

What can provenance do for me?

Stian Soiland-Reyes

myGrid, University of Manchester



This work is licensed under a

[Creative Commons Attribution 3.0 Unported License](https://creativecommons.org/licenses/by/3.0/)

Ocean Sampling Day planning Bremen 2013-03-21



Provenance of Stian Soiland-Reyes

- Developer/researcher in **myGrid** team, School of Computer Science, University of Manchester since 2006
- Involved with:
 - **Taverna** - *Scientific workflow system*
 - **myExperiment** – *sharing workflows and artefacts*
 - **Wf4Ever** - *digital preservation (of workflows and workflow runs)*
 - **W3C Provenance WG** – *standards for describing provenance*
 - **Open Annotation** – *standard for tracking who said what about something*



experiment



Workflow

4Ever



Semantic
Web

<http://soiland-reyes.com/stian/work/>

Overview

What is provenance?

- Attribution
- Derivation
- Activities
- PROV model

Aggregating and sharing

Why you want provenance

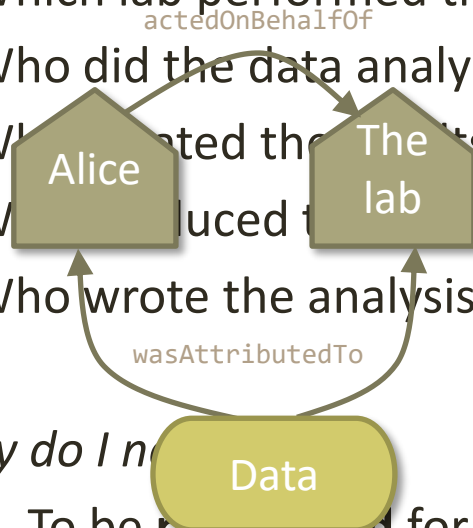
← Newer Older →

g Request to license Dr Stephen Dann's photos via Getty Images

By [Dr Stephen Dann](#)
licensed under Creative Commons Attribution-ShareAlike 2.0 Generic
<http://www.flickr.com/photos/stephendann/3375055368/>

Attribution

- Who collected this sample? Who helped?
- Which lab performed the sequencing?
- Who did the data analysis?
- Who created the datasets?
- Who produced the data this analysis is based on?
- Who wrote the analysis workflow?



Why do I need

- To be **recognized** for my work
- Who should I give **credits** to?
- Who should I **complain** to?
- Can I **trust** them?
- Who should I make **friends** with?

Roles

prov:wasAttributedTo
 prov:actedOnBehalfOf
 dct:creator
 dct:publisher
 pav:authoredBy
 pav:contributedBy
 pav:curatedBy
 pav:createdBy
 pav:importedBy
 pav:providedBy
 ...

Agent types

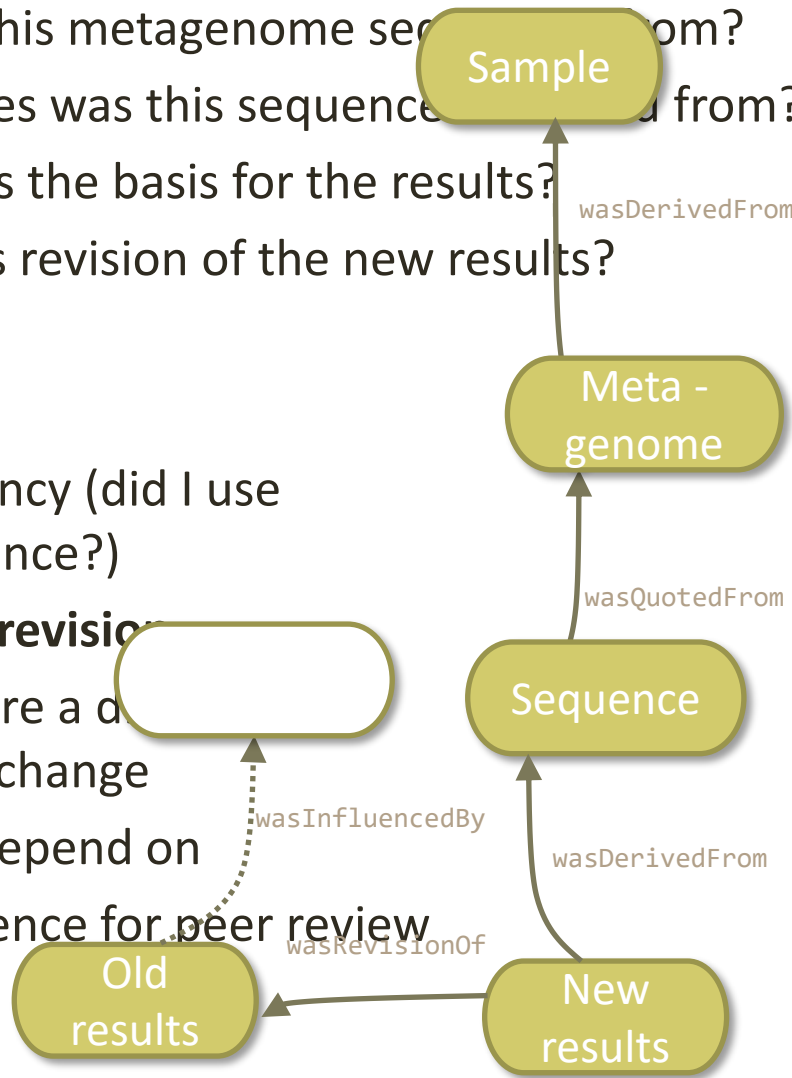
Person
 Organization
 SoftwareAgent

Derivation

- Which sample was this metagenome sequenced from?
- Which meta-genomes was this sequence derived from?
- Which sequence was the basis for the results?
- What is the previous revision of the new results?

Why do I need this?

- To **verify** consistency (did I use the correct sequence?)
- To find the latest **revision**
- To **backtrack** where a change appeared after a change
- To **credit** work I depend on
- Auditing** and defence for peer review

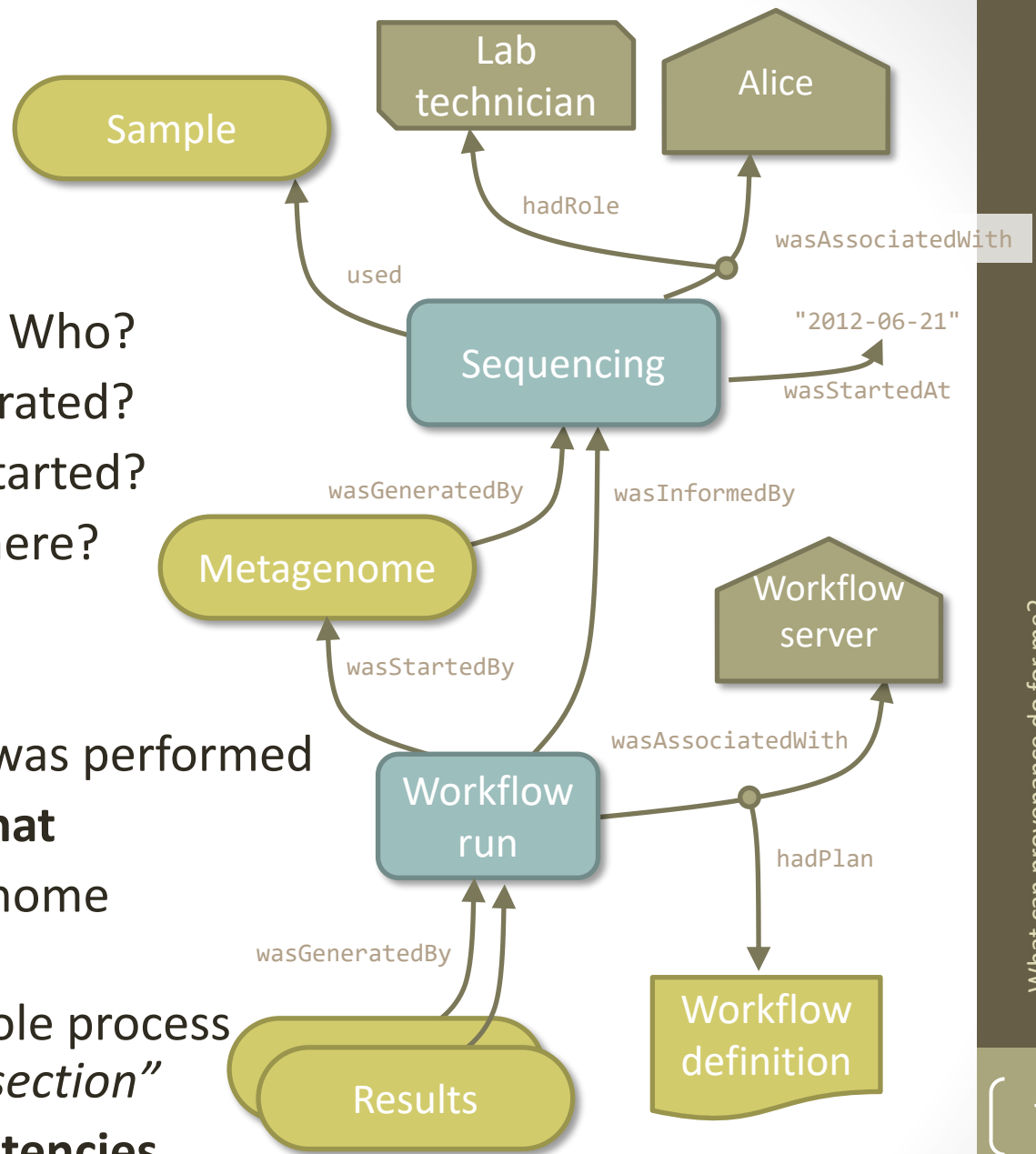


Activities

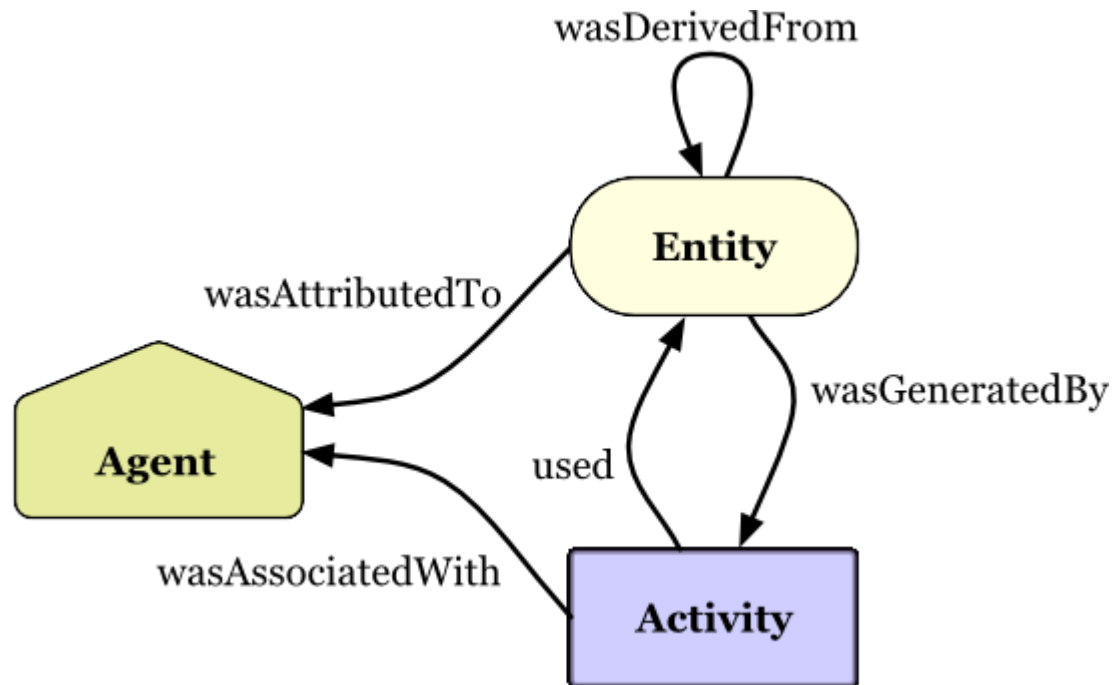
- What happened? When? Who?
- What was used and generated?
- Why was this workflow started?
- Which workflow ran? Where?

Why do I need this?

- To see which **analysis** was performed
- To find out **who** did **what**
- What was the metagenome **used** for?
- To **understand** the whole process
"make me a Methods section"
- To track down **inconsistencies**



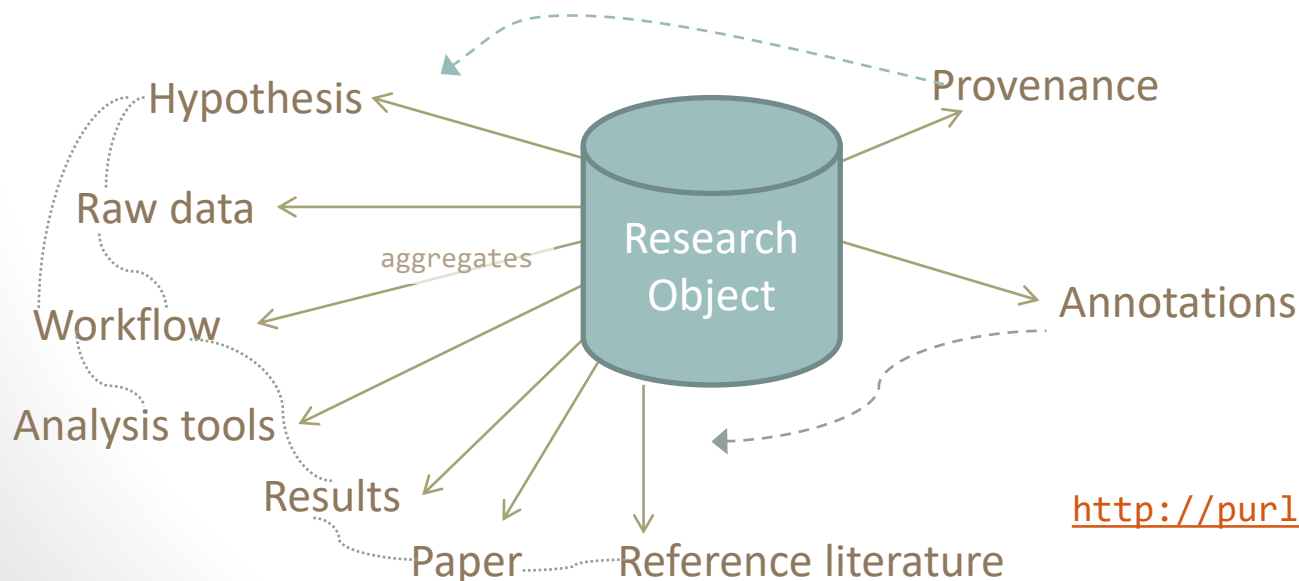
Core PROV model



Copyright © 2013 W3C® (MIT, ERCIM, Keio, Beihang), All Rights Reserved.

Gathering everything

- **Research Objects (RO)** aggregate related resources, their provenance and annotations
- Conveys “everything you need to know” about a study/experiment/analysis/dataset/workflow
- Shareable, evolvable, contributable, citable
- ROs have their own provenance and lifecycles



<http://purl.org/wf4ever/model>



- ## What can provenance do for me?

myExperiment Research Objects



[About](#) | [Mailing List](#) | [Publications](#)

[Logout](#) | [Give us Feedback](#) | [Invite](#)

[Home](#)

[Users](#)

[Groups](#)

[Workflows](#)

[Files](#)

[Packs](#)

[Services](#)

All ▼

[Home](#) > [Packs](#) > [GWAS to pathway](#)

Pack: GWAS to pathway

Created at: 07/02/13 @ 08:32:47

[Tags \(0\)](#) | [Featured in Packs \(0\)](#) | [Favourited By \(0\)](#) | [Comments \(0\)](#)

Title: GWAS to pathway

Research object: <http://sandbox.wf4ever-project.org/rod/ROs/Pack384/>

Description

This pack is for a workflow that finds KEGG pathways for genes from a GWAS.

Uploaded by

- [Stian Soiland-Reyes](#) (last uploaded on 2012-12-24 18:40)
- [Khalid Belhajjame](#) (last uploaded on 2012-11-02 18:40)

Authors

- [Kristina Hettne](#) (last authored on 2012-12-24 18:40)
- [Marco Roos](#) (last authored on 2012-10-15 11:24)

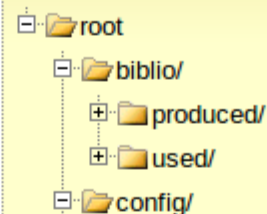
Creator



[Marco Roos](#)

6 items in this pack

Navigate RO



New/Upload

▼



[Stian Soiland-Reyes](#)

- [My Profile](#) [edit]
- [My Messages \(3\)](#)
- [My Memberships](#)
- [My History](#)
- [My News](#)

3 new messages

- [RE: hehe..](#)
- [RE: Hello](#)
- [RE: testing](#)

1 new friendship request

- [Raul Palma](#)

Why you want provenance

- i. To **acknowledge** sources you have based your work on
- ii. Receive **credit** when others uses your work
- iii. Build **trust** (*who did it?*) and **verify** consistency (*was it done correctly?*)
- iv. To **audit** and defend for peer review
- v. Keep track of resources that change over time (**versioning**)
- vi. Investigate and **compare** data (*where did that strange value come from?*)
- vii. Gather everything you need for that **Methods section**
- viii. Facilitate **reproducibility** by tracking activities and their outcomes
- ix. To **prevent decay** by aggregating related resources and their descriptions

Thank you

Questions?

Twitter: @soilandreyes

Skype: soiland

<http://soiland-reyes.com/stian/work/>

<http://www.wf4ever-project.org/>