IN34B-03
From Observation to Impacts: Provenance for Earth Science Resources

IN34B: Semantic Web and Provenance: Distributed Earth Science Resources in the Data Life Cycle II

AGU 2014

Wednesday, 17 December 2014

Earth Science Provenance Ontology (PROV-ES)

OVERVIEW

Reproducible Science

Reproducibility

Scientific Verifiability

Understandability

Collaboration

Provenance

Transparency

Motivation

- Improve their **understanding and trust** of the science results.
- Provide more **transparency** and understanding in the science process
- Need a formal recommendation for an **interoperable standard for provenance** representation for use in Earth Science Data Systems.

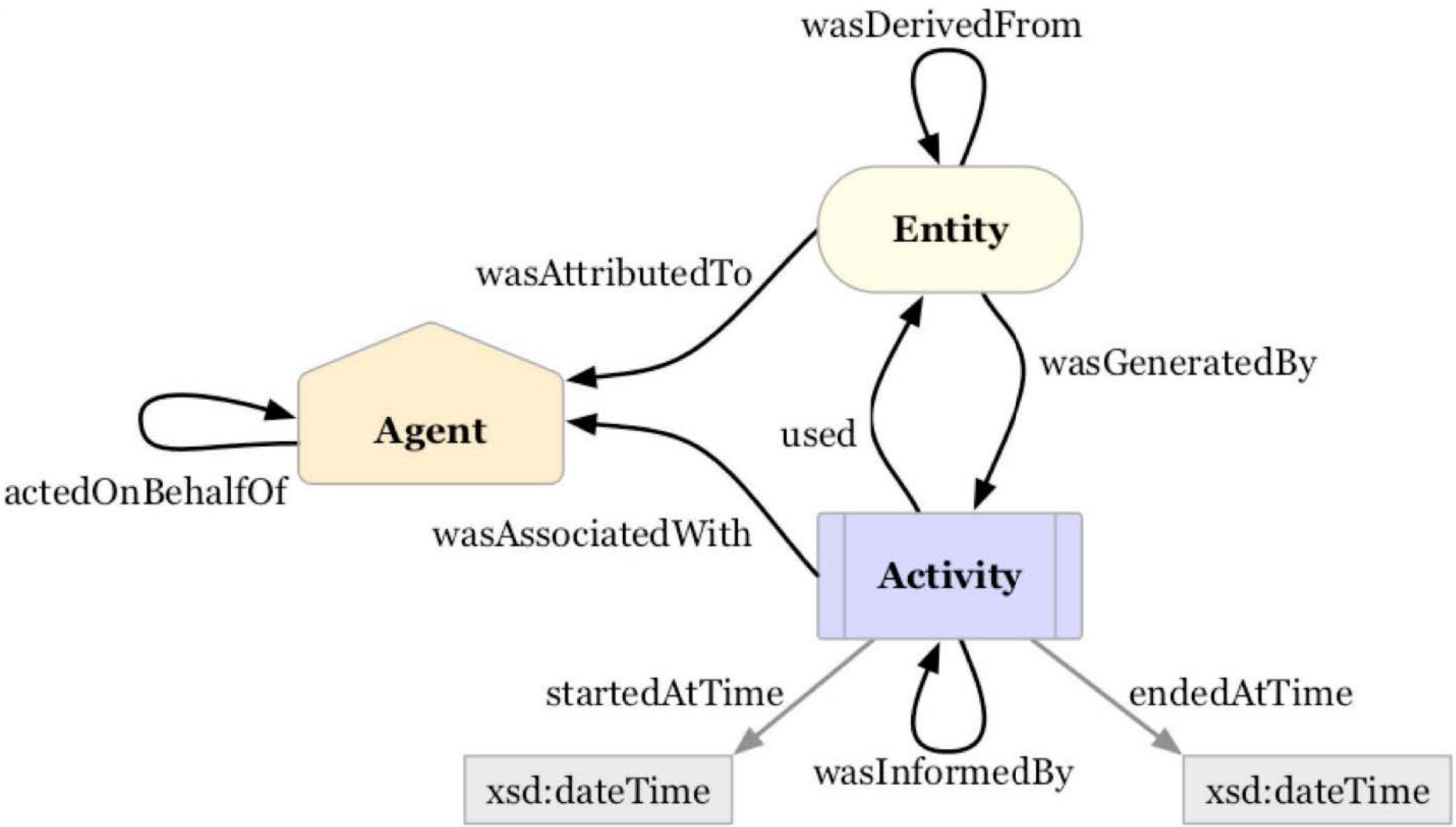
Provenance Recommendation

W3C PROV

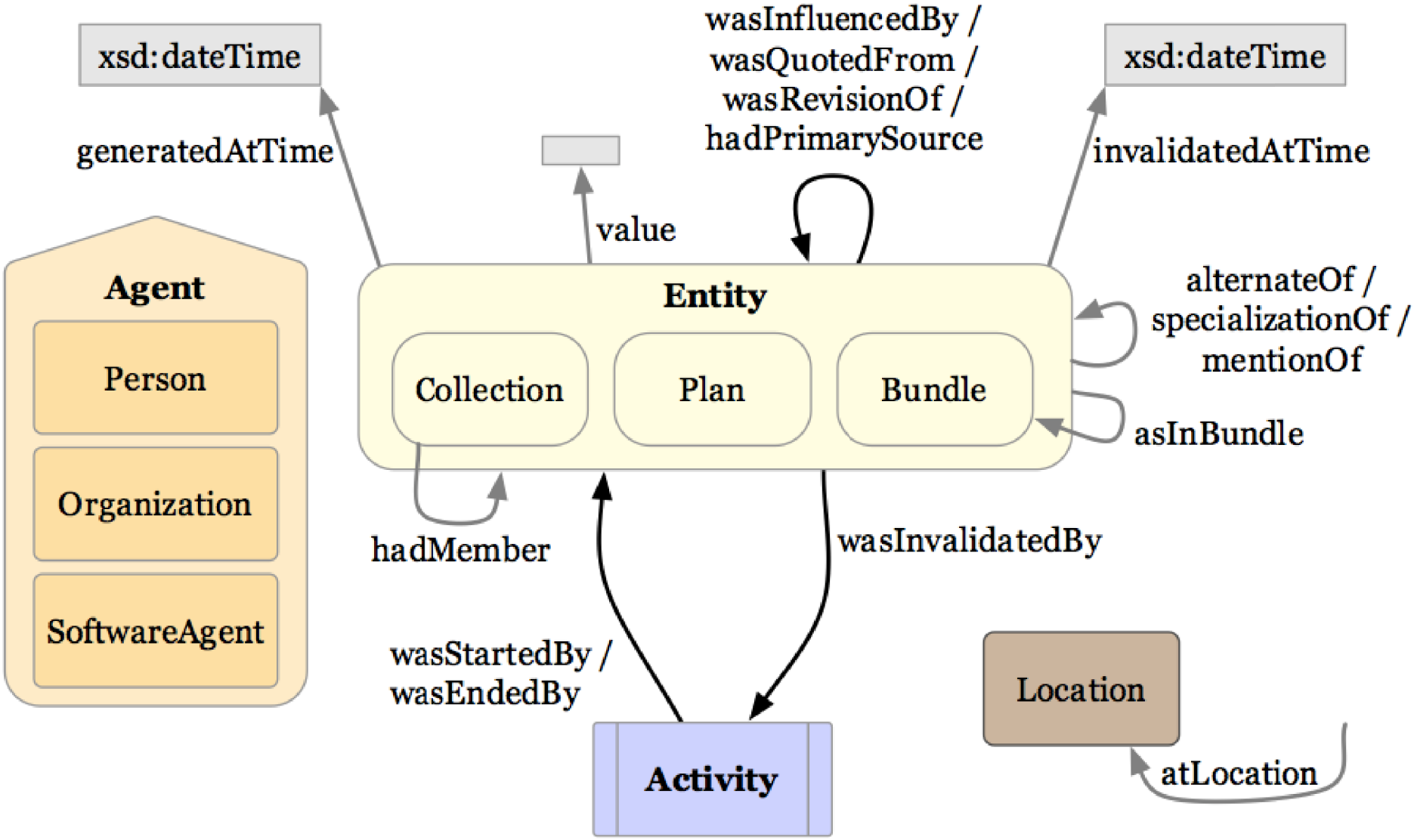
W3C PROV Family

- Model
  - PROV-DM (Recommendation)
    - PROV data model for provenance
  - PROV-CONSTRAINTS (Recommendation)
    - a set of constraints applying to the PROV data model
  - PROV-SEM (Note)
    - declarative specification in terms of first-order logic of the PROV data model
  - PROV-LINKS (Note)
    - mechanism to link across bundles
- Encodings
  - PROV-N (Recommendation)
    - Notation for provenance aimed at human consumption
  - PROV-O (Recommendation)
