

NetKarma: tool for obtaining a provenance-based record of experimentation

Beth Plale, School of Informatics and Computing

Chris Small, Global Research Network Operations Center

Indiana University

Provenance of art works



- Trace of history of work of art from moment it was made until it comes into a collection.
- Impartial and authoritative information on authenticity, ownership, theft, and other artistic, legal, and ethical issues concerning art objects.



It can be impossible to determine complete provenance for work of art:

- provenance records often reflect an owner or former owner's wish for anonymity.
- ephemeral nature of historical records, which are often lost or destroyed over time

Saint John the Baptist in the Wilderness;
circa 1535 Moretto da Brescia Italy,
Brescia 1498–1554



Provenance:

Oskar Bondy [d. 1944], Vienna¹
[Jacob M. Heimann, Beverly
Hills];² Philip Yordan,
Beverly Hills (until 1951),³
gift 1951; to Los Angeles
County Museum of Art

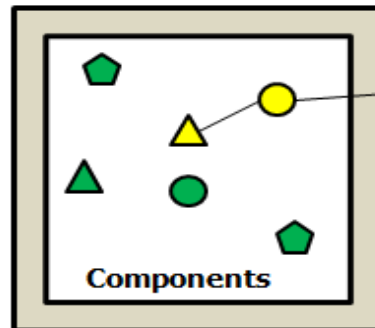
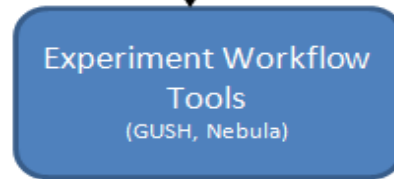
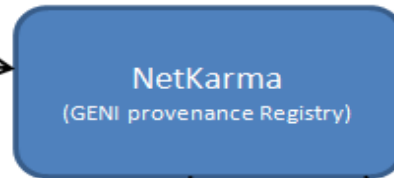
- The ***GENI Provenance Registry (NetKarma) project***,
 - funded in Spiral 2,
 - tool for capturing and representing the workflow of GENI slice creation, topology of the slice, operational status and other measurement statistics and correlate it with remaining experimental data.

- The ***provenance of an experiment***
 - ***is its lineage or historical trace that can capture*** experiment conditions, time ordering, and relationships within the experiment and across the application and infrastructure layer.

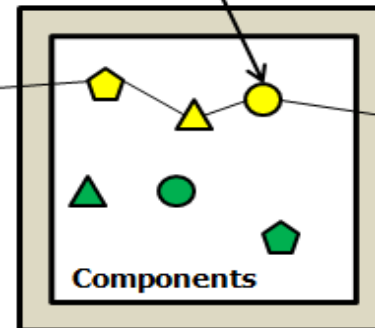
Use scenario



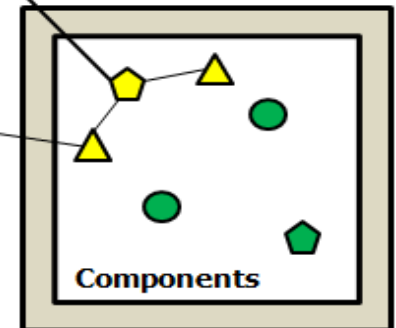
User interacts with NetKarma using a portal which gives access to various provenance information.



Aggregate A
Compute Nodes



Aggregate B
Storage Nodes



Aggregate C
Backbone Net

Provenance Query

- A. Get Topology and Slice Creation process messages.
- B. Get Experiment Workflow and Data
- C. Get Data from GIMS
- D. Researcher can push data directly from slice.
- E. Data is "curated" and pushed to permanent storage.

Forms of provenance capture

User Annotation	Scavenging	Full Provenance Instrumentation
Low application burden	Low application burden	High application burden
High human burden	Low human burden	Low human burden (high developer setup cost)
Error rates and omissions can be high leading to incomplete information	Incomplete information	Complete information

Project approach

NetKarma will use a scavenging approach to information collection (push or pull).

