_Update-1,-,-,[nm:CreateTable="false",nm:FileName="D:/lumass_models/ESModels/EcoServi
wasAssociatedWith(nm:TableReader8_Update-1,nm:TableReader8,-,[prov:role="operator"])
agent(nm:DataBuffer8,[prov:type='prov:SoftwareAgent',nm:objectName="DataBuffer8",nm:UserID="AnnMeatProd",nm:D
actedOnBehalfOf(nm:DataBuffer8,nm:AggrComp3,-,[prov:type="delegation"])
activity(nm:DataBuffer8_Update-1,-,-)
wasAssociatedWith(nm:DataBuffer8_Update-1,nm:DataBuffer8,-,[prov:role="operator"])
wasStartedBy(nm:DataBuffer8_Update-1,-,nm:AggrComp3_Update-1,2019-09-02T04:30:41.094,[nm:objectName="DataBuff
wasStartedBy(nm:TableReader8_Update-1,-,nm:AggrComp3_Update-1,2019-09-02T04:30:41.094)
entity(db:annmeatprod,[db:DbFileName="D:/lumass_models/ESModels/EcoServices.db",db:TableName="annmeatprod"])
used(nm:TableReader8_Update-1,db:annmeatprod,2019-09-02T04:30:41.094)
wasEndedBy(nm:TableReader8_Update-1,-,nm:AggrComp3_Update-1,2019-09-02T04:30:41.094)
wasEndedBy(nm:DataBuffer8_Update-1,-,nm:AggrComp3_Update-1,2019-09-02T04:30:41.094,[nm:objectName="DataBuffer
wasStartedBy(nm:DataBuffer20_Update-1,-,nm:AggrComp15_Update-1,2019-09-02T04:30:41.094,[nm:objectName="DataBu
wasStartedBy(nm:TableReader20_Update-1,-,nm:AggrComp15_Update-1,2019-09-02T04:30:41.094)
entity(db:Output,[db:DbFileName="D:/lumass_models/ESModels/EcoServices.db",db:TableName="Output"])
used(nm:TableReader20_Update-1,db:Output,2019-09-02T04:30:41.094)
wasEndedBy(nm:TableReader20_Update-1,-,nm:AggrComp15_Update-1,2019-09-02T04:30:41.094)
wasEndedBy(nm:DataBuffer20_Update-1,-,nm:AggrComp15_Update-1,2019-09-02T04:30:41.094,[nm:objectName="DataBuff
wasStartedBy(nm:DataBuffer25_Update-1,-,nm:AggrComp15_Update-1,2019-09-02T04:30:41.328,[nm:objectName="DataBu
```

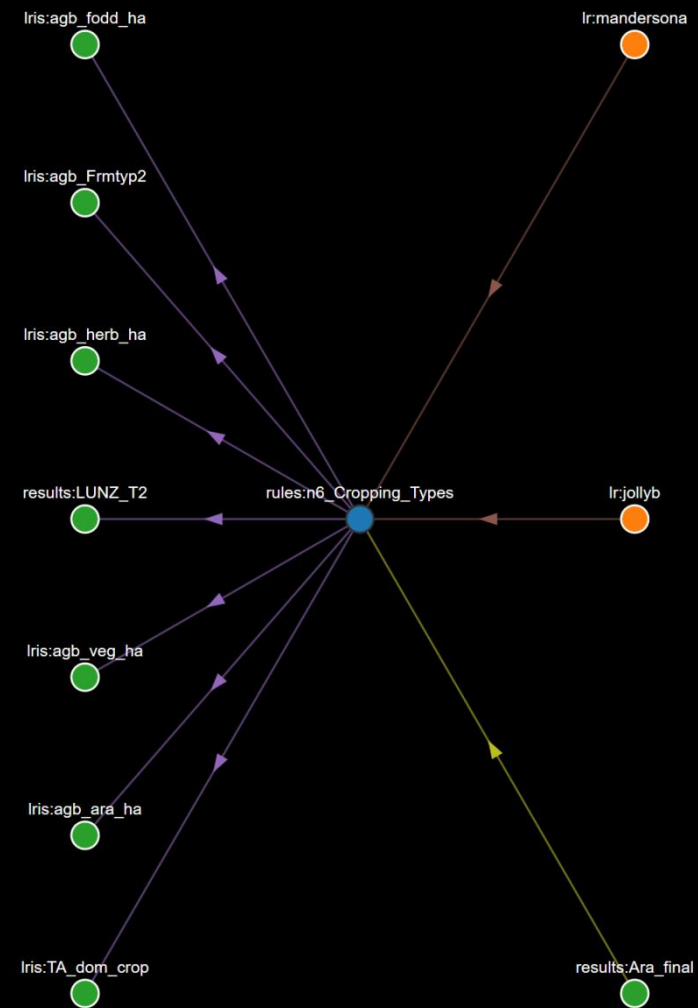
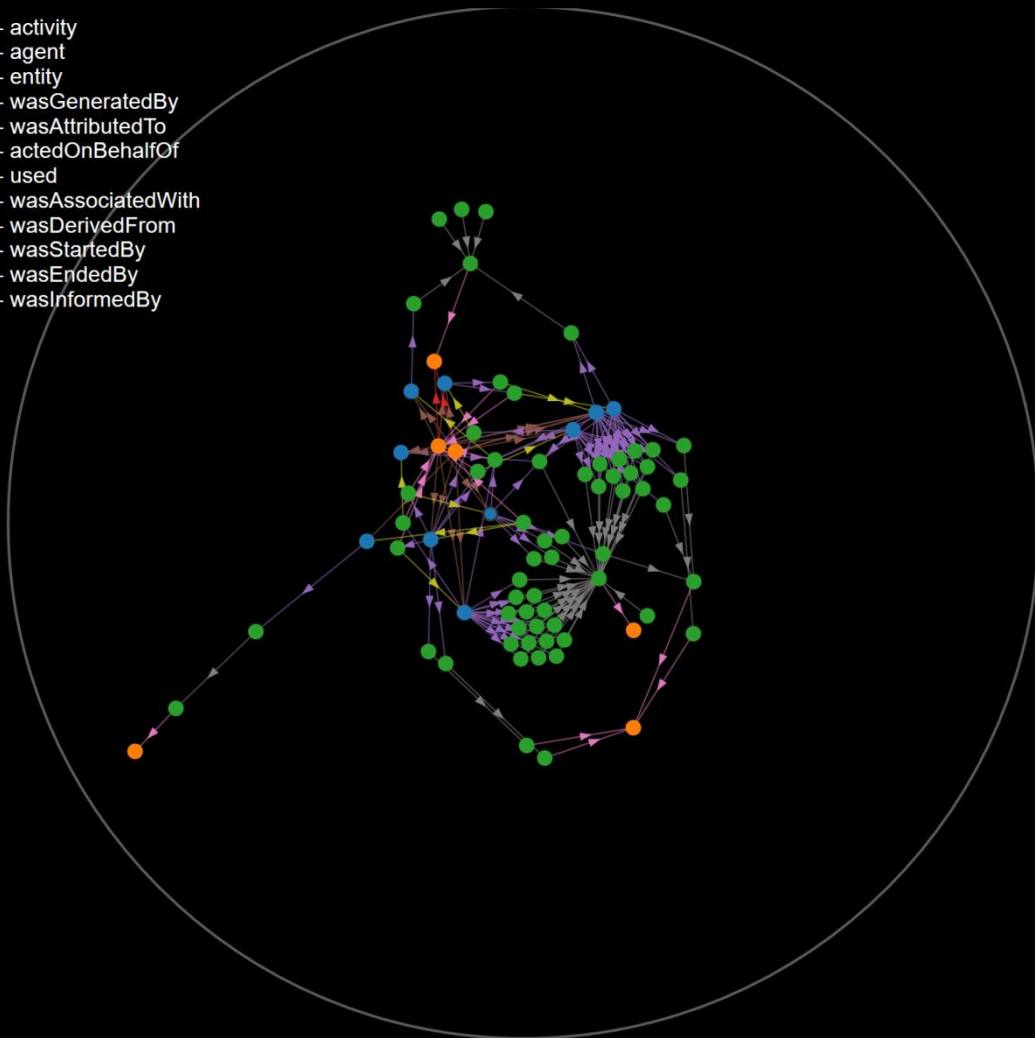
Visualisation of Data Provenance





Visualisation of Data Provenance

- activity
- agent
- entity
- wasGeneratedBy
- wasAttributedTo
- actedOnBehalfOf
- used
- wasAssociatedWith
- wasDerivedFrom
- wasStartedBy
- wasEndedBy
- wasInformedBy





Conclusions

- Successful fine-grained data provenance tracking
- Fine-grained data provenance produces lots of data
- Visualisation of fine-grained data provenance challenging
- Integration of different provenance records needs to be worked on
- What is the right level of granularity (for which application)?
- (How) would you use (consume) data provenance records?



References

Behlhajjame K, B'Far R, Cheney J, Coppens S, Cresswell S, Gil Y, Growth P, Klyne G, Lebo T, McCusker J, Miles S, Myers J, Sahoo S, Tilmes C 2013. PROV-DM: The PROV Data Model, W3C Recommendation. <http://www.w3.org/TR/prov-dm>

Lanter DP 1991. Design of a lineage-based meta-data base for GIS. Cartogr. Geogr. Inf. Syst. 18: 255–261.

Ram S, Liu J 2009. A new perspective on semantics of data provenance. CEUR Workshop Proceedings 526, 1–6.

Spiekermann R, Jolly B, Herzig A, Burleigh T, Medyckyj-Scott D 2019. Implementations of fine-grained automated data provenance to support transparent environmental modelling. Environmental Modelling and Software 118: 134–145. <https://doi.org/10.1016/j.envsoft.2019.04.009>

W3C 2010. Provenance XG Final Report. Report of the W3C. <http://www.w3.org/2005/Incubator/prov/XGR-prov-20101214>

Links

Forced-based interactive provenance visualisation: <http://vizdemo.landcareresearch.co.nz/provis-d3/?example=0>

Python Land Use Classification Tool pyluc: <https://bitbucket.org/landcareresearch/pyluc/src/master/>

Spatial Modelling and Optimisation Framework LUMASS: <https://bitbucket.org/landcareresearch/lumass/wiki/Home>