    - PROV ontology, an OWL2 ontology allowing the mapping of the PROV data model to RDF
  - PROV-DC (Note)
    - mapping between PROV-O and Dublin Core Terms
  - PROV-XML (Note)
    - XML schema for the PROV data model
  - PROV-JSON
- Access
  - PROV-AQ (Note)

W3C PROV: Standard Terms



W3C PROV: Expanded Terms



PROV-DM Cross-References

Table 10 ◊: Cross-References to PROV-O and PROV-N

PROV-DM	PROV-O	PROV-N	Component
Entity	Entity	entityExpression	Component 1: Entities/Activities
Activity	Activity	activityExpression	
Generation	wasGeneratedBy, Generation	generationExpression	
Usage	used, Usage	usageExpression	
Communication	wasInformedBy, Communication	communicationExpression	
Start	wasStartedBy, Start	startExpression	
End	wasEndedBy, End	endExpression	
Invalidation	wasInvalidatedBy, Invalidation	invalidationExpression	
Derivation	wasDerivedFrom, Derivation	derivationExpression	Component 2: Derivations
Revision	wasRevisionOf, Revision	type Revision	
Quotation	wasQuotedFrom, Quotation	type Quotation	
Primary Source	hadPrimarySource, PrimarySource	type PrimarySource	
Agent	Agent	agentExpression	Component 3: Agents, Responsibility, Influence
Attribution	wasAttributedTo, Attribution	attributionExpression	
Association	wasAssociatedWith, Association	associationExpression	
Delegation	actedOnBehalfOf, Delegation	delegationExpression	