Collect from different data sources to obtain information describing GENI experiments and the conditions of the experiment. These include:

- Year 1 focus: the experiment tools.
- Year 2 focus: Control plane and Measurement Plane layers, and GMOC.

For instance:

- Topology created using the control frameworks
- Experiment tool commands
- Operational status on substrate/infrastructure
- Code and data contained in the experimental slice
- Measurement data obtained from the GENI
- Annotations by experimenters

Provenance Views

Life Science Grid

File Applications Tools Help

GeneBrowser

Search Reset

Entrez ID	Gene Symbol	Gene Name	Chromosome
7677	ZNF123	zinc finger p...	11q
7678	ZNF124	zinc finger p...	1q4
7679	ZNF125	zinc finger p...	11q
7680	ZNF126	zinc finger p...	11q
7681	MKRNP3	makorin, ring...	15q
7682	MKRNP5	makorin, ring...	Xp2
7683	MKRNP1	makorin, ring...	Xq1
7684	MKRNP5	makorin, ring...	12q
7685	MKRNP4	makorin, ring...	1p3
7686	MKRNP3	makorin, ring...	20q
7690	ZNF131	zinc finger p...	5p1
7691	ZNF132	zinc finger p...	19q
7692	ZNF133	zinc finger p...	20p
7693	ZNF134	zinc finger p...	19q
7694	ZNF135	zinc finger p...	19q
7695	ZNF136	zinc finger p...	19q
7696	ZNF137	zinc finger p...	19q
7697	ZNF138	zinc finger p...	7q1

GeneOntology RDFViewer

Visualize Refresh

```

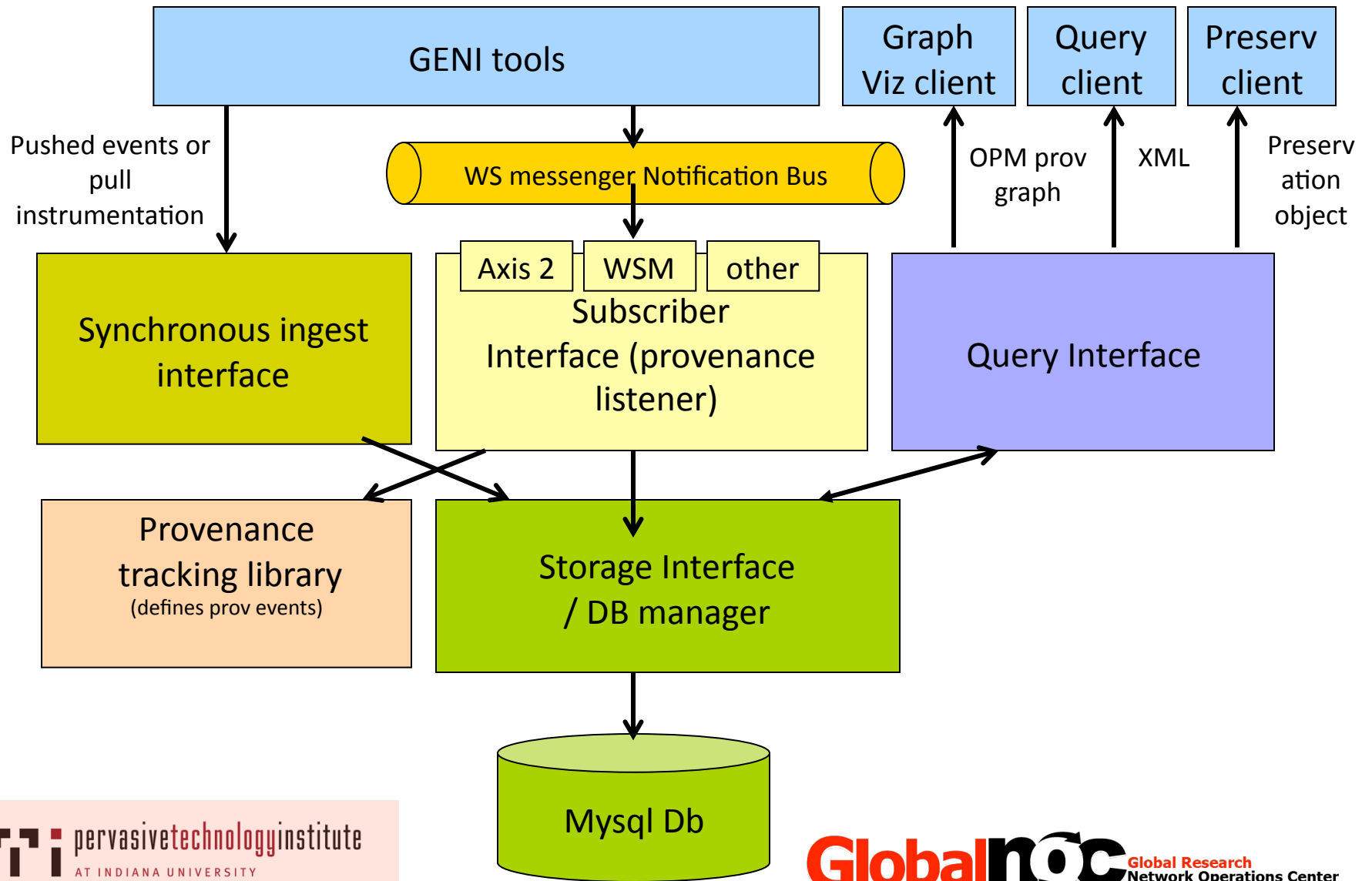
graph TD
    A(Entrez_genbank_protein_record) -- wasGeneratedBy --> B[NCBI_Entrez_Plugin]
    A -- wasGeneratedBy --> C[NCBI_Web_Entrez_GeneLookup_Service]
    D(gene_ontology_term_record) -- wasGeneratedBy --> E[Gene_ontology_GO]
    D -- wasGeneratedBy --> F[Gene_Ontology_Plugin]
    B -- used --> G(Entrez_Gene_ID)
    C -- used --> G
    E -- used --> H(Gene_Ontology_term_id)
    F -- used --> H
    G -- wasTriggeredBy --> I[Gene_Browser_Plugin]
    H -- wasTriggeredBy --> I
    I -- wasGeneratedBy --> J(Gene_Data.csv)
  
