

Sharing massive data analysis : from provenance to linked experiment reports

Scientific workflows, provenance and linked data to the rescue

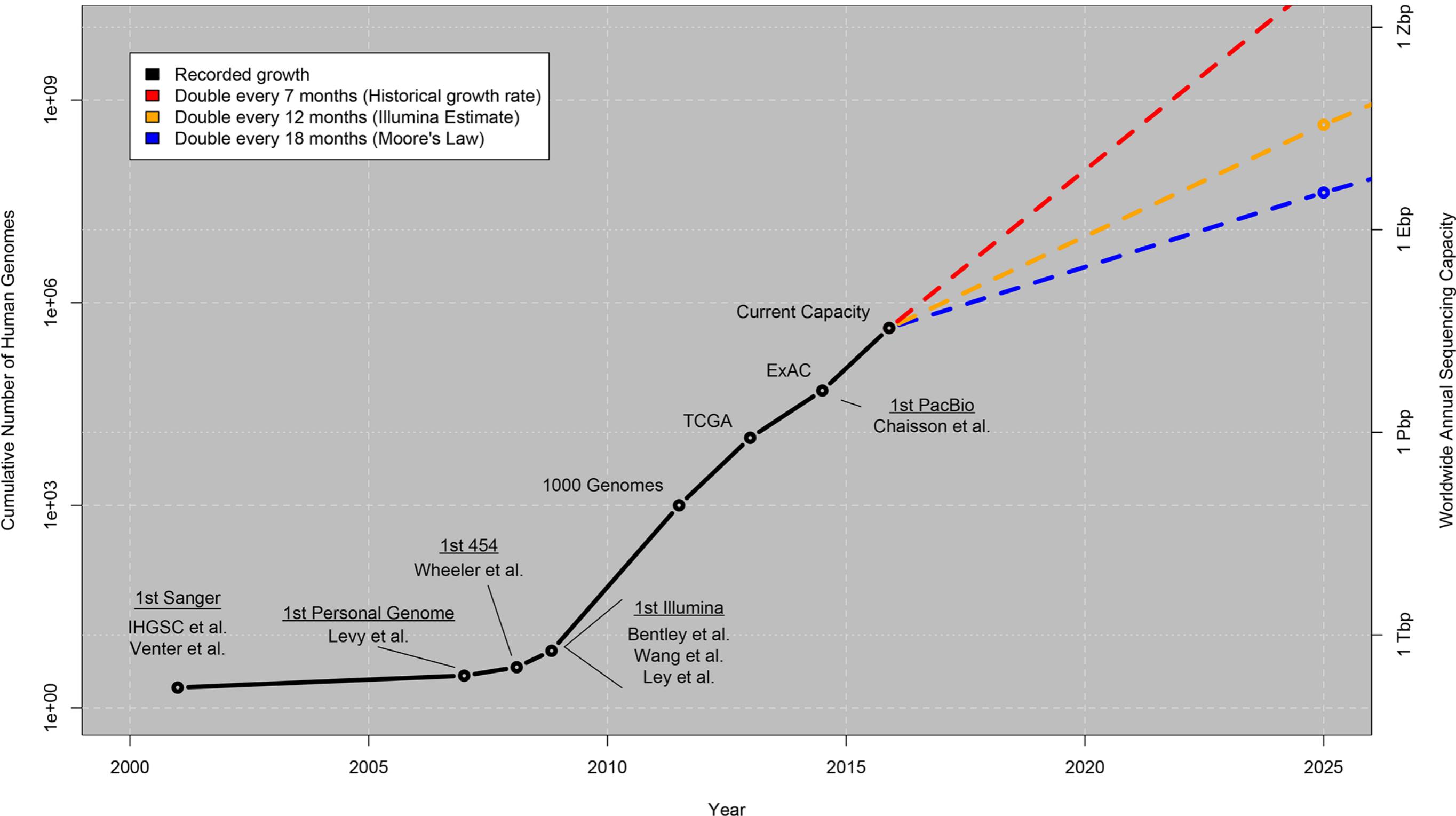
Alban Gaignard, PhD, CNRS

13 november 2018
APSEM 2018

Reproducibility at the age of **massive data**

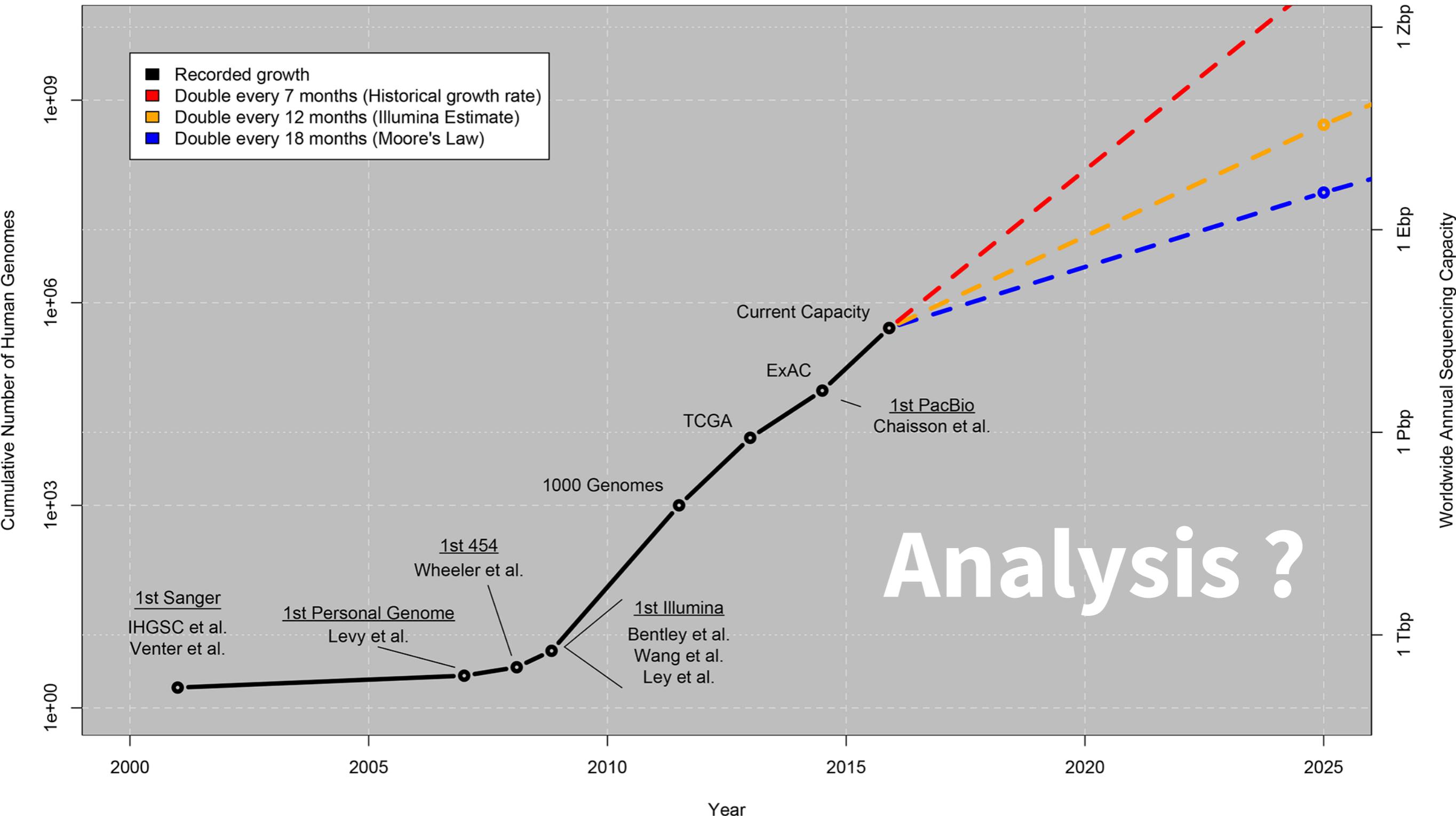
Massive life science data production

Growth of DNA Sequencing



Massive life science data production

Growth of DNA Sequencing



Knowledge production



COMMENT

How science goes wrong

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

© nature

COMMENT

612 | NATURE | VOL 505 | 30 JANUARY 2014

NIH plans to enhance reproducibility

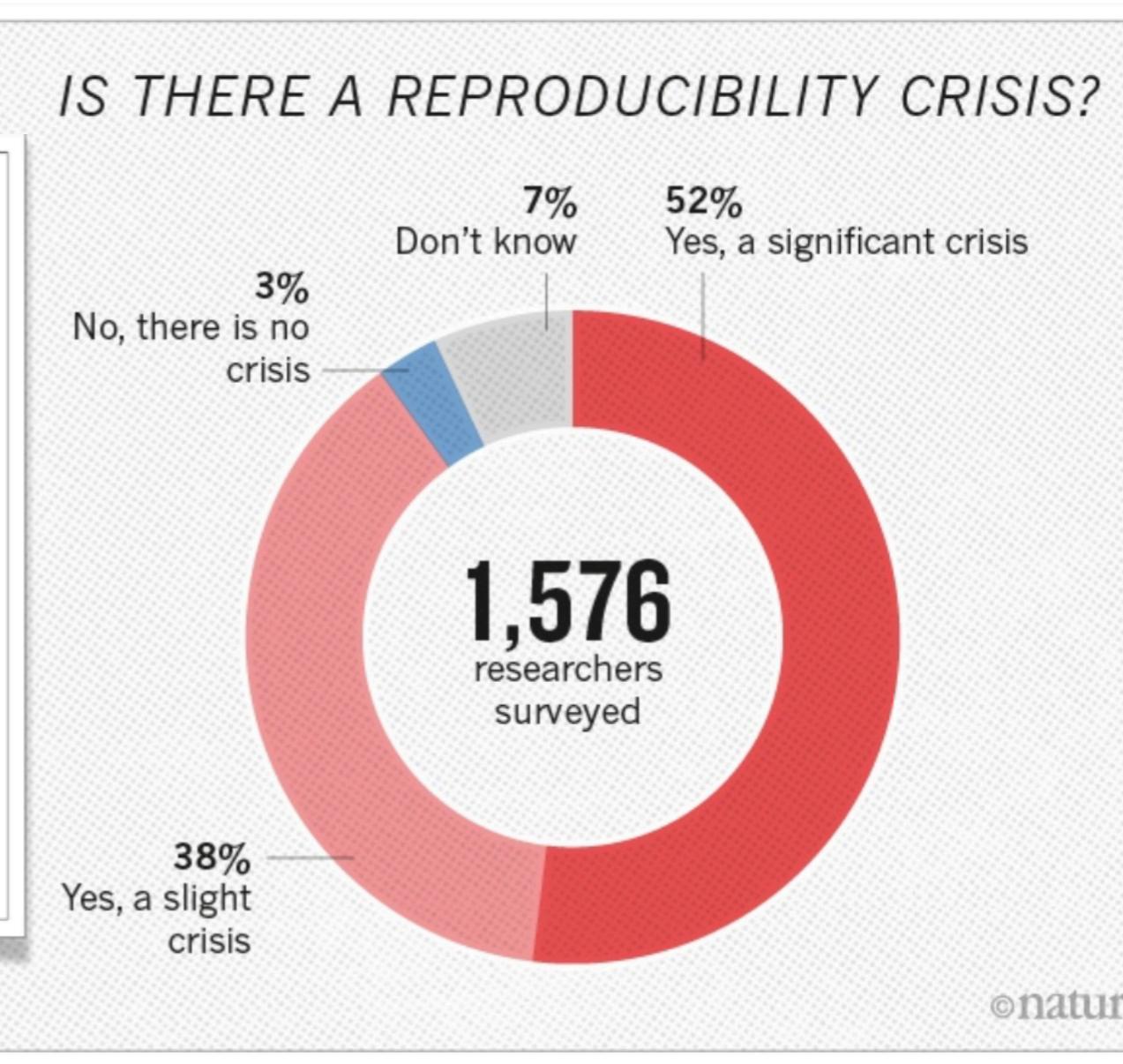
Francis S. Collins and Lawrence A. Tabak discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

A growing chorus of concern, from scientists and laypeople, contends that the complex system for ensuring the reproducibility of biomedical research is failing and is in need of restructuring^{1,2}. As leaders of the US National Institutes of Health (NIH), we share this concern and here explore some of the significant interventions that we are planning.

Science has long been regarded as 'self-correcting', given that it is founded on the

shorter term, however, the checks and balances that once ensured scientific fidelity have been hobbled. This has compromised the ability of today's researchers to reproduce others' findings.

Let's be clear: with rare exceptions, we have no evidence to suggest that irreproducibility is about scientific misconduct. In 2011, the Office of Research Integrity of the US Department of Health and Human Services pursued only 12 such cases³.



« In 2012, Amgen researchers made headlines when they declared that they had been **unable to reproduce the findings in 47 of 53 'landmark' cancer papers** » (doi:10.1038/nature.2016.19269)