Plan	Plan	type Plan	
Person	Person	type Person	
Organization	Organization	type Organization	
SoftwareAgent	SoftwareAgent	type SoftwareAgent	
Influence	wasInfluencedBy, Influence	influenceExpression	
Bundle constructor	bundle description	bundle	Component 4: Bundles
Bundle type	Bundle	type Bundle	
Alternate	alternateOf	alternateExpression	Component 5: Alternate
Specialization	specializationOf	specializationExpression	
Collection	Collection	type Collection	Component 6: Collections
EmptyCollection	EmptyCollection	type EmptyCollection	
Membership	hadMember	membershipExpression	

PROV-N Example

- Provenance Notation
  - a notation for provenance aimed at human consumption
  - http://www.w3.org/TR/prov-n/
- *"An entity e1"*
  - `entity(e1)`
- *"An activity a2, which occurred between 2011-11-16T16:00:00 and 2011-11-16T16:00:01".*
  - `activity(a2, 2011-11-16T16:00:00, 2011-11-16T16:00:01)`
- *"Entity e2 was derived from entity e1"*
  - `wasDerivedFrom(e2, e1)`
- Optional attributes
  - `used(ex:id1; ex:act2, ex:e1, 2011-11-16T16:00:00, [ex:fct="load"])`

ESDSWG Provenance for Earth Science Working Group

PROV-ES

Provenance for Earth Science (PROV-ES) Data Systems

- Started as an ESDSWG 2013 Working Group
- *Want to look at it from a broader ESIP perspective*