```

NCBIEntrez

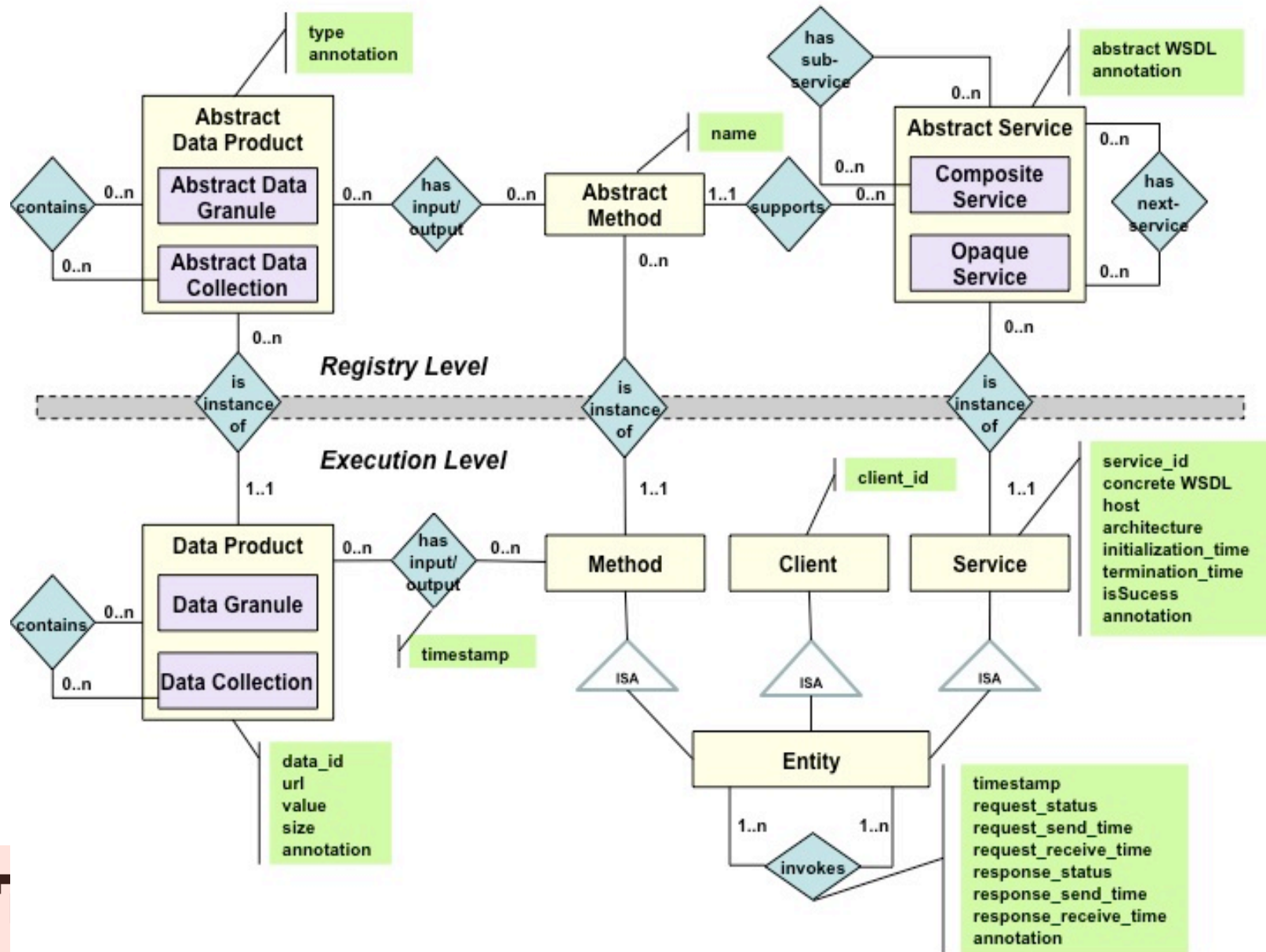
Gene Symbol	MKRNP5 [Entrez Gene Id: 7682]
Species	Homo sapiens
Synonyms	MKRNP4 RNF64 ZNF-Xp ZNF127L1 ZNF127-Xp
Description	makorin ring finger protein pseudogene 5
Summary	
Chromosome	Xp21.1

Girish localhost jobs

netKarma architecture

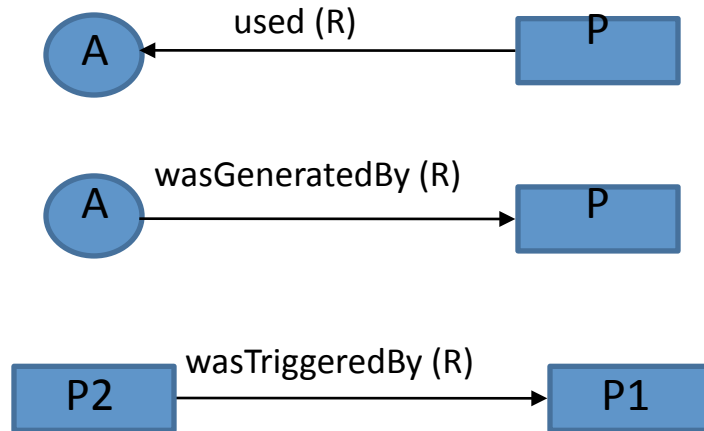


Data model for provenance representation



Provenance Representation for exchange

- Standard representation of simplest provenance graph
- Open Provenance Model (OPM)
- Entities
 - Artifact: an immutable piece of state
 - Process: action resulting in new artifact
- Explicit relationships
 - used
 - wasGeneratedBy
 - wasTriggeredBy



Value to Experiment Owner

- Workflow description
- Understanding your own data
- Sharing: making your data understandable to others
- Handles (DOI)
- Reproduceability