Repeat < Replicate < Reproduce < Reuse

Repeat < Replicate < Reproduce < Reuse

Same experiment

Same setup

Same lab

Repeat < Replicate < Reproduce < Reuse

Same experiment

Same setup

~~Same lab~~

Repeat < Replicate < Reproduce < Reuse

Same experiment

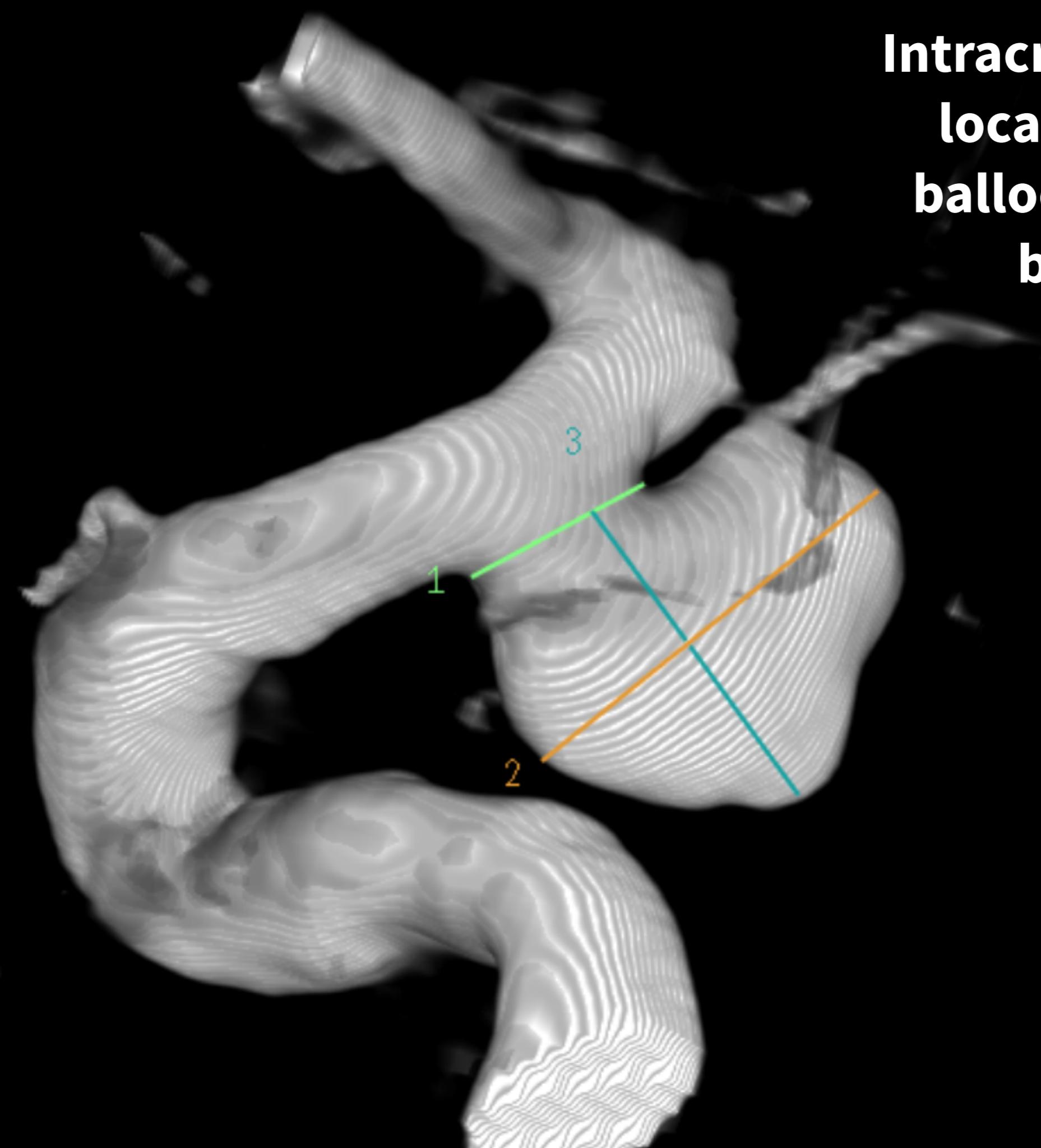
~~Same setup~~

~~Same lab~~

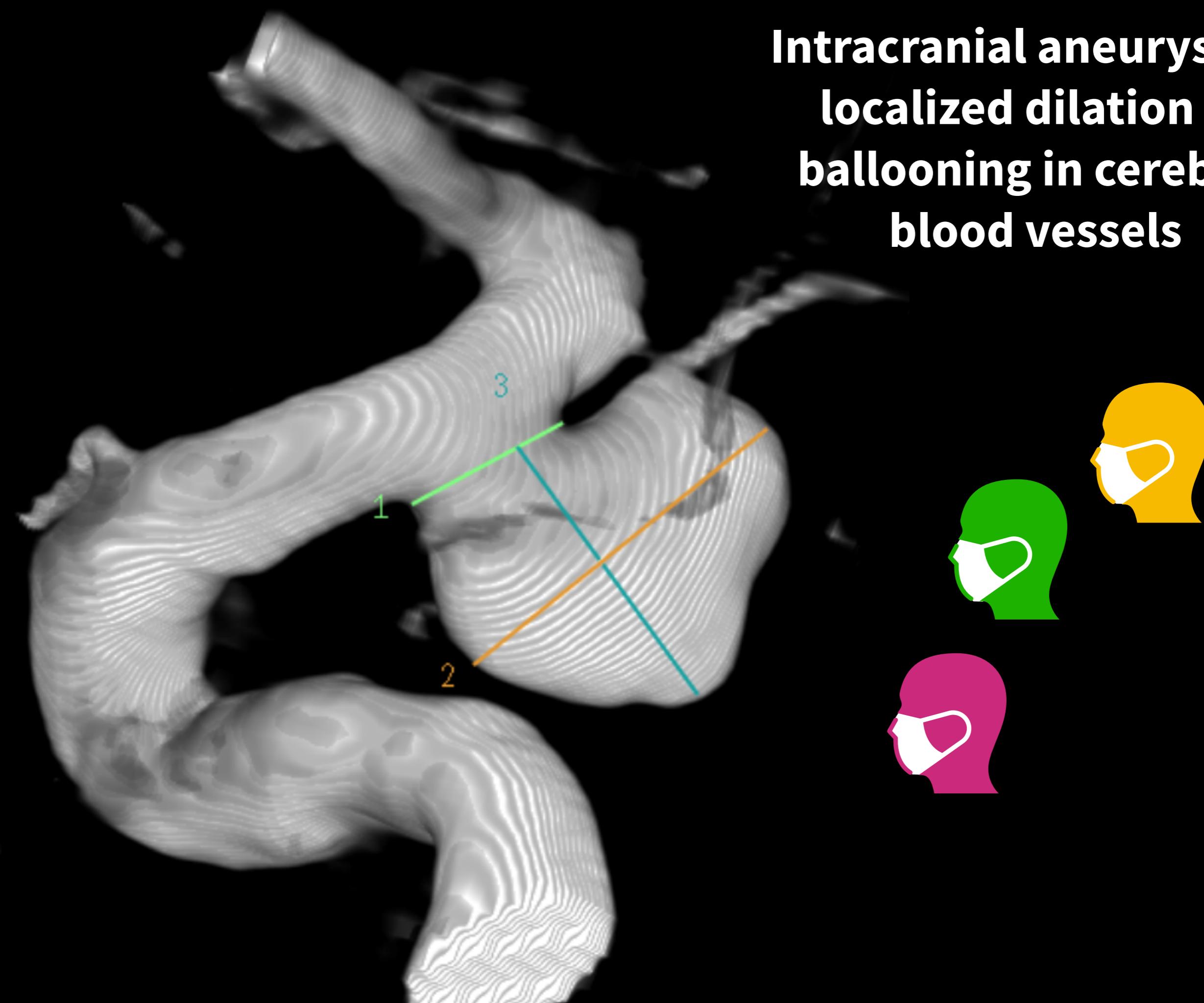
Repeat < Replicate < Reproduce < Reuse

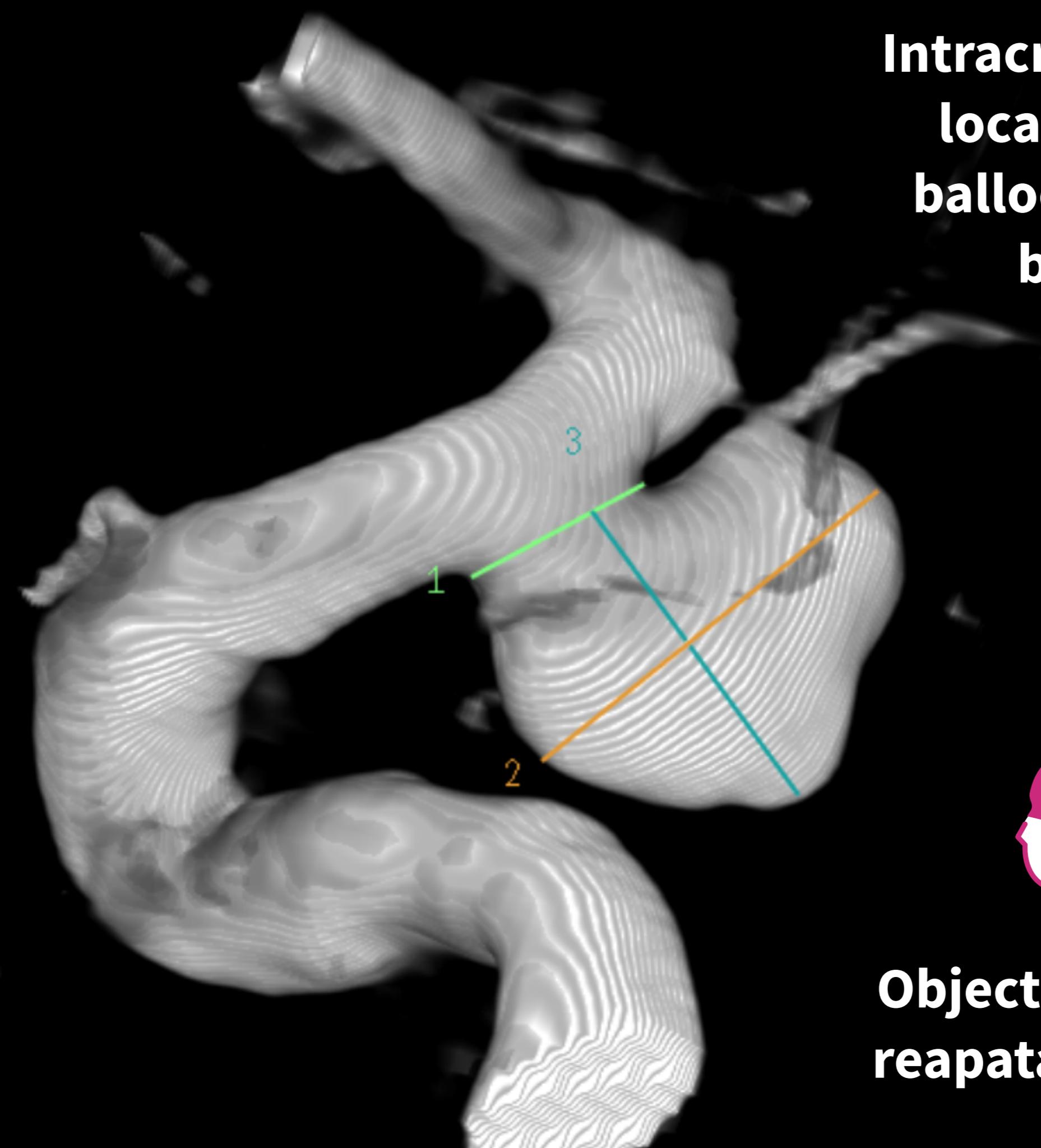
New ideas,
new experiment,
some commonalities

Intracranial aneurysms : localized dilation or ballooning in cerebral blood vessels

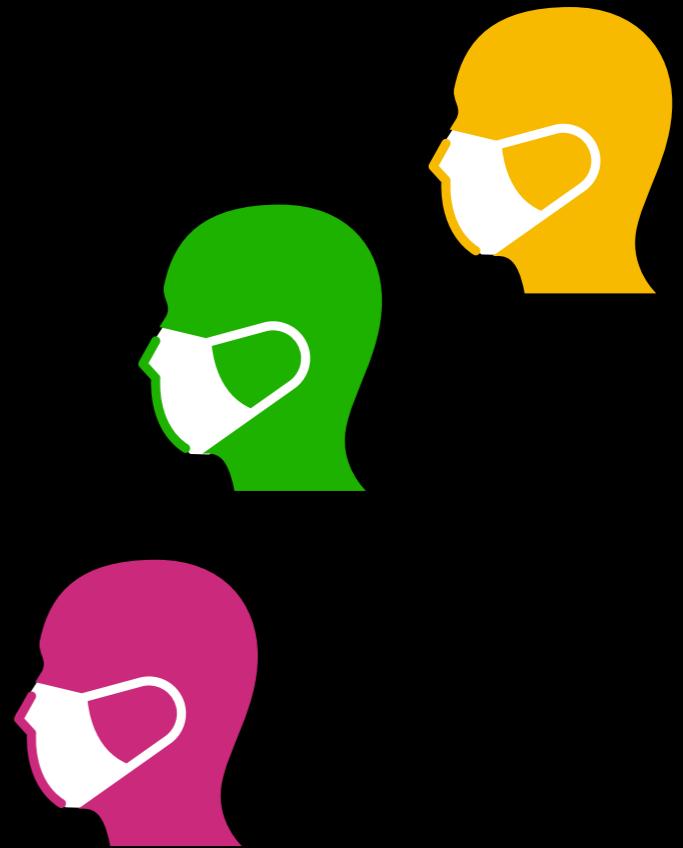


Intracranial aneurysms : localized dilation or ballooning in cerebral blood vessels





Intracranial aneurysms : localized dilation or ballooning in cerebral blood vessels



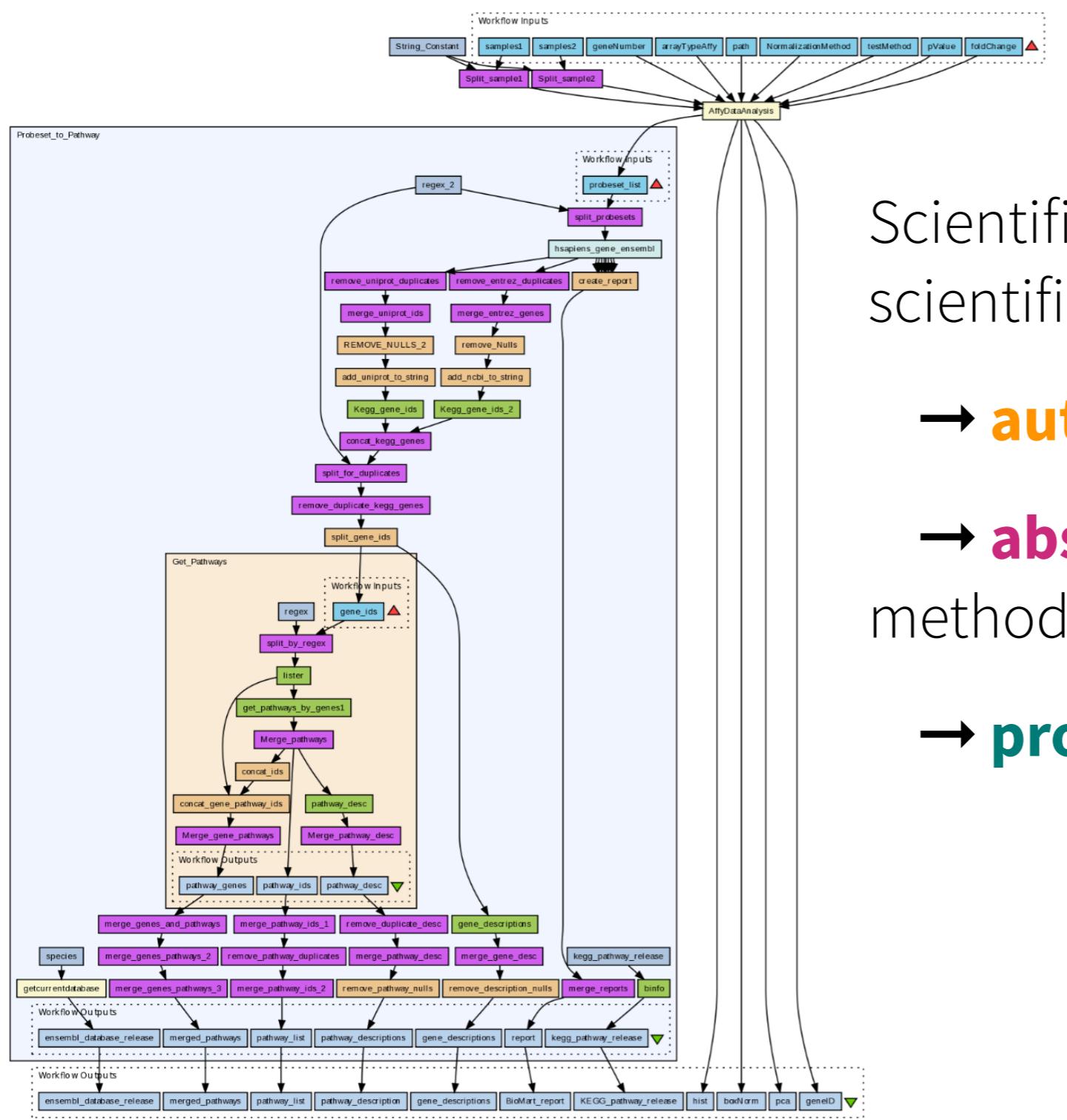
Objective, systematic,
reapatable measures ?

Scientific **workflows** to
the rescue . . .

What is a workflow ?

« Workflows provide a systematic way of describing the **methods** needed and provide the **interface** between **domain specialists** and **computing infrastructures**. »

« Workflow management systems (WMS) **perform** the complex analyses on a variety of **distributed resources** »



Scientific workflows to enhance **trust** in scientific results :

- **automate** data analysis (at scale)
- **abstraction** (describe/share methods)
- **provenance** (~transparency)

[pdtommaso / awesome-pipeline](#)
A curated list of awesome pipeline toolkits inspired by Awesome Sysadmin
#awesome-list #workflow

228 commits · 0 branch · 0 releases · 42 contributors

Branch: master · New pull request · Create new file · Upload files · Find file · Clone or download · Latest commit @7ewf6a 25 days ago

[pdtommaso](#) Update README.md
[CONTRIBUTING.md](#) Added contributing
[README.md](#) Update README.md 4 years ago
25 days ago

[README.md](#)

Awesome Pipeline

A curated list of awesome pipeline toolkits inspired by Awesome Sysadmin

Pipeline frameworks & libraries

- ActionChain - A workflow system for simple linear success/failure workflows.
- Adage - Small package to describe workflows that are not completely known at definition time.
- Airflow - Python-based workflow system created by Airbnb.
- Antha - High-level language for biology.
- Anduri - Component-based workflow framework for scientific data analysis.
- Bds - Scripting language for data pipelines.
- BioMake - GNU-Make-like utility for managing builds and complex workflows.
- BioQueue - Explicit framework with web monitoring and resource estimation.
- Bistro - Library to build and execute typed scientific workflows.
- Bipipe - Tool for running and managing bioinformatics pipelines.
- Briefly - Python Meta-programming Library for Job Flow Control.
- Cluster Flow - Command-line tool which uses common cluster managers to run bioinformatics pipelines.
- ClusterJob - Automated reproducibility and hassle-free submission of computational jobs to clusters.
- Compss - Programming model for distributed infrastructures.
- Conan2 - Light-weight workflow management application.
- Consecution - A Python pipeline abstraction inspired by Apache Storm topologies.
- Cosmos - Python library for massively parallel workflows.
- Cromwell - Workflow Management System geared towards scientific workflows from the Broad Institute.
- Cuneiform - Advanced functional workflow language and framework, implemented in Erlang.
- Dagobah - Simple DAG-based job scheduler in Python.
- Dagr - A scala based DSL and framework for writing and executing bioinformatics pipelines as Directed Acyclic Graphs.
- Dask - A task-based parallel computing library for analytics.
- Dockerflow - Workflow runner that uses Dockerfile to run a series of tasks in Docker.
- DoIt - Task management & automation tool.
- Drake - Robust DSL akin to Make, implemented in Clojure.
- Drake R package - Reproducible and high-performance computing with an easy R-focused interface. Unrelated to Factual's Drake.
- Dray - An engine for managing the execution of container-based workflows.
- Fusion Workflows - A fast, lightweight workflow engine for serverless/FaaS functions.



TensorFlow



Provenance : a way to **reuse**
produced & analysed data

Definition: Oxford dictionary

« The beginning of something's existence; something's origin. »

Definition: Computer Science

« Provenance information describes the **origins** and the **history of data in its life cycle**. »

« Today, data is often made available on the Internet with **no centralized control over its integrity**: data is constantly being created, copied, moved around, and combined indiscriminately. Because information sources (or different parts of a single large source) may vary widely in terms of quality, it is essential to provide **provenance** and other context information which can **help end users** judge whether query results are **trustworthy**. »

Representing provenance



PROV-O: The PROV Ontology

W3C Recommendation 30 April 2013

This version:

<http://www.w3.org/TR/2013/REC-prov-o-20130430/>

Latest published version:

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

Implementation report:

<http://www.w3.org/TR/2013/NOTE-prov-implementations-20130430/>

Previous version:

<http://www.w3.org/TR/2013/PR-prov-o-20130312/>

Editors:

[Timothy Lebo](#), Rensselaer Polytechnic Institute, USA

[Satya Sahoo](#), Case Western Reserve University, USA

[Deborah McGuinness](#), Rensselaer Polytechnic Institute, USA

Contributors:

(In alphabetical order)

[Khalid Belhajjame](#), University of Manchester, UK

[James Cheney](#), University of Edinburgh, UK

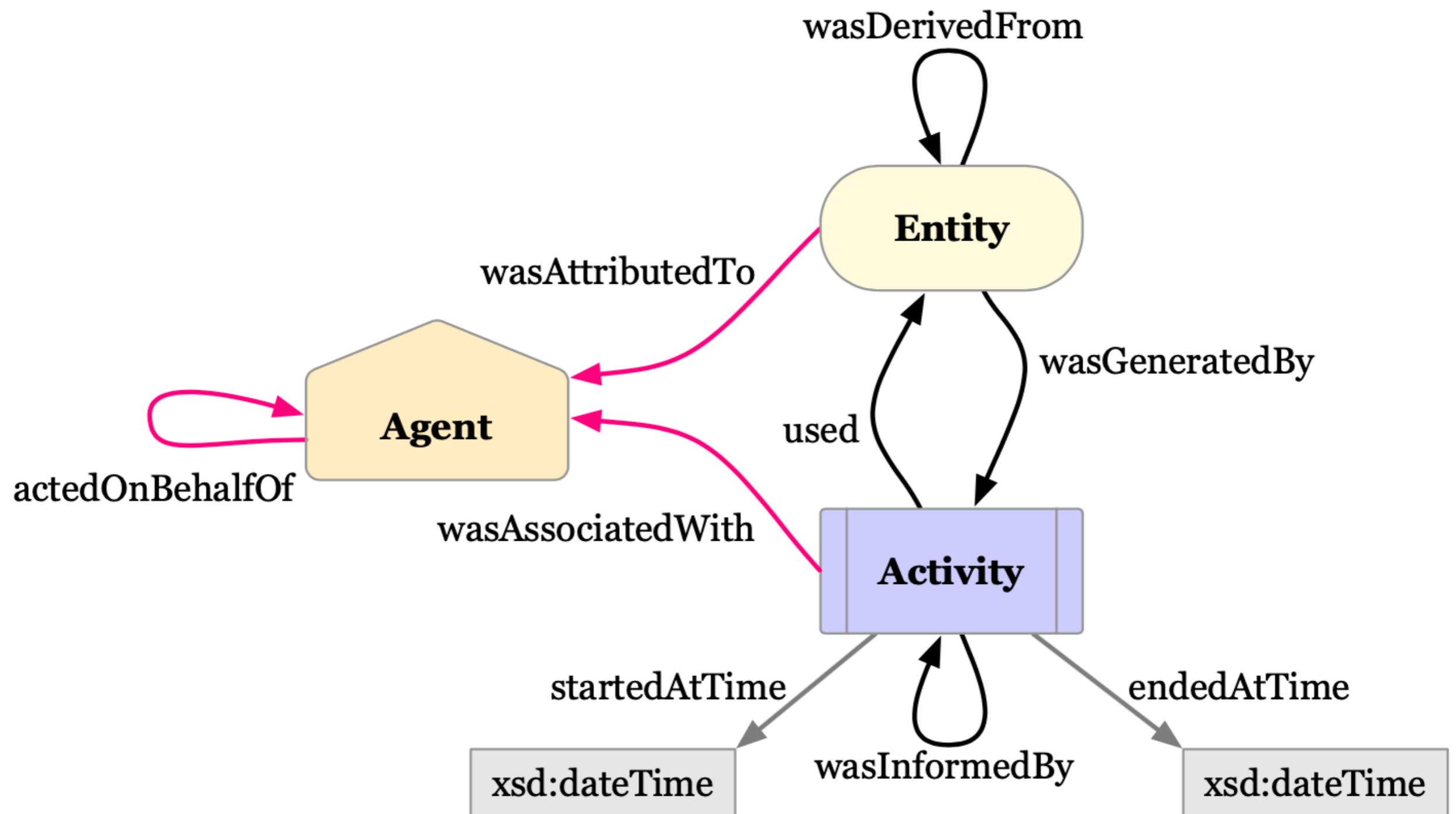
[David Corsar](#), University of Aberdeen, UK

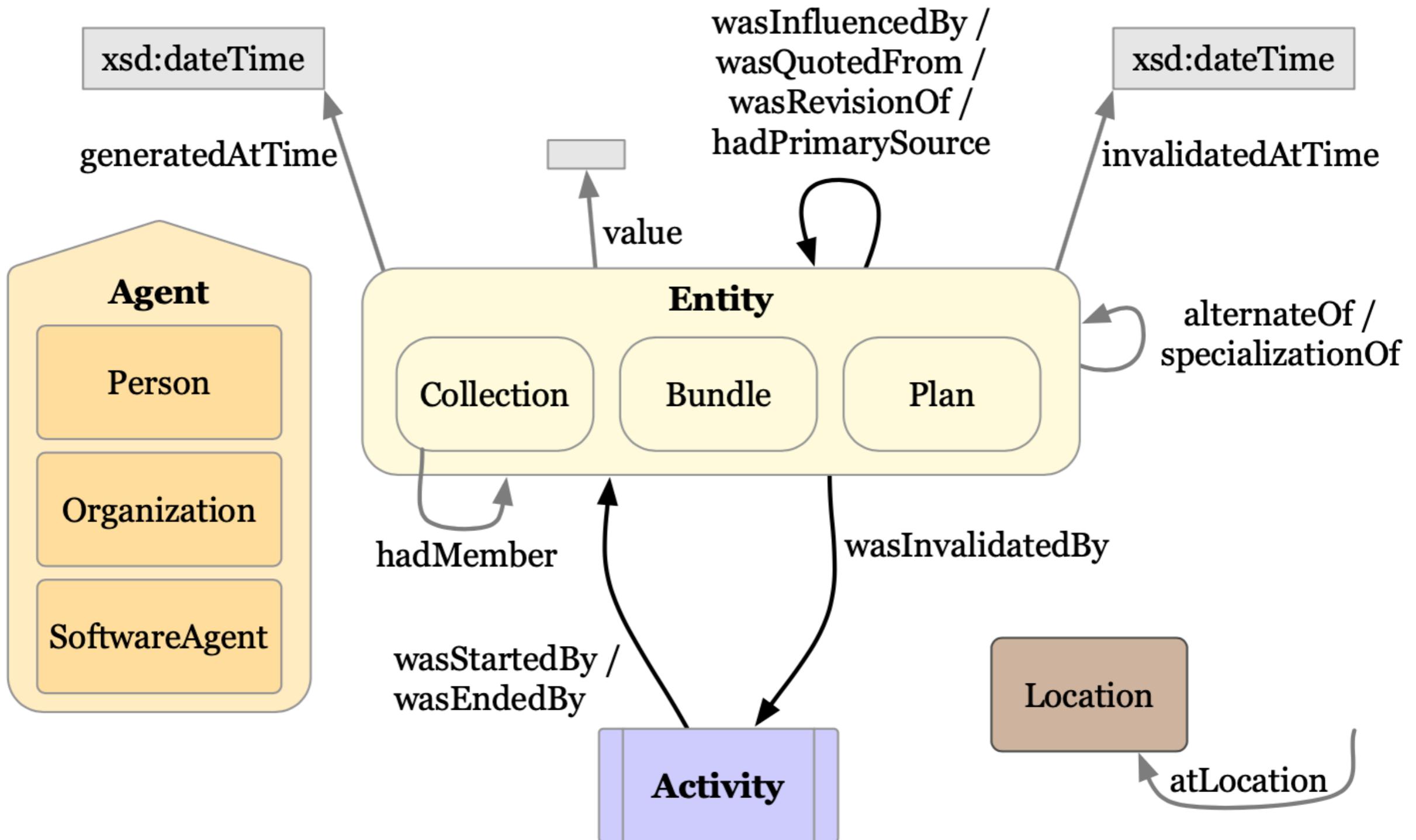
[Daniel Garijo](#), Ontology Engineering Group, Universidad Politécnica de Madrid, Spain

[Stian Soiland-Reyes](#), University of Manchester, UK

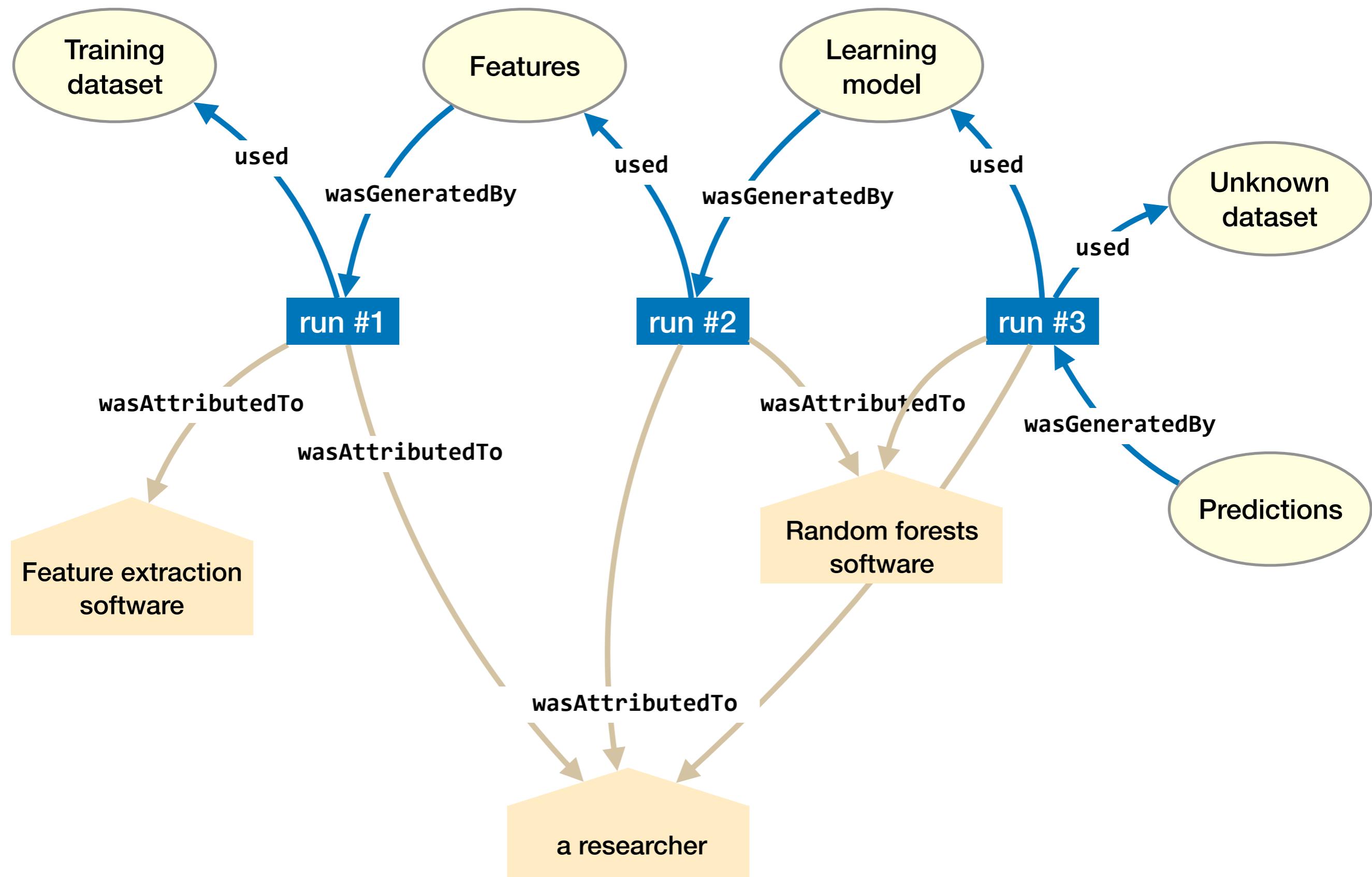
[Stephan Zednik](#), Rensselaer Polytechnic Institute, USA

[Jun Zhao](#), University of Oxford, UK









Reasoning with provenance



Constraints of the PROV Data Model

W3C Recommendation 30 April 2013

This version:

<http://www.w3.org/TR/2013/REC-prov-constraints-20130430/>

Latest published version:

<http://www.w3.org/TR/prov-constraints/>

Test suite:

<http://dvcs.w3.org/hg/prov/raw-file/default/testcases/process.html>

Implementation report:

<http://www.w3.org/TR/2013/NOTE-prov-implementations-20130430/>

Previous version:

<http://www.w3.org/TR/2013/PR-prov-constraints-20130312/> (color-coded diff)

Editors:

[James Cheney](#), University of Edinburgh

[Paolo Missier](#), Newcastle University

[Luc Moreau](#), University of Southampton

Author:

[Tom De Nies](#), iMinds - Ghent University

Please refer to the [errata](#) for this document, which may include some normative corrections.

The English version of this specification is the only normative version. Non-normative [translations](#) may also be available.

5.3 Derivations

Derivations with explicit activity, generation, and usage admit the following inference:

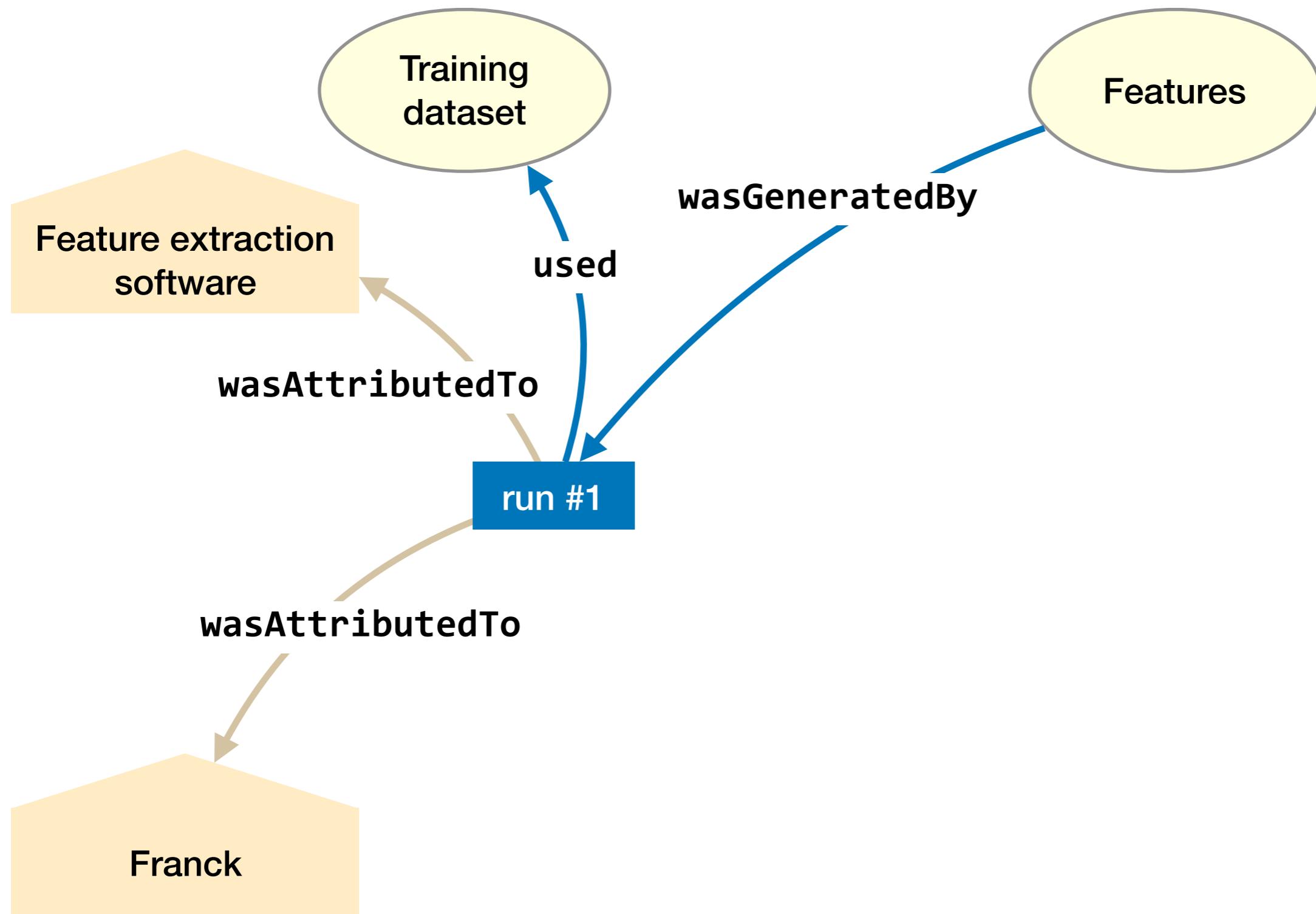
Inference 11 (derivation-generation-use-inference)

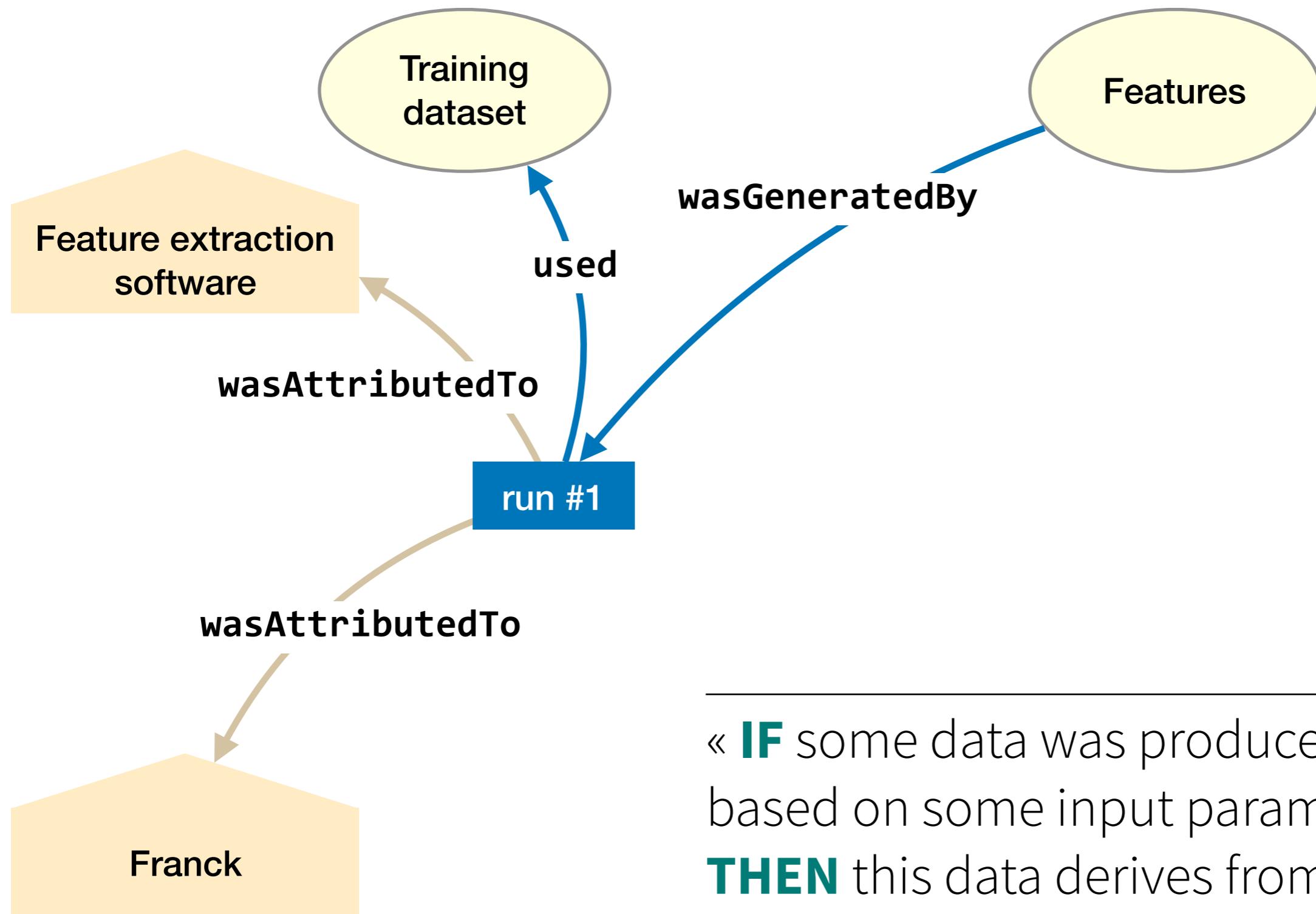
In this inference, none of `a`, `gen2` or `use1` can be placeholders -.

IF `wasDerivedFrom(_id; e2,e1,a,gen2,use1,_attrs)`, **THEN** there exists `_t1` and `_t2` such that `used(use1; a,e1,_t1,[])` and `wasGeneratedBy(gen2; e2,a,_t2,[])`.

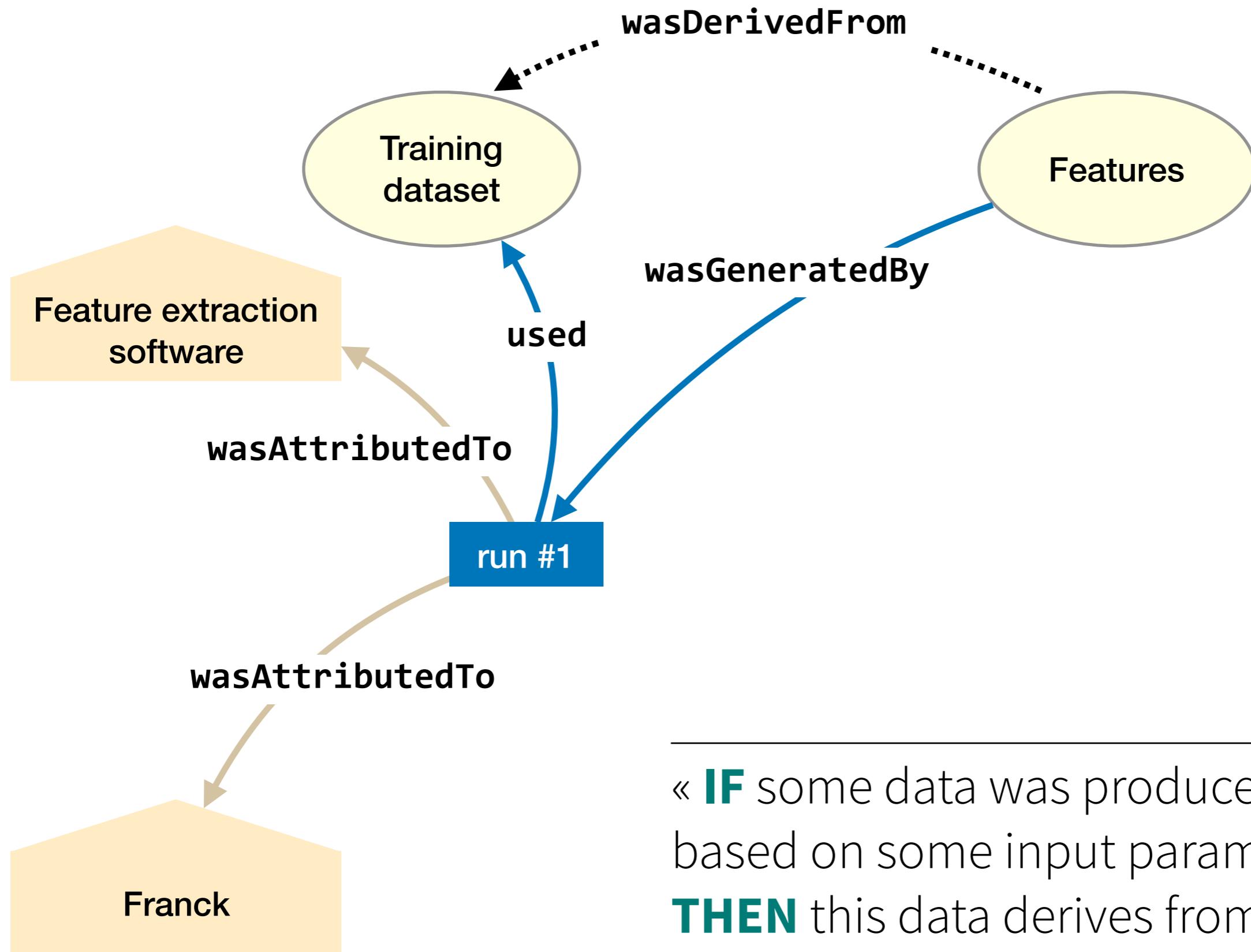
Inference 15 (influence-inference)

1. **IF** `wasGeneratedBy(id; e,a,_t,attrs)` **THEN** `wasInfluencedBy(id; e, a, attrs)`.
2. **IF** `used(id; a,e,_t,attrs)` **THEN** `wasInfluencedBy(id; a, e, attrs)`.
3. **IF** `wasInformedBy(id; a2,a1,attrs)` **THEN** `wasInfluencedBy(id; a2, a1, attrs)`.
4. **IF** `wasStartedBy(id; a2,e,_a1,_t,attrs)` **THEN** `wasInfluencedBy(id; a2, e, attrs)`.
5. **IF** `wasEndedBy(id; a2,e,_a1,_t,attrs)` **THEN** `wasInfluencedBy(id; a2, e, attrs)`.
6. **IF** `wasInvalidatedBy(id; e,a,_t,attrs)` **THEN** `wasInfluencedBy(id; e, a, attrs)`.
7. **IF** `wasDerivedFrom(id; e2, e1, _a, _g, _u, attrs)` **THEN** `wasInfluencedBy(id; e2, e1, attrs)`. Here, `_a, _g, _u` **MAY** be placeholders -.
8. **IF** `wasAttributedTo(id; e,ag,attrs)` **THEN** `wasInfluencedBy(id; e, ag, attrs)`.
9. **IF** `wasAssociatedWith(id; a,ag,_pl,attrs)` **THEN** `wasInfluencedBy(id; a, ag, attrs)`. Here, `_pl` **MAY** be a placeholder -.
10. **IF** `actedOnBehalfOf(id; ag2,ag1,_a,attrs)` **THEN** `wasInfluencedBy(id; ag2, ag1, attrs)`.





« **IF** some data was produced by a tool based on some input parameters,
THEN this data derives from the input parameters »



« **IF** some data was produced by a tool based on some input parameters,
THEN this data derives from the input parameters »

PROV,
how-to record/query ?

Writing PROV statements

```
<Samples/Sample1/VCF/Sample1.hapcaller.g.vcf>
  a prov:Entity;
prov:wasGeneratedBy <http://snakemake-provenance#activity-945311db-8f46-440d-ad05-e8c93832006e>;
prov:wasAttributedTo <#haplotypeCaller>;
rdfs:label "Samples/Sample1/VCF/Sample1.hapcaller.g.vcf";
crypto:sha512
"28df1599bd0053ac08ce195db5750cf5c5f3a05597695449a2acee49b72245aeee290bbc3f9ffcbef3f43b8c07156a175a424a25f2de5
3c326906340081d1c7b"^^xsd:string ;
prov:wasDerivedFrom <BED/capture.extended1000.bed> ;
prov:wasDerivedFrom <Samples/Sample1/BAM/Sample1.final.bam> ;
prov:wasDerivedFrom <../testdata/dbsnp_138.b37.chr22.recode.vcf.gz> ;
prov:wasDerivedFrom <gatkPresent.txt> ;
prov:wasDerivedFrom <../testdata/human_g1k_v37.chr22.fasta> ;
prov:wasDerivedFrom <../testdata/human_g1k_v37.chr22.dict> ;
prov:wasDerivedFrom <Samples/Sample1/BAM/Sample1.final.bai> ;
```

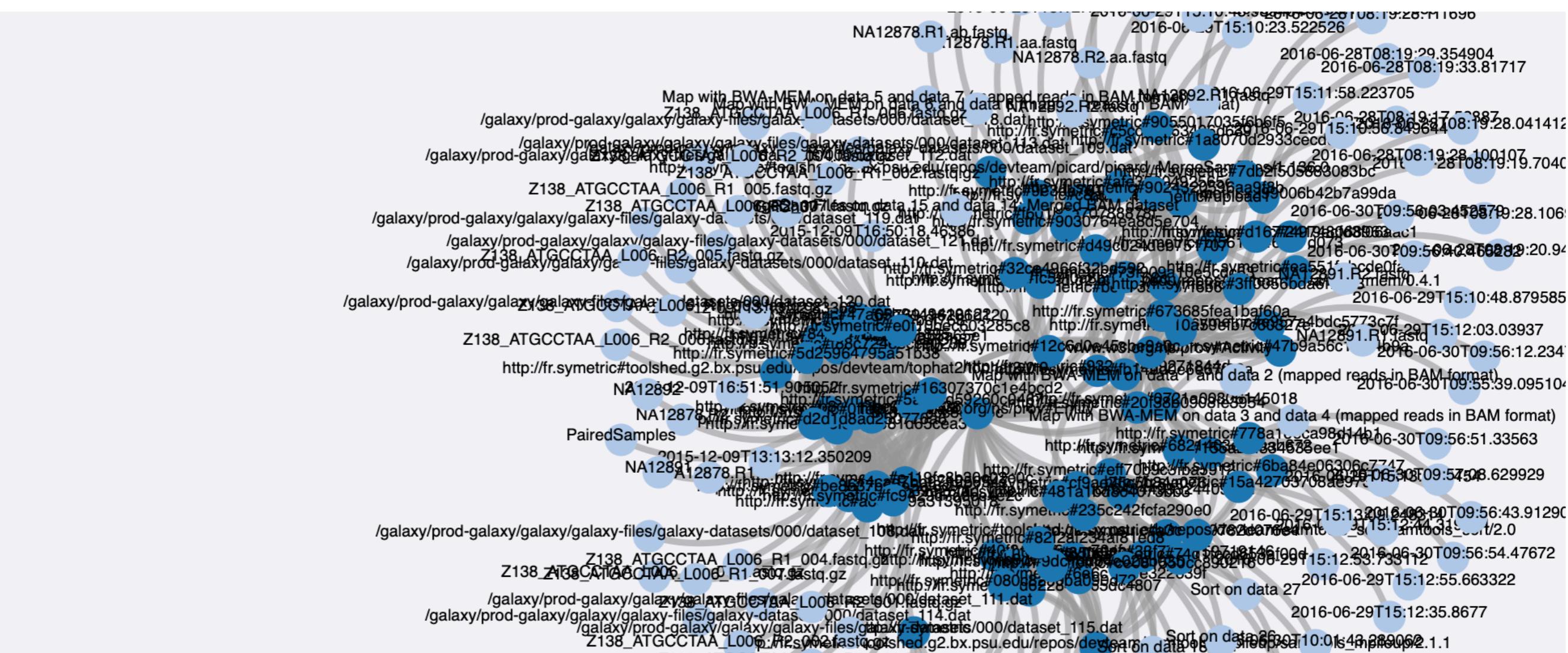
Writing PROV statements

```
<http://snakemake-provenance#activity-945311db-8f46-440d-ad05-e8c93832006e>
a prov:Activity ;
rdfs:comment """
    gatk -Djava.io.tmpdir=Samples/Sample1/VCF -XX:ParallelGCThreads=5 -Xmx3g -T
        HaplotypeCaller                               -R ../testdata/human_g1k_v37.chr22.fasta
        --emitRefConfidence GVCF                      -variant_index_type LINEAR
        --variant_index_parameter 128000                -stand_call_conf 30.0           -nct 1
        -rf ReadLength                   -minRead 0
        -L:capture,BED BED/capture.extended1000.bed
        Samples/Sample1/BAM/Sample1.final.bam          -maxRead 10000
        ../testdata/dbsnp_138.b37.chr22.recode.vcf.gz   -I
        Samples/Sample1/VCF/Sample1.hapcaller.g.vcf      --dbsnp:dbsnp,VCF
                                                -o
"""
prov:wasAssociatedWith <#haplotypeCaller> ;
prov:startedAtTime "2018-04-04T16:19:57.054054"^^xsd:dateTime;
prov:endedAtTime "2018-04-04T16:19:57.054076"^^xsd:dateTime;
prov:used <BED/capture.extended1000.bed> ;
prov:used <Samples/Sample1/BAM/Sample1.final.bam> ;
prov:used <../testdata/dbsnp_138.b37.chr22.recode.vcf.gz> ;
prov:used <gatkPresent.txt> ;
prov:used <../testdata/human_g1k_v37.chr22.fasta> ;
prov:used <../testdata/human_g1k_v37.chr22.dict> ;
prov:used <Samples/Sample1/BAM/Sample1.final.bai> ;
```

Querying PROV graphs

```
CONSTRUCT {?x ?p ?y} WHERE {?x ?p ?y}
```

Query



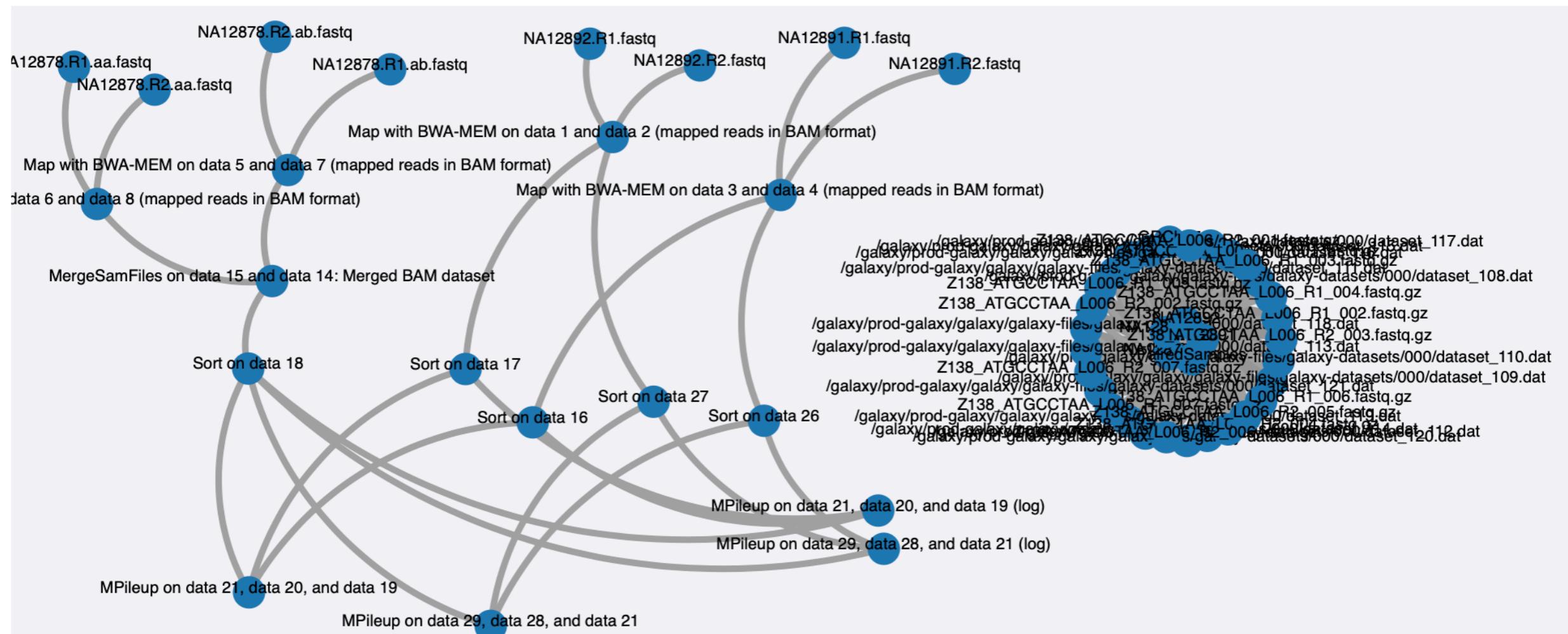
Querying PROV graphs

```
PREFIX prov: <http://www.w3.org/ns/prov#>
```

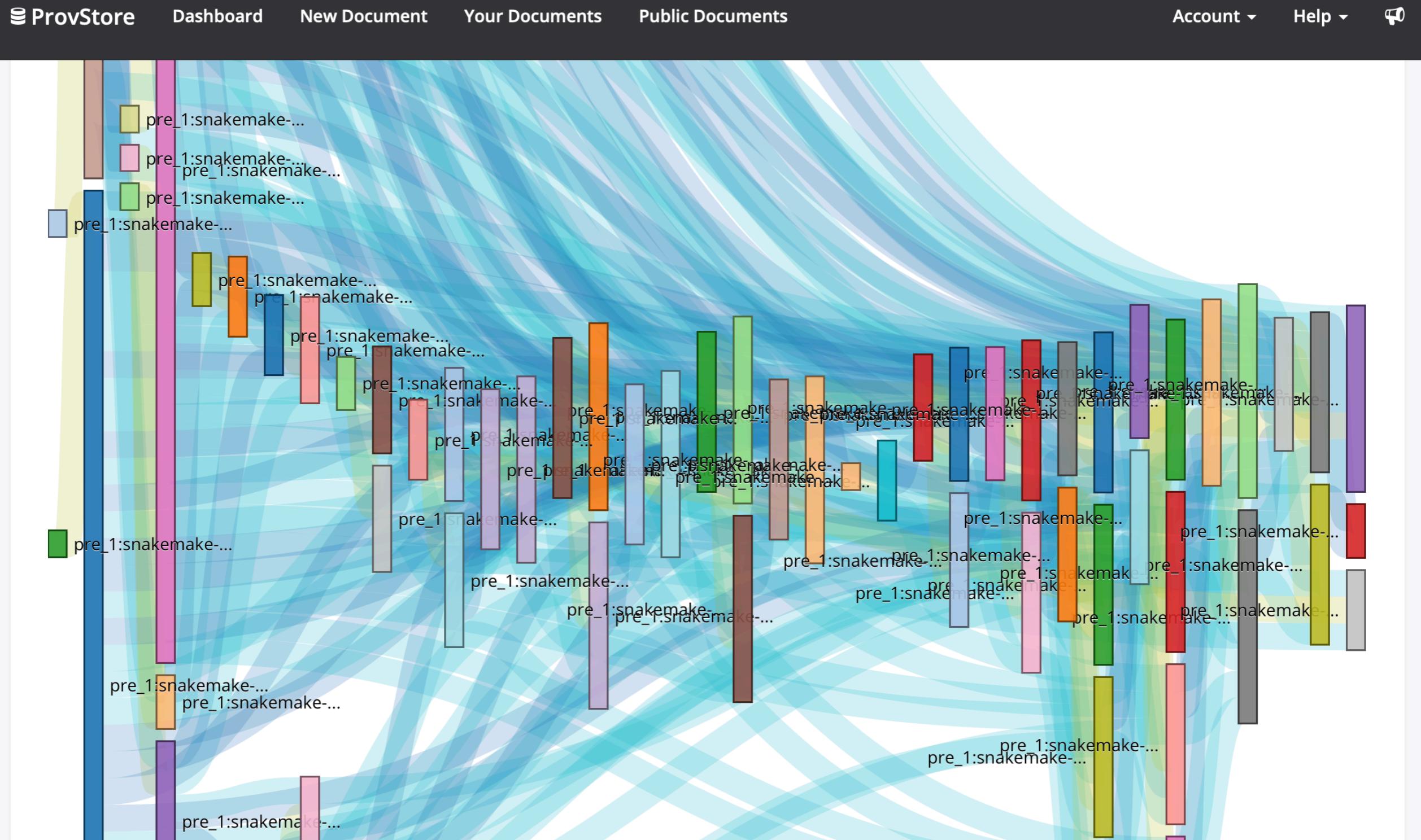
```
CONSTRUCT {  
    ?x prov:wasDerivedFrom ?y  
}
```

```
WHERE {  
    ?x_uri prov:wasDerivedFrom ?y_uri .  
    ?x_uri rdfs:label ?x .  
    ?y_uri rdfs:label ?y .  
}
```

Query



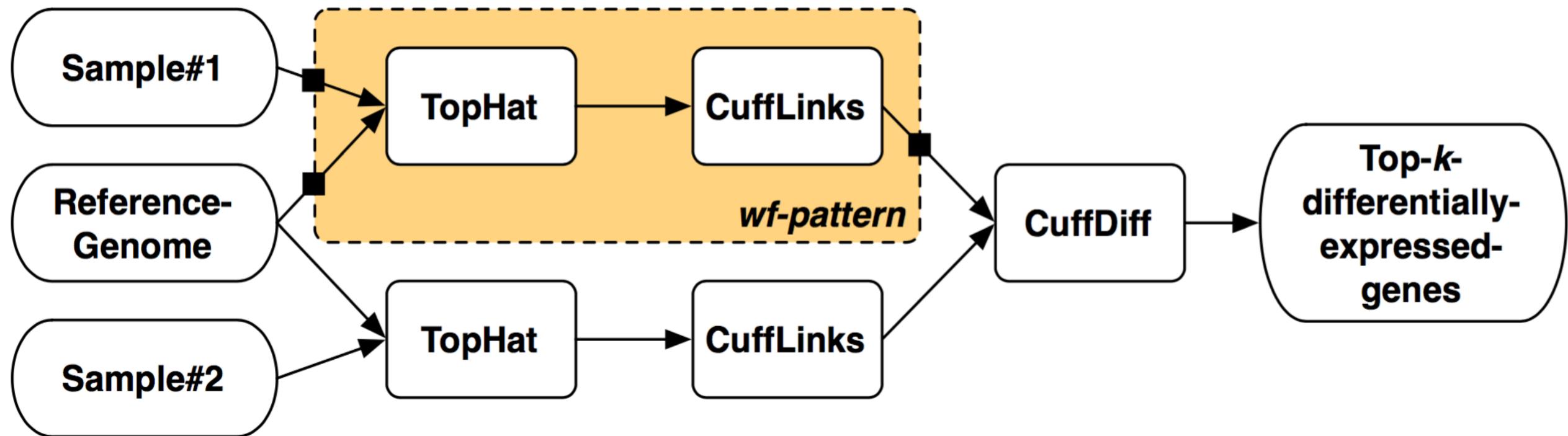
PROV store



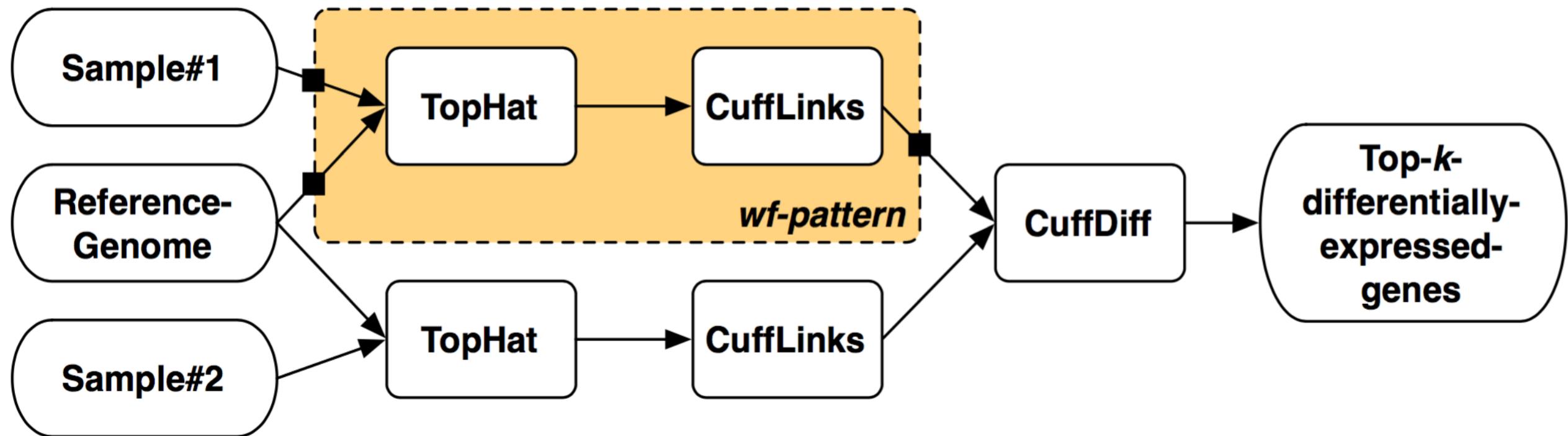
Still open issues ...

Reuse instead of
re-execution ?

RNA-seq data analysis workflow

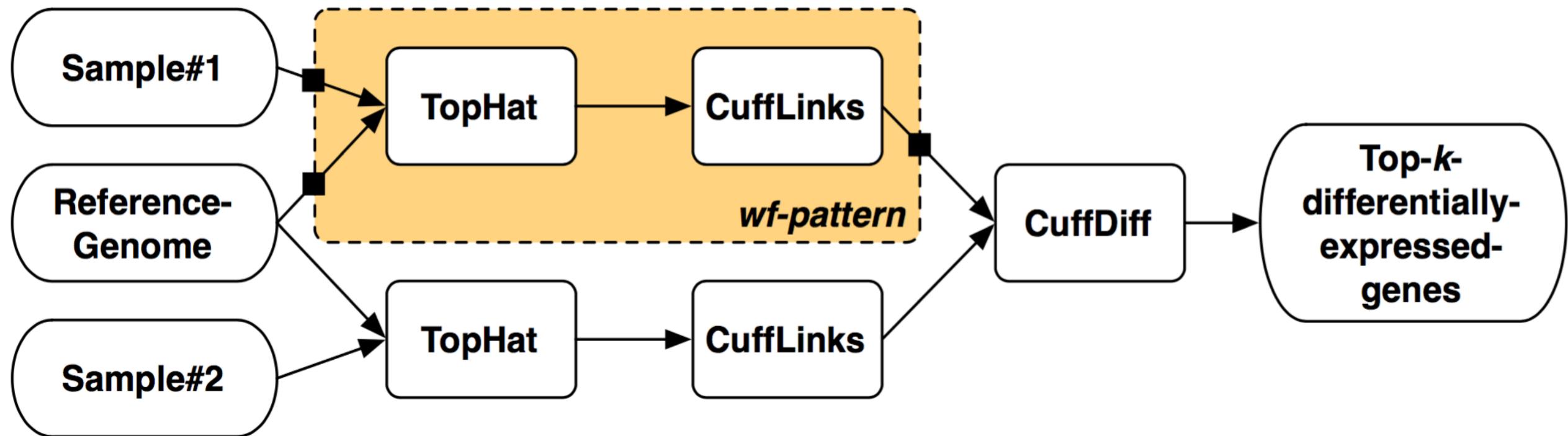


RNA-seq data analysis workflow



**Compute and
storage intensive**

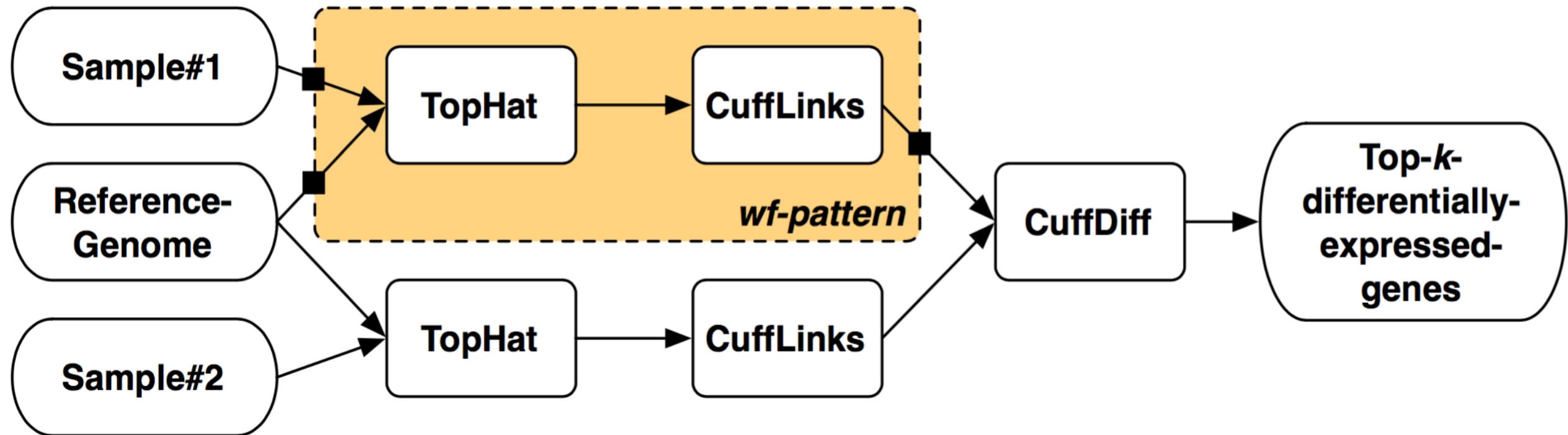
RNA-seq data analysis workflow



TopHat	1 sample	300 samples
Input data	2 x 17 Gb	10.2 Tb
1-core CPU	170 hours	5.9 years
32-cores CPU	32 hours	14 months
Output data	12 Gb	3.6 Tb

**Compute and
storage intensive**

RNA-seq data analysis workflow



TopHat	1 sample	300 samples
Input data	2 x 17 Gb	10.2 Tb
1-core CPU	170 hours	5.9 years
32-cores CPU	32 hours	14 months
Output data	12 Gb	3.6 Tb

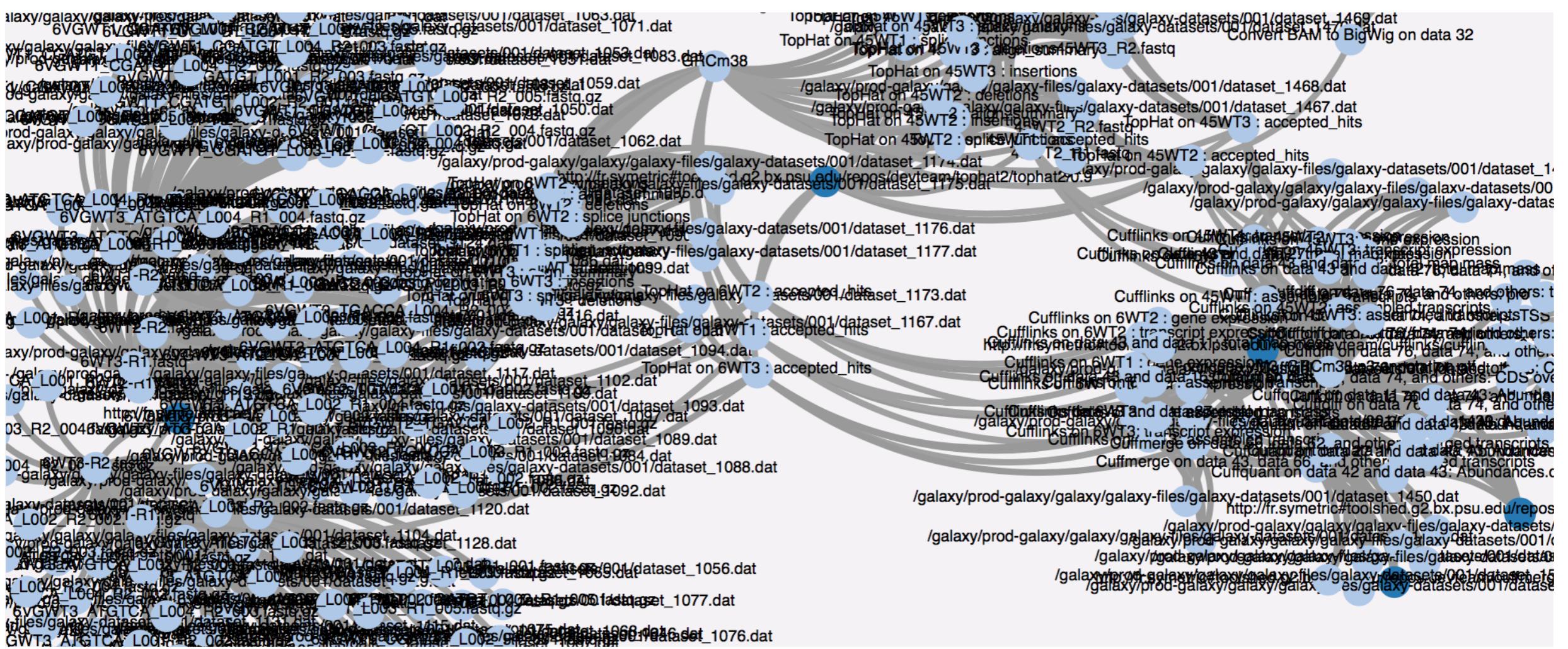
**Compute and
storage intensive**

**Avoid duplicated
storage / computing**

Is provenance enough for reuse ?

```
11     a prov:Bundle, prov:Entity;
12     prov:wasAttributedTo <#galaxy2prov>;
13     prov:generatedAtTime "2016-04-14T18:18:37.000409"^^xsd:dateTime;
14 .
15
16 <#72486b583fe152f0>
17     a prov:Activity ;
18     prov:wasAssociatedWith <#cat1> ;
19     prov:startedAtTime "2015-12-15T12:54:50.749845"^^xsd:dateTime;
20     prov:endedAtTime "2015-12-15T12:55:57.016799"^^xsd:dateTime;
```

Visualise



Is provenance enough for reuse ?

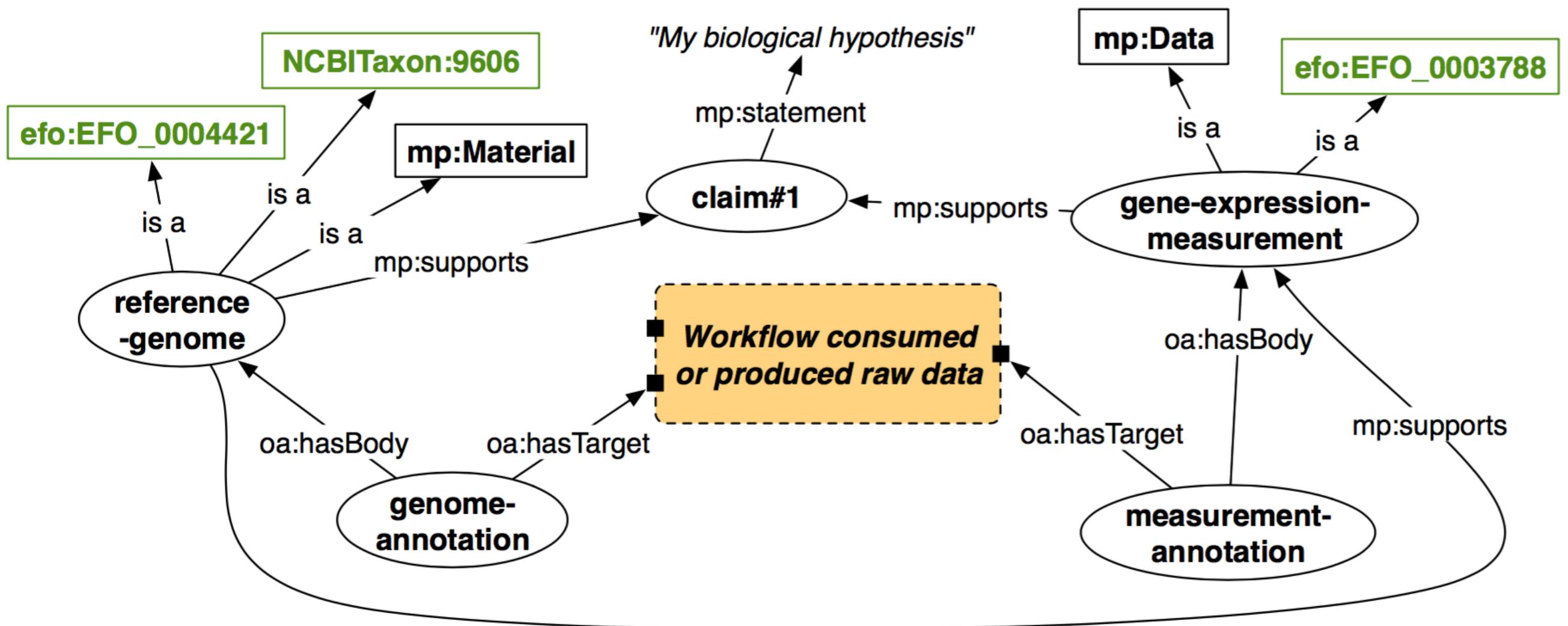
```
11     a prov:Bundle, prov:Entity;
12     prov:wasAttributedTo <#galaxy2prov>;
13     prov:generatedAtTime "2016-04-14T18:18:37.000409"^^xsd:dateTime;
14 .
15 .
16 <#72486b583fe152f0>
17     a prov:Activity ;
18     prov:wasAssociatedWith <#cat1> ;
19     prov:startedAtTime "2015-12-15T12:54:50.749845"^^xsd:dateTime;
20     prov:endedAtTime "2015-12-15T12:55:57.016799"^^xsd:dateTime;
```

Visualise



Human & machine-tractable report needed !

Annotated paper's "**Material & Methods**" with links to **some** workflow artifacts (algorithms, data).



Problem statement & objectives

Problem statement

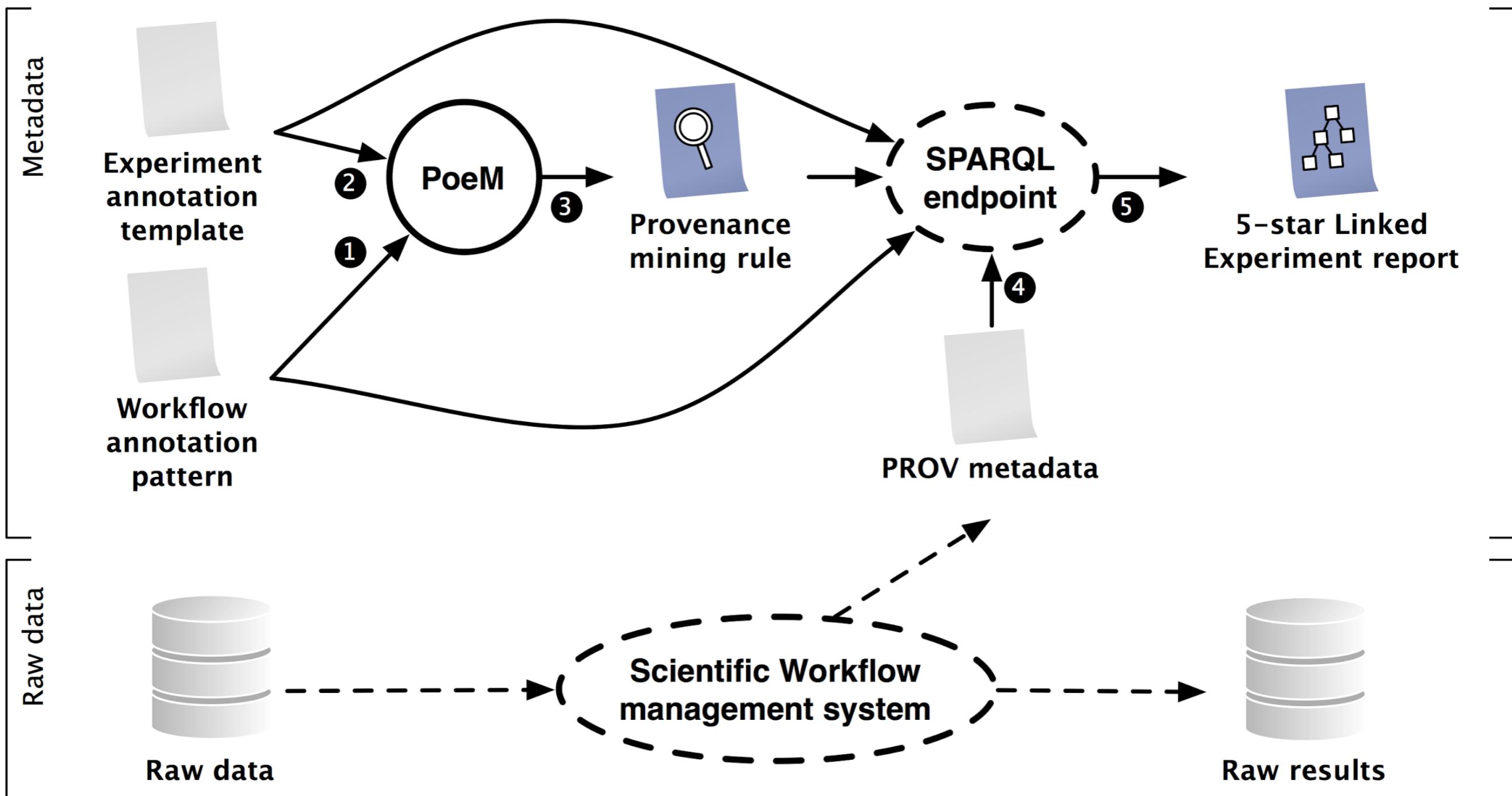
Scientific workflows produce massive raw results. Their publication into curated query-able linked data repositories requires lot of time and expertise.

Can we exploit provenance traces to ease the publication of scientific results as Linked Data ?

Objectives

- (1) Leverage annotated workflow patterns to generate **provenance mining rules**.
- (2) Refine provenance traces into **linked experiment reports**.

Approach



PoeM: generating PrOvEnhance Mining rules ③

Input : W : Workflow annotated pattern ①,
 S_1 : First step of W ,
 S_2 : Last step of W ,
 A : Annotation template ②.

Output: $Rule$: Provenance mining rule.

```
1 begin
2    $IN_{S1} \leftarrow getInputs(S_1)$ 
3    $OUT_{S2} \leftarrow getOutputs(S_2)$ 
4
5    $provGraph \leftarrow genDataLineage(OUT_{S2}, IN_{S1})$ 
6    $reportGraph \leftarrow bindReportTargets(provGraph, A)$ 
7
8    $Rule \leftarrow \frac{provGraph.edge_1 \wedge \dots \wedge provGraph.edge_N}{reportGraph}$ 
```

SPARQL Property path
SPARQL Basic graph pattern
SPARQL Construct query

Demo



PoeM: Provenance Mining for scientific linked data publishing

Contact : Alban Gaignard, Hala Skaf, Audrey Bihouee, {firstname.lastname}@univ-nantes.fr.

For demonstration purpose only.

This web page illustrates our semi-automated approach for mining provenance traces and assembling linked experiment reports. PoeM generates Semantic Web rules from (i) annotated workflow patterns, (ii) domain-specific annotations, and (iii) provenance traces of a workflow run. The rules finally match provenance subgraphs and produce linked experiment reports.

We illustrate our approach in the context of an RNAseq bioinformatics workflow.

This work reuses existing linked open vocabularies, namely PROV-O, P-PLAN, Micropublications Ontology, Experimental Factors Ontology, and EDAM.

This demo is supported by the [Corese](#) Semantic Web factory, [Apache Jena](#), [D3.js](#), [Codemirror.js](#) and [Twitter Bootstrap](#).

Show input annotations

Show provenance trace

Show generated rule

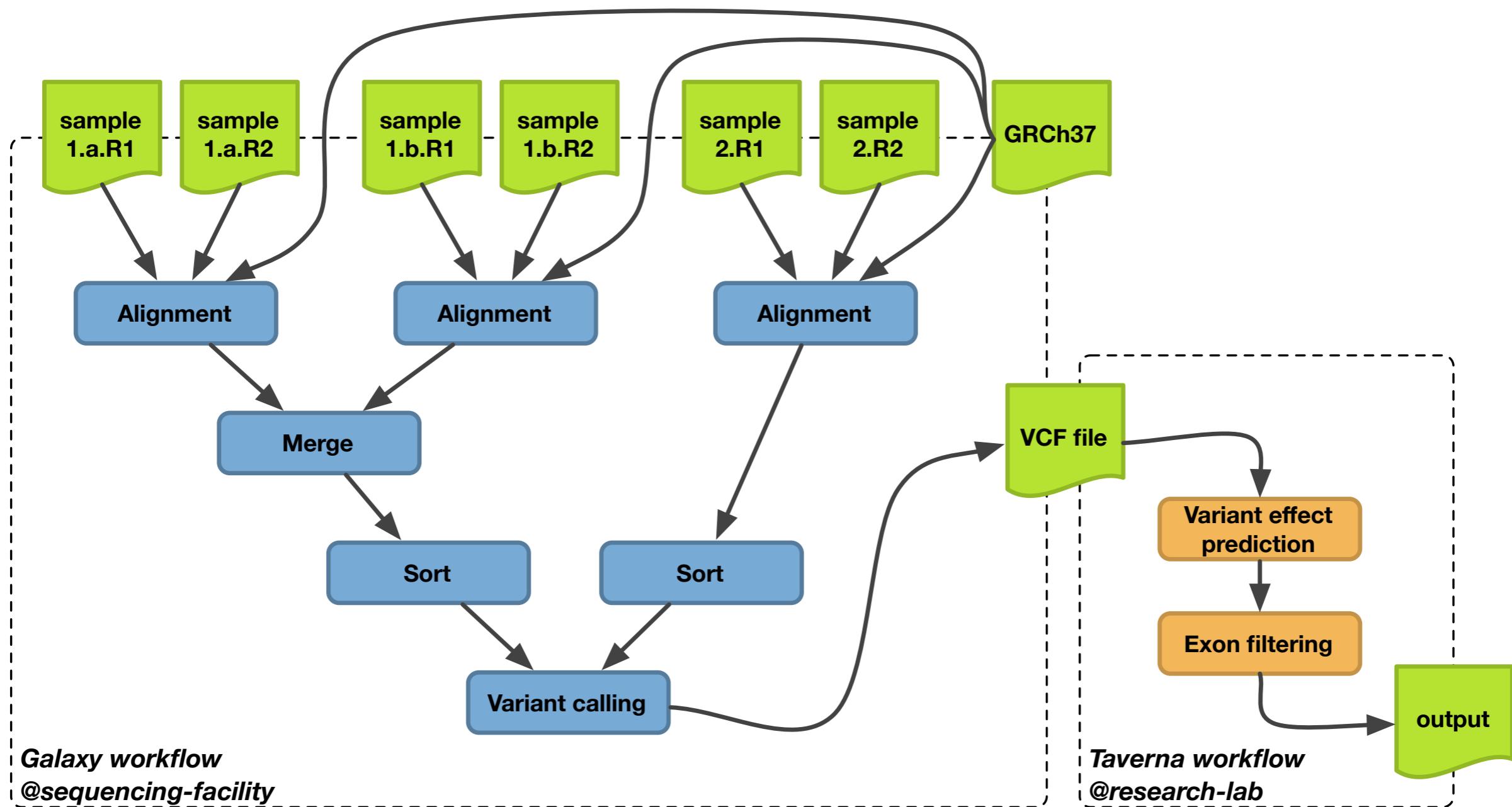
Show resulting report

Resulting linked experiment report

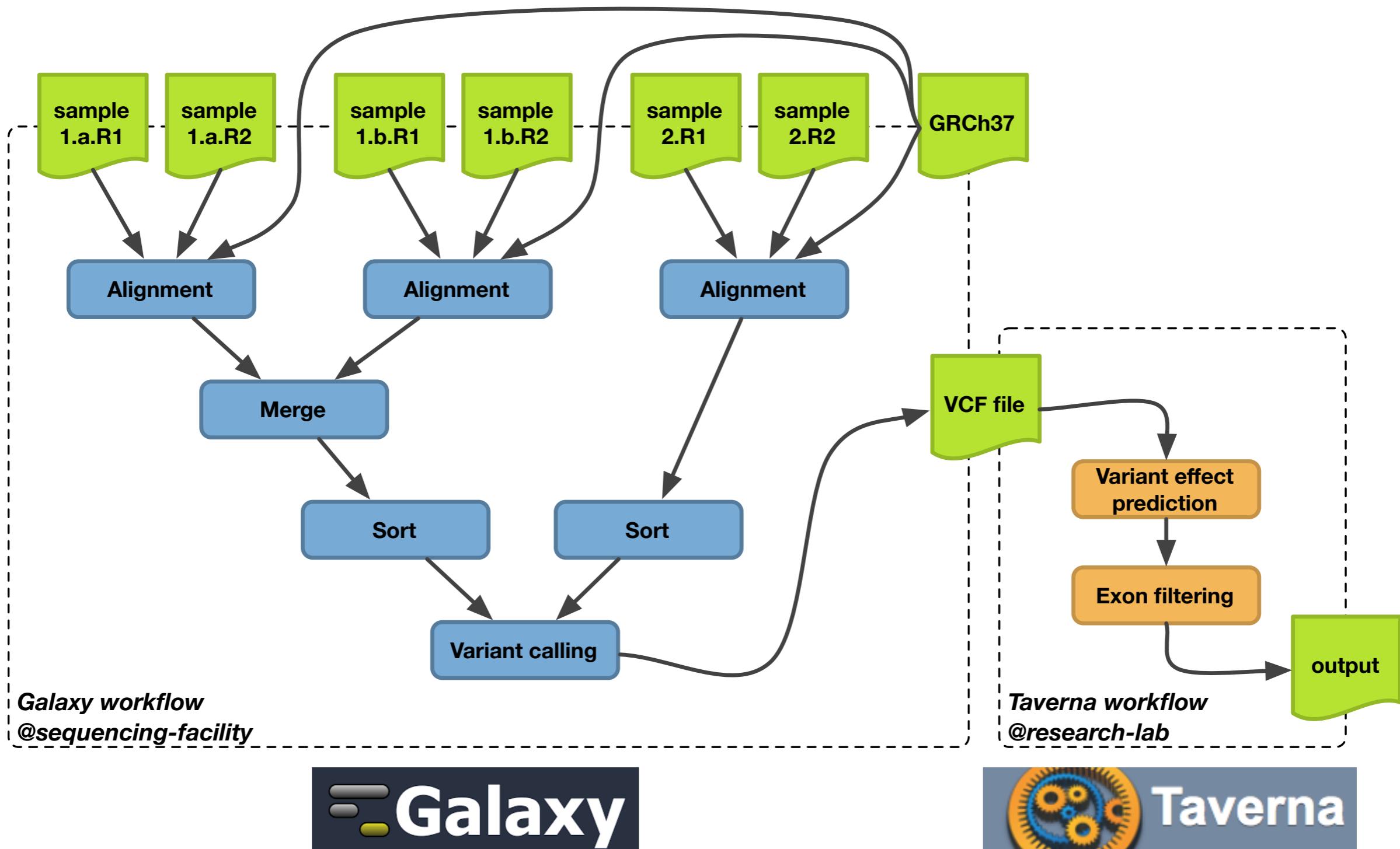
Provenance
in **multi-site** studies ?

Multi-site studies → ≠ workflow engines !

Multi-site studies → ≠ workflow engines !

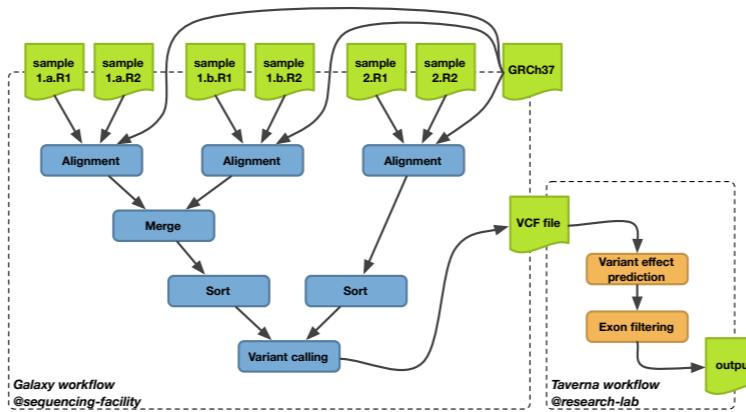


Multi-site studies → ≠ workflow engines !



Scattered provenance capture ?

Provenance issues



« Which alignment algorithm was used when predicting these effects ? »

« A new version of a reference genome is available, which genome was used when predicting these phenotypes ? »

Need for an overall tracking of provenance over both Galaxy and Taverna workflows !

Provenance « heterogeneity »

Galaxy PROV predicates	counts
prov:wasDerivedFrom	118
rdf:type	76
rdfs:label	62
prov:used	61
prov:wasAttributedTo	34
prov:wasGeneratedBy	33
prov:endedAtTime	26
prov:startedAtTime	26
prov:wasAssociatedWith	26
prov:generatedAtTime	1

Taverna PROV predicates	counts
rdf:type	54
rdfs:label	13
prov:atTime	8
wfprov:describedByParameter	6
rdfs:comment	6
prov:hadRole	6
prov:activity	5
dcterms:hasPart	4
prov:agent	4
prov:endedAtTime	4
prov:hadPlan	4
prov:qualifiedAssociation	4
prov:qualifiedEnd	4
prov:qualifiedStart	4
prov:startedAtTime	4
prov:wasAssociatedWith	4
tavernaprov:content	3
wfprov:usedInput	3
wfprov:wasEnactedBy	3
wfprov:wasOutputFrom	3

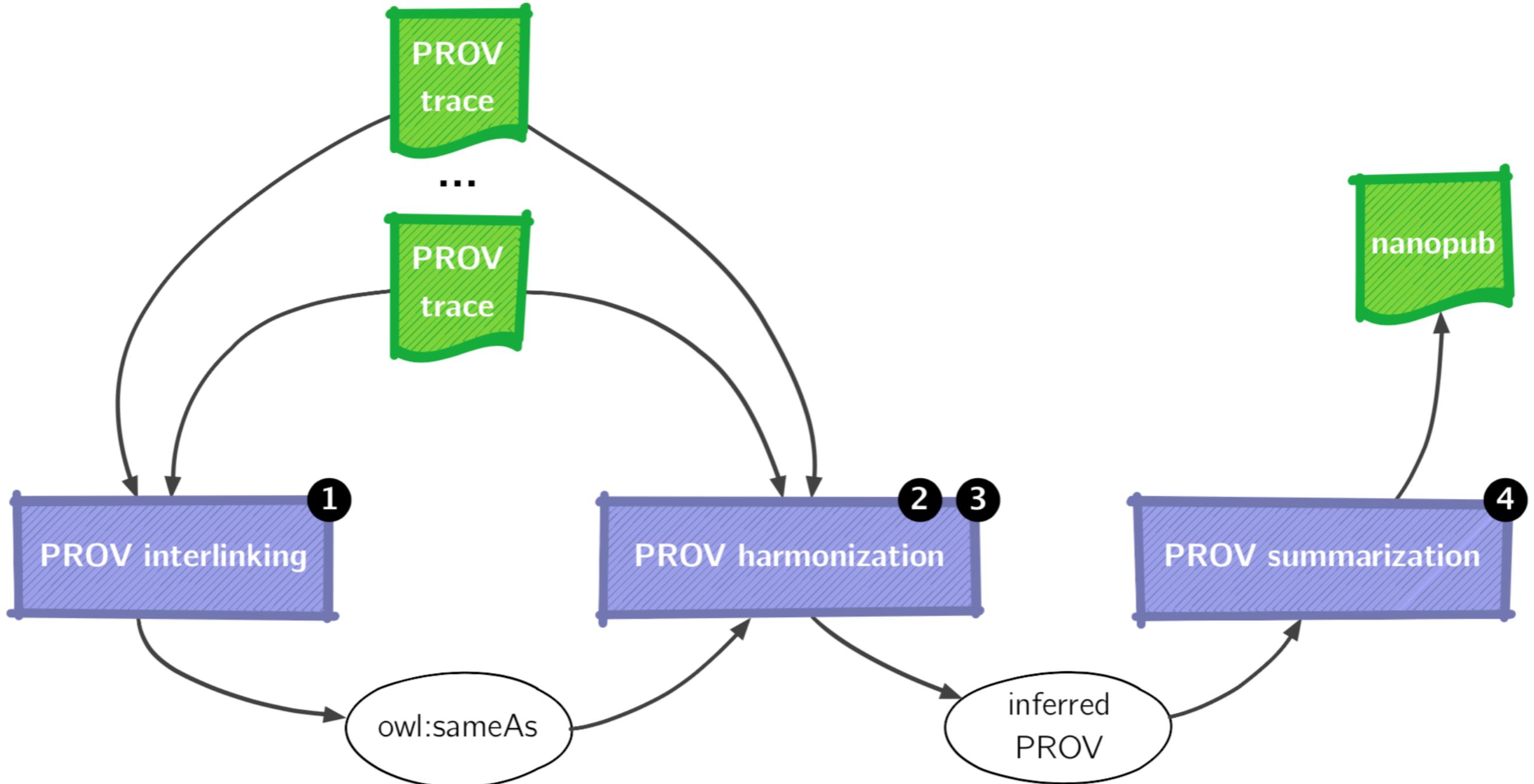
Provenance « heterogeneity »

Galaxy PROV predicates	counts
prov:wasDerivedFrom	118
rdf:type	76
rdfs:label	62
prov:used	61
prov:wasAttributedTo	34
prov:wasGeneratedBy	33
prov:endedAtTime	26
prov:startedAtTime	26
prov:wasAssociatedWith	26
prov:generatedAtTime	1

Taverna PROV predicates	counts
rdf:type	54
rdfs:label	13
prov:atTime	8
wfprov:describedByParameter	6
rdfs:comment	6
prov:hadRole	6
prov:activity	5
dcterms:hasPart	4
prov:agent	4
prov:endedAtTime	4
prov:hadPlan	4
prov:qualifiedAssociation	4
prov:qualifiedEnd	4
prov:qualifiedStart	4
prov:startedAtTime	4
prov:wasAssociatedWith	4
tavernaprov:content	3
wfprov:usedInput	3
wfprov:wasEnactedBy	3
wfprov:wasOutputFrom	3

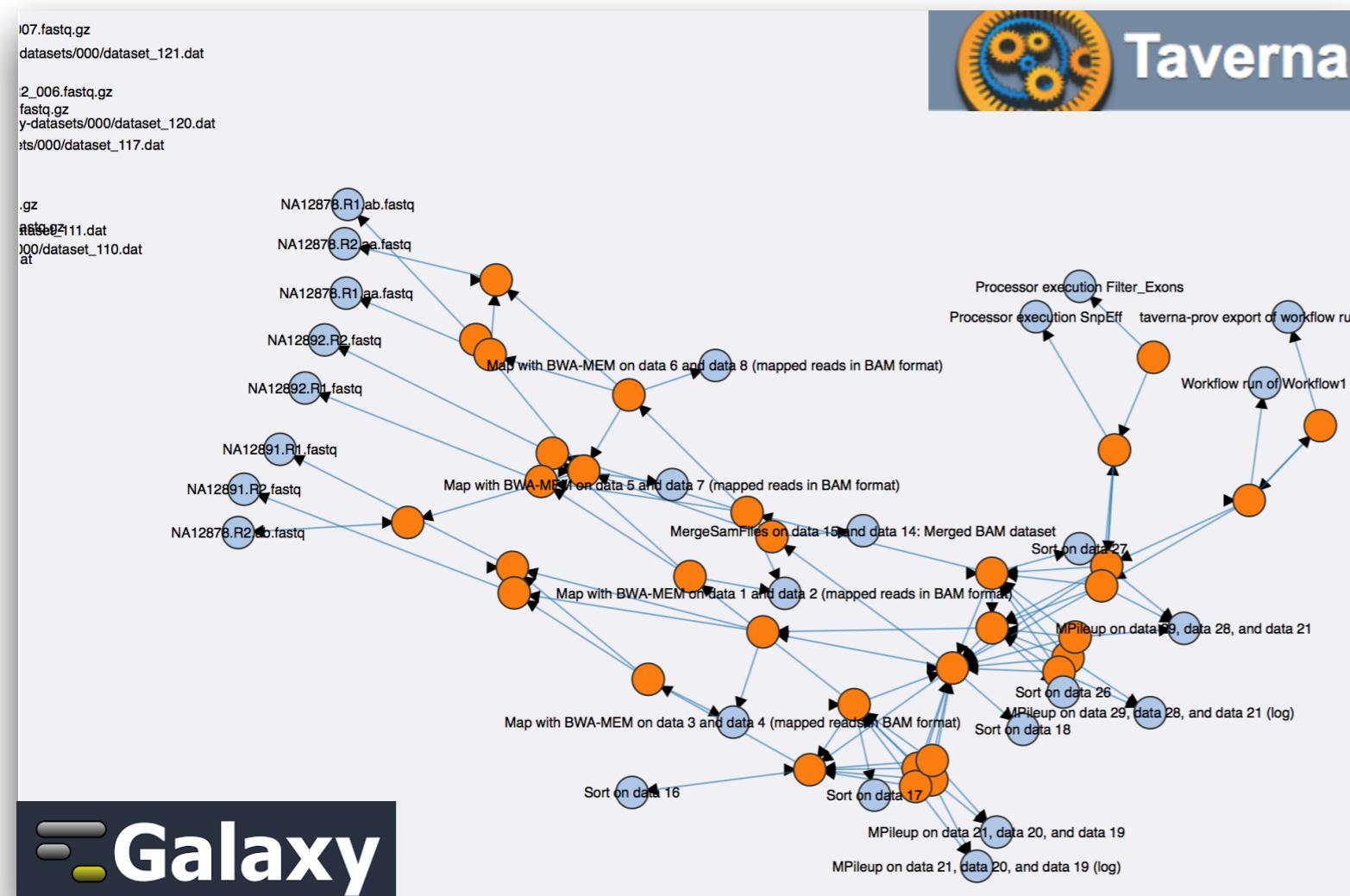
How to reconcile these provenance traces ?

Approach



Results

Results



Reconciled provenance as
an « influence graph »

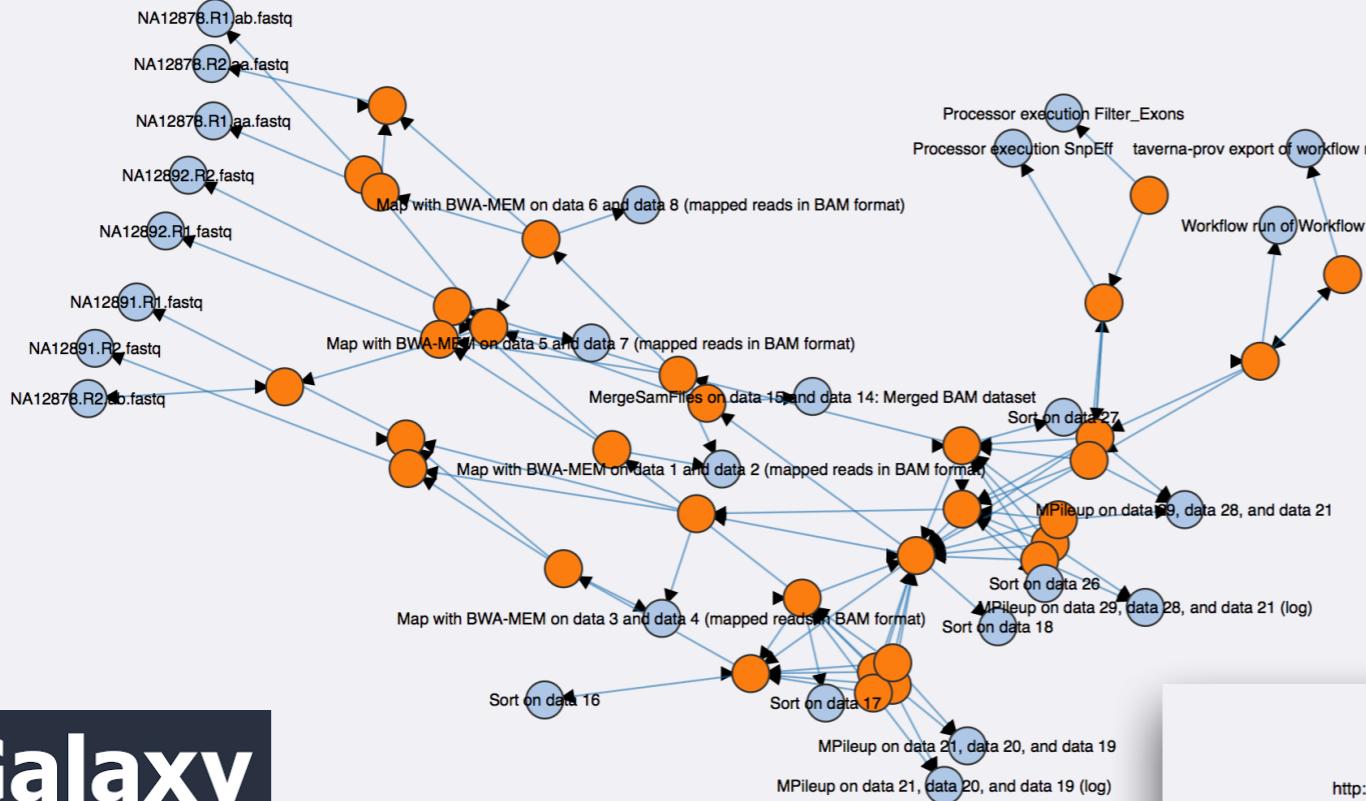
Results

```
007.fastq.gz  
datasets/000/dataset_121.dat  
  
12_006.fastq.gz  
fastq.gz  
y-datasets/000/dataset_120.dat  
ets/000/dataset_117.dat
```

```
.gz  
00902111.dat  
000/dataset_110.dat  
at  
  
NA12878.R1.ab.fastq  
NA12878.R2.aa.fastq  
NA12878.R1.aa.fastq  
NA12892.R2.fastq  
NA12892.R1.fastq  
NA12891.R1.fastq  
NA12891.R2.fastq  
NA12878.R2.ab.fastq
```

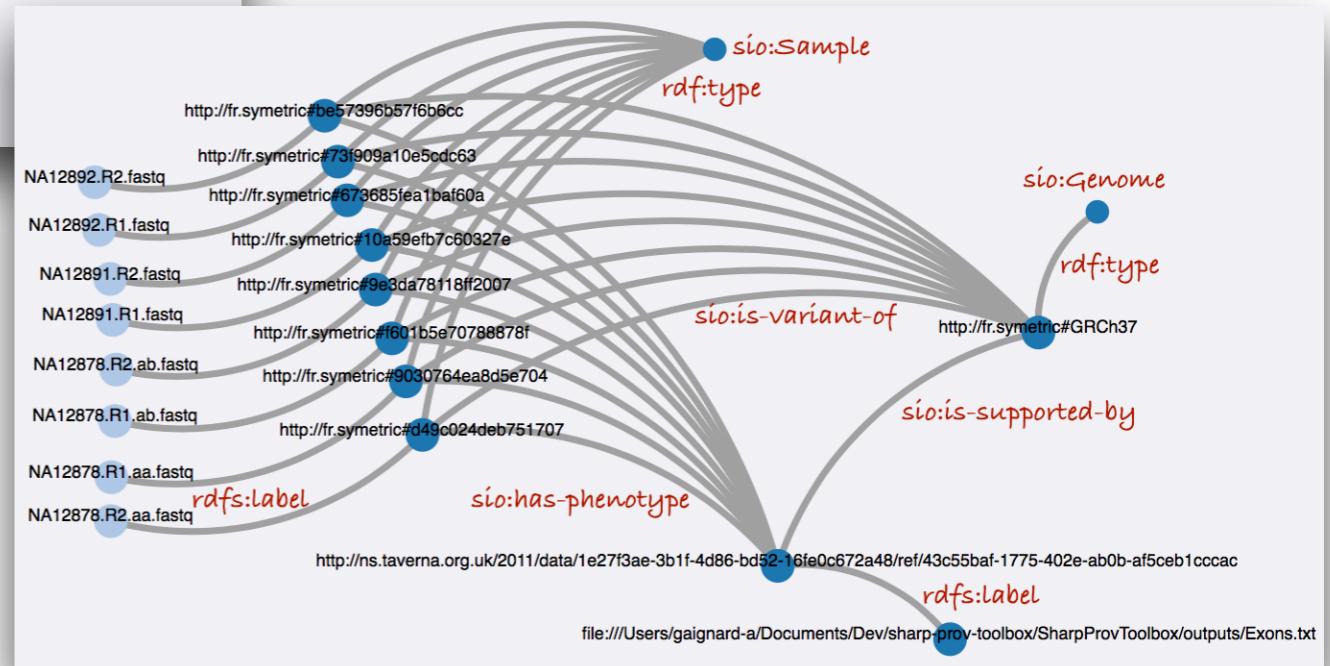


Taverna



Reconciled provenance as
an « influence graph »

Linked experiment report
with Nanopublication,
domain-specific concepts



Summary

Take home message & perspectives

- **Scientific Workflows** → automation, abstraction, provenance
- Standards for **provenance representation** and **reasoning**
- Better handle **multi-site studies** (ESWC'17 satellite event paper)
- Linked experiment reports = **contextualized** and **summarized** provenance (TaPP'16 paper)
- Distributed data analysis → **Distributed provenance, reasoning** ?
- **Learning patterns** in provenance graphs ?
- **Predicting domain-specific annotation** for workflow results ?

Acknowledgments



Audrey Bihouée, Institut du Thorax, BiRD Bioinformatics facility, University of Nantes



Hala Skaf-Molli, LS2N, University of Nantes



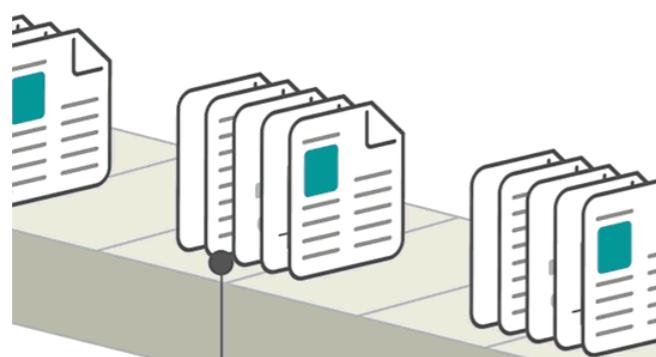
Khalid Belhajjame, LAMSADE, University of Paris-Dauphine, PSL

GDR **MaDiCS**
action **ReProVirtuFlow**

Backup slides

Research Objects

 Enabling **reproducible**, transparent research.



 **scientific**hypothesis

 **PUBLICATIONS**

 **SLIDES**

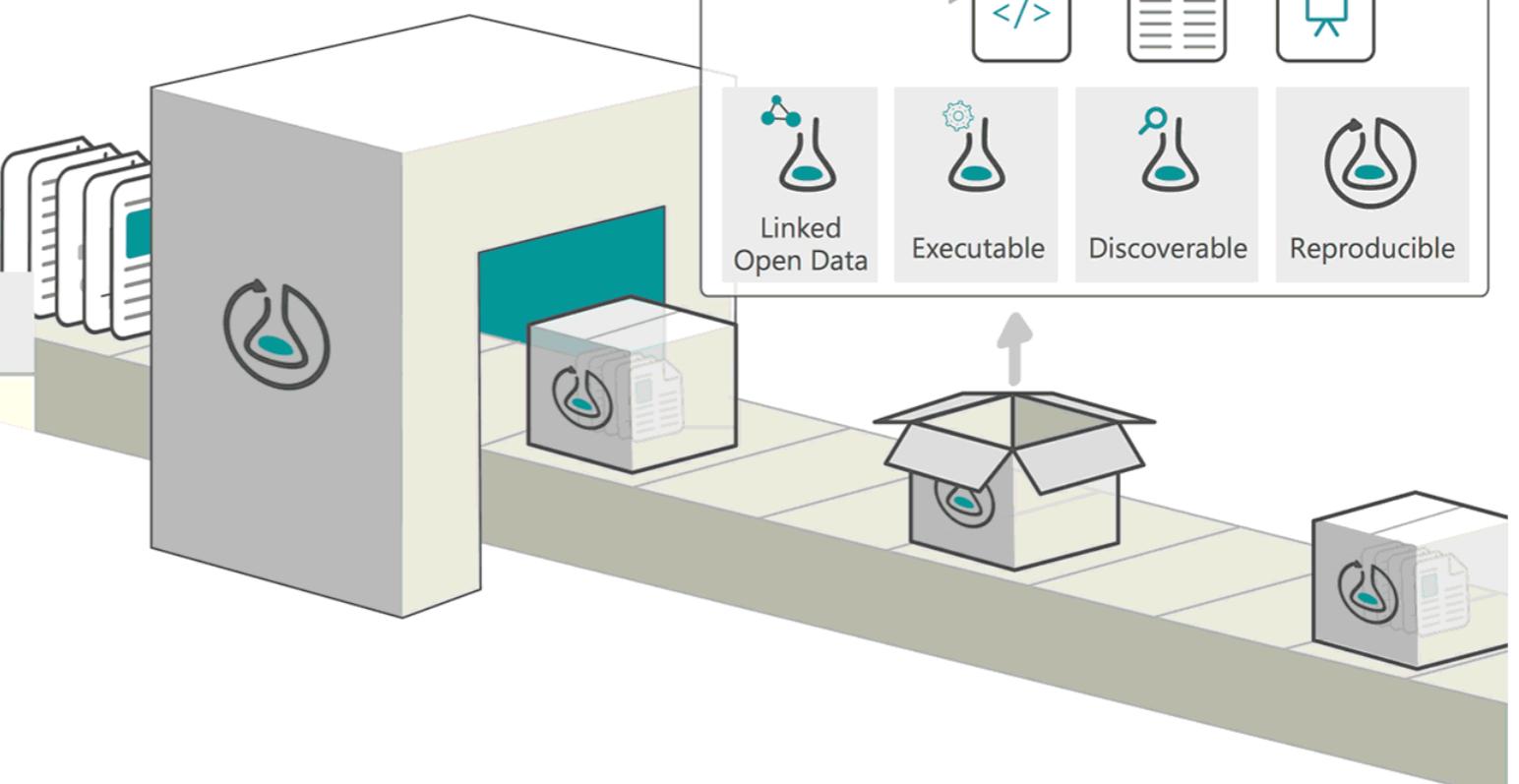
 **DATA**

 **METADATA**

 **RESULTS**

 **LOGS**

 **WORKFLOWS**



Sean Bechhofer, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Phillip Couch, Don Cruickshank, Mark Delderfield, Ian Dunlop, Matthew Gamble, Danius Michaelides, Stuart Owen, David Newman, Shoaib Sufi, Carole Goble (2013) **Why Linked Data is Not Enough for Scientists**, Future Generation Computer Systems **29**(2), February 2013, Pages 599-611, ISSN 0167-739X, <https://doi.org/10.1016/j.future.2011.08.004>

Khalid Belhajjame, Jun Zhao, Daniel Garijo, Matthew Gamble, Kristina Hettne, Raul Palma, Eleni Mina, Oscar Corcho, José Manuel Gómez-Pérez, Sean Bechhofer, Graham Klyne, Carole Goble (2015) **Using a suite of ontologies for preserving workflow-centric research objects**, Web Semantics: Science, Services and Agents on the World Wide Web, <https://doi.org/10.1016/j.websem.2015.01.003>

schema.org Action

Action

Canonical URL: <http://schema.org/Action>

Thing > Action

An action performed by a direct agent and indirect participants upon a direct object. Optionally happens at a location with the help of an inanimate instrument. The execution of the action may produce a result. Specific action sub-type documentation specifies the exact expectation of each argument/role.

See also [blog post](#) and [Actions overview document](#).

Usage: Between 100 and 1000 domains

[\[more...\]](#)

Property	Expected Type	Description
Properties from Action		
actionStatus	ActionStatusType	Indicates the current disposition of the Action.
agent	Organization or Person	The direct performer or driver of the action (animate or inanimate). e.g. <i>John</i> wrote a book.
endTime	DateTime	The endTime of something. For a reserved event or service (e.g. FoodEstablishmentReservation), the time that it is expected to end. For actions that span a period of time, when the action was performed. e.g. John wrote a book from January to <i>December</i> .
		Note that Event uses startDate/endDate instead of startTime/endTime, even when describing dates with times. This situation may be clarified in future revisions.

error	Thing	For failed actions, more information on the cause of the failure.
instrument	Thing	The object that helped the agent perform the action. e.g. John wrote a book with <i>a pen</i> .
location	Place or PostalAddress or Text	The location of for example where the event is happening, an organization is located, or where an action takes place.
object	Thing	The object upon which the action is carried out, whose state is kept intact or changed. Also known as the semantic roles patient, affected or undergoer (which change their state) or theme (which doesn't). e.g. John read <i>a book</i> .
participant	Organization or Person	Other co-agents that participated in the action indirectly. e.g. John wrote a book with <i>Steve</i> .
result	Thing	The result produced in the action. e.g. John wrote <i>a book</i> .
startTime	DateTime	The startTime of something. For a reserved event or service (e.g. FoodEstablishmentReservation), the time that it is expected to start. For actions that span a period of time, when the action was performed. e.g. John wrote a book from <i>January</i> to December.
target	EntryPoint	Note that Event uses startDate/endDate instead of startTime/endTime, even when describing dates with times. This situation may be clarified in future revisions. Indicates a target EntryPoint for an Action.