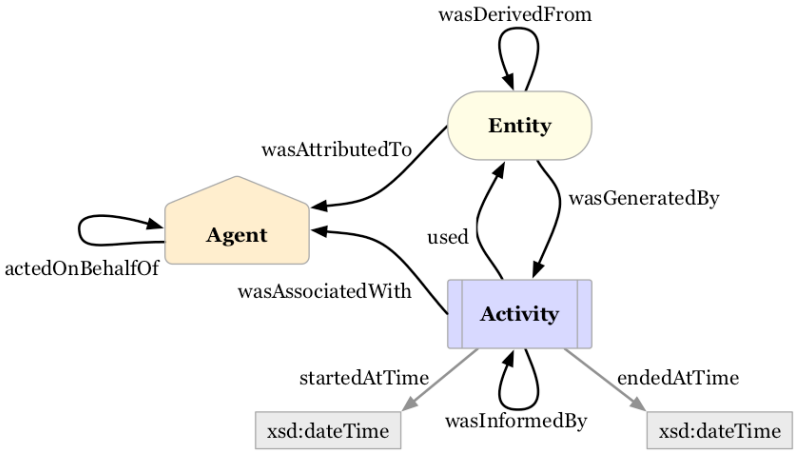
- Mission Statement
  - We will create an interoperable provenance specification for use in Earth Science Data Systems to enable:
    - Capturing the increasing amount of contextual processing information of Earth Science Data Records (ESDRs).
    - Improving transparency, understanding, and trust of ESDRs .
    - Providing an interoperable representation of provenance for NASA EOS data products.
    - Reducing the barrier of entry to infusing provenance capture and consumption in Earth Science Data Systems.
- Stakeholders
  - Data Producers - NASA Missions and projects
  - Data Processing Systems – Enable capturing and storing provenance
  - Data Archive Centers – Enable sharing and seamlessly linking of provenance between data centers.
  - Data Users – Consuming provenance of data products.
  - Tool Makers – Enable provenance exploration, visualization, and analysis.

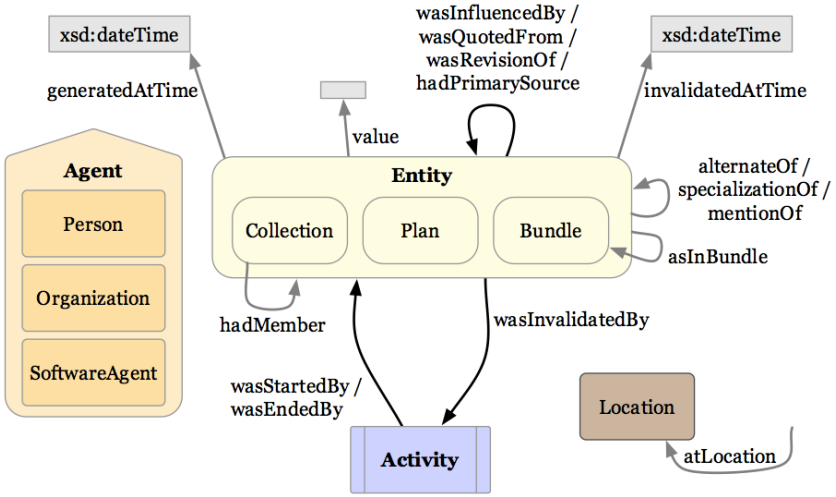
Methodology

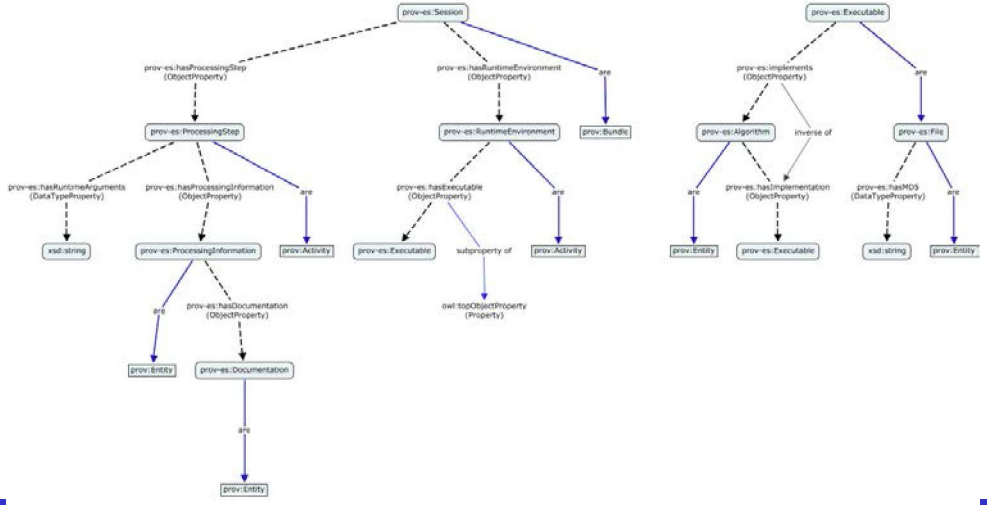
1. Collected relevant use cases
2. Extracted terms/concepts and mapped onto existing W3C PROV
3. Determine frequency counts to see what in W3C PROV is in scope
4. Generate overall taxonomy of concepts based on relevant use cases
5. Map taxonomy back to use cases via properties and relationships of concepts
6. Generate concept map
7. Generate encoding(s)
8. Harmonize GCIS provenance approach with ESDSWG PROV-ES provenance approach
9. Provide mapping of W3C PROV to PROV-ES

Provenance for Earth Science

- ESDSWG PROV-ES Working Group
  - PI Hua is chair of working group
- Leverage W3C international standards
- Encodes PROV Data Model (PROV-DM) in Web Ontology Language (OWL2).
- Good decoupling of provenance models and encodings
- Supports Qualified terms enabling more custom metadata
- Extend W3C PROV-Ontology for modeling Earth Science Data System Processing (PROV-ES)







APPROACH

Earth Science Data Processing Use Cases

- Creating a data set
- Using a data set
- Use Case Result Differences
- Dataset regridding across different instruments
- NASA Earth Science Data Preservation Content Specification (NASA ESD PCS)
- Sea Level Rise: Past, Present and Future: where it came from and why they should believe (trust)
  - Someone reading the NCA sees Figure 1.2, Sea Level Rise: Past, Present and Future, page 21 of the draft and wants to know where it came from and why they should believe (trust) it.
- IceBridge-derived product: Ice sheet modeller using IceBridge-derived bedmap and/or surface DEM
  - A modeller is making use of an IceBridge-derived bedmap and/or surface DEM. Due to the incomplete/sparse coverage of airborne data, they find a discontinuity in the data and need to figure out what **version** of input data were used for this product.
- IceBridge metadata: Archivist finds a problem with spatial metadata and needs to know how the metadata were produced
- Multi-Satellite Ultraviolet Solar Spectral Irradiance Composite.
  - Create an ultraviolet SSI composite using data from SORCE, SVUB/2, GOME, SCIAMACHY, and FISM covering from 1978 until approximately now.
- A Multi-Sensor Water Vapor, Temperature and Cloud Climate Data Record (MEaSUREs)
  - Create a combined data record of clouds, water vapor and atmospheric temperature incorporating all water vapor measurements by A-Train sensors. Combines AIRS/AMSU water vapor and CloudSat cloud observations of the total available water vapor observations in the A-Train swath.
- More examples at https://wiki.earthdata.nasa.gov/display/ESDSWG/Use+Case+Concepts%3A+Provenance+For+Earth+Science

Use Case Mapping to W3C PROV

- Extract core concepts from use case and map to W3C PROV concept

		mapping count -->	7	1	4	4	1	2	2
		base class concept -->			Agent	Agent	Agent		
Use Case	**Reference**	**Description**	**Activity**	**Agent**	**Person**	**Organization**	**SoftwareAgent**	**Entity**	**Bundle**
Creating a data set	http://wiki.esipfed.org/index.php/Preservation_Use_Case_Activity				data producers	citations to organizations	production software		
Using a data set	http://wiki.esipfed.org/index.php/Preservation_Use_Case_Activity								
Use Case Result Differences	http://www.w3.org/2005/Incubator/prov/wiki/Use_Case_Result_Differences		Determine or disambiguate why two processes produced different results.		user B performs a process using D as input			user B downloads data from D using the same query and performs the same process.	bundle captures the provenance of the original workflow. Representing the provenance of data which is derived from large and complex data.
Dataset regridding across different instruments			Gridding in space/time/parameter /etc.						
Identifying differences between "near real-time" and "final" passive microwave gridded data		A user is looking at granules from these 2 data sets, which look really similar, but needs to understand why they both exist, and which one is better for the user's purposes. Ref 423-RQMT-003 NASA Metadata Reqs Base Reference document for terminology used here.							
NASA Earth Science Data Preservation Content Specification (NASA ESD PCS)	http://earthdata.nasa.gov/sites/default/files/field/document/423-SPEC-001_NASA%20ESD_Preservation_Spec_OriginalCh01_0.pdf								
		Preflight/Pre-Operations Calibration							

Use Case Mapping to W3C PROV

- Extend W3C PROV concepts where applicable

		mapping count -->	0	3	2	2	1	3	2	1
		base class concept -->		Activity	Activity	Activity	Activity	Activity	Activity	Activity
Use Case	**Reference**	**Description**		**Data Access**	**Data Discovery**	**Data Analysis**	**Data Sharing**	**Data Comparison**	**Data Verification**	**Data Transformation**
Creating a data set	http://wiki.esipfed.org/index.php/Preservation_Use_Case_Activity									
Using a data set	http://wiki.esipfed.org/index.php/Preservation_Use_Case_Activity			obtaining data	finding the data	Analysis of new version of data set How do the improvements affect our use of the data?	Sharing data set for collaboration	Comparing multiple data sets	Verification of an experiment	
Use Case Result Differences	http://www.w3.org/2005/Incubator/prov/wiki/Use_Case_Result_Differences			user B downloads data from source D				user B compares the two process outputs and notices a difference. Determining the differences between two complex provenance records.		
Dataset regridding across different instruments										Gridding
Identifying differences between "near real-time" and "final" passive microwave gridded data		A user is looking at granules from these 2 data sets, which look really similar, but needs to understand why they both exist, and which one is better for the user's purposes. Ref 423-RQMT-003 NASA Metadata Reqs Base Reference document for terminology used here.								
NASA Earth Science Data Preservation Content Specification (NASA ESD PCS)	http://earthdata.nasa.gov/sites/default/files/field/document/423-SPEC-001_NASA%20ESD_Preservation_Spec_OriginalCh01_0.pdf									
		Preflight/Pre-Operations Calibration								

Concepts and Relationships

- Transpose mapping Taxonomy

					Use Case Result Differences	Dataset regridding across different instruments	Identifying differences between "near real-time" and "final" passive microwave gridded data	NASA Earth Science Data Preservation Content Specification (NASA ESD PCS)	Sea Level Rise: Past, Present and Future: where it came from and why they should believe (trust)	IceBridge-derived product: Ice sheet modeller using IceBridge-derived bedmap and/or surface DEM
Activity								NOTE: Activities can be regarded as "activities needed" as well as "activities enabled". That is, activities needed in order to meet the PCS requirements , and activities enabled if PCS requirements are satisfied.		
	DataAccess							DOI or other unique identifier		
	DataDiscovery							DOI or other unique identifier		
	DataAnalysis							Enabled by PCS (for future investigators)		
	DataSharing							Enabled by PCS (for future investigators)		
	DataComparison						X	Enabled by PCS (for future investigators)		
	DataVerification							Enabled by PCS (for future investigators)		X
	DataTransformation							Enabled by PCS (for future investigators)		
		Geophysical retrieval								
		Data Gridding								X
		Averaging								
	Preflight/Pre-operational Calibration							The activity itself is not part of meeting PCS, but preserving results of the activity is.		
	Calibration (During mission)							The activity itself is not part of meeting PCS, but preserving results of the activity is.		
	Science Data Product Validation							The activity itself is not part of meeting PCS, but preserving results of the activity is.		
	Processing Steps						X	The activity itself is not part of meeting PCS, but preserving results of the activity is.		X
	Planning							preservation planning should be part of overall data management plan		
	Information gathering						X	Information called for in the PCS may be in places other than final destination for post-mission preservation. Ensuring that they are gathered for preservation is a key activity.		X
	Allocation of information storage to organizations							Per Preservation Information Architecture		
Entity										
	Agent									
		Person(s)						wasAttributedTo/wasAssociatedWith Unique IDs needed	x	
		Organization(s)						wasAttributedTo/wasAssociatedWith Unique IDs needed	x	
		Software Agent						wasAttributedTo/wasAssociatedWith		
		Product Team						wasAttributedTo/wasAssociatedWith		
		Data Producer						wasAttributedTo/wasAssociatedWith	x	
		Data Consumer						Needed if we store users' comments for the benefit of other users		role=modelUser used DEM
	Data									
		PreoperationalCalibrationData						wasGeneratedBy/used/wasDerivedFrom		
		ScienceDataProducts						wasGeneratedBy/used/wasDerivedFrom		
			Raw Data Product					wasGeneratedBy/used/wasDerivedFrom		
			DerivedDataProduct (attributes: spatial & temporal extents)					wasGeneratedBy/used/wasDerivedFrom		wasGeneratedBy DataProducer(role=modelGenerator_
				DEM				wasGeneratedBy/used/wasDerivedFrom		
			Science Data Image (raster)					wasGeneratedBy/used/wasDerivedFrom		
				Figure				wasGeneratedBy/used/wasDerivedFrom	x	
		Geolocation data						wasGeneratedBy/used/wasDerivedFrom		
		Calibration Data						wasGeneratedBy/used/wasDerivedFrom		
		Model Data						wasGeneratedBy/used/wasDerivedFrom		
								wasGeneratedBy/used/wasDerivedFrom		

Extending W3C PROV for Earth Science

- Prototype extending **W3C PROV**-Ontology
- Focus on modeling Earth Science Data System Processing
- **PROV-ES** supports capturing
  - A runtime environment has sessions which have one or more processing steps
  - Each processing step has executables, which have command-line arguments
  - Executables implement algorithms that follow documentation
- Providing lineage derivations leading to more trust

Interoperability with W3C PROV

- Maximize reuse of available tools interoperable with PROV
  - Examples include (but not limited to):
    - ProvToolbox (Java)
    - Prov (Python)
    - Prov-O-Viz (Javascript)
- Provenance capture instrumented into processing steps
  - Entities, activities, and agents are kept loosely-coupled to enable more versatile provenance curation
- Published prov-es traces are ingested into ElasticSearch and explored via Faceted Search

PROV-ES Mapping to W3C PROV

Define *additional attributes* to W3C PROV to encode PROV-ES constructs

ES concept	PROV concept	Additional Attributes		
		key	value	cardinality
eos:dataset	prov:Entity			
		prov:type	*eos:dataset*	
		info:doi	doi	[0..1]
		gcis:accessURL	xsd:anyURI	[0..*]
		gcis:downloadURL	xsd:anyURI	[0..*]
		gcis:sourceInstrument	eos:instrument	[0..*]
		eos:level	xsd:string	[0..1]
eos:platform	prov:Entity			
		prov:type	*eos:platform*	
		gcis:hasInstrument	eos:instrument	[0..*]
eos:instrument	prov:Entity			
		prov:type	*eos:instrument*	
		gcis:inPlatform	eos:platform	
		gcis:hasSensor	eos:sensor	[0..*]
		gcis:hasGoverningOrganization	prov:Organization	[0..*]
eos:sensor	prov:Entity			
		prov:type	*eos:sensor*	
		gcis:inInstrument	eos:instrument	
eos:software	prov:Entity			
		prov:type	*eos:software*	
		gcis:implements	eos:algorithm	[0..*]
eos:algorithm	prov:Entity			
		prov:type	*eos:algorithm*	
		gcis:implementedIn	eos:software	[0..*]
		eos:describedBy	bibo:Document	[0..*]
eos:document	prov:Entity			
		prov:type	*bibo:Document*	
		info:doi	doi	[0..1]
		gcis:downloadURL	xsd:anyURI	[0..*]
eos:processStep	prov:Activity			
		prov:type	*eos:processStep*	
		eos:usesSoftware	eos:software	[0..*]
		prov:wasAssociatedWith	prov:SoftwareAgent	[0..1]
		eos:runTimeContext	eos:runTimeContext	[0..1]
		prov:used	eos:dataset	[0..*]
		prov:generated	eos:dataset	[0..*]
eos:softwareAgent	prov:SoftwareAgent			
		prov:type	*prov:SoftwareAgent*	
		hysds:pid	xsd:string	[0..1]
		hysds:host	xsd:string	[0..1]
eos:runtimeContext	prov:Entity			
		prov:type	*eos:runtimeContext*	
		eos:hasRuntimeParameter	xsd:string	[0..*]
eos:download	prov:Activity	prov:type	*eos:download*	

Extending PROV: Via Subclassing

- Leverage extended concepts for Earth science data systems
- Retains for ontological concepts
- May need PROV-ES-specific implementations
- Could leverage rules / classifications to assert extended classes
- PROV-N example

```
L3DailyAverage(e1)
L3DailyAveraging(a1)
wasGeneratedBy(e1, a1)
wasStartedBy(a1, e1, -, 2013-08-29T10:00:00)
organization(ag1, [name="NASA"])
authored(e1, ag1)
```

Extending PROV: Via Attributes

- Does not retain hierarchy from ontology/taxonomy
- Can reuse generic existing W3C PROV tools as-is
  - http://www.w3.org/TR/prov-implementations/
- Could leverage rules / classifications to assert attributes to extended classes
- PROV-N

```
entity(e1, [prov:type="eos:L3DailyAverage"] )
activity(a1, [prov:type="eos:L3DailyAveraging"] )
wasGeneratedBy(e1, a1)
wasStartedBy(a1, e1, -, 2013-08-29T10:00:00)
agent(ag1, [prov:label=”NASA”,
prov:type='prov:Organization’])
wasAttributedTo(e1, ag1, [prov:type="authorship"])
```

PROV-ES Encoding: Example 1

- Science Data Product Validation
  - *Agent ag1 performs activity a1 to generate science data products e1 and e2*

```
entity(e1, [prov:type="eos:ScienceDataProduct"] )
entity(e2, [prov:type="eos:ScienceDataProduct"] )
activity(a1, [prov:type="eos:ScienceDataProductValidation”])
wasGeneratedBy(e1, a1)
wasGeneratedBy(e2, a1)
wasStartedBy(a1, e1, -, 2013-08-29T10:00:00)
wasEndedBy(a1, e1, -, 2013-09-29T10:00:00)
wasStartedBy(a1, e2, -, 2013-09-29T10:00:00)
wasEndedBy(a1, e2, -, 2013-08-29T10:00:00)
agent(ag1, [prov:label=”NASA”, prov:type='prov:Organization’])
wasAttributedTo(e1, ag1, [prov:role="authorship"])
```

PROV-ES Encoding: Example 2

- Perform a daily means over the entire collection of *Aqua AIRS Level 2 Standard Physical Retrieval (AIRS+AMSU) V006 (AIRX2RET) collection* for the variable *total cloud fraction* (3 by 3 AIRS Fields of View)

```
entity(airx2ret_6, [prov:type="eos:DerivedDataProduct”, eos:collection="AIRX2RET",
eos:version="6"] )
entity(airx2ret_dailymeans_6, [prov:type="eos:DerivedDataProduct”,
eos:data_level="3"] )
entity(airx_daily_means_software_6, [prov:type=”eos:PGE_software”])

activity(daily_averaging_6, [prov:type="eos:Averaging", period="daily”,
variables=”CldFrcTot”] )

wasGeneratedBy(airx2ret_dailymeans_6, daily_averaging_6)
used(daily_averaging_6, airx2ret_6)

wasStartedBy(daily_averaging_6, airx_daily_means_software_6, -, 2013-08-29T10:00:00)

Agent(project, [prov:label=”MEaSUREs”, prov:type='project’])

wasAttributedTo(airx2ret_dailymeans_6, project, [prov:role="authorship"])

wasDerivedFrom(airx2ret_dailymeans_6, airx2ret_6, daily_averaging_6)
```

PROV-ES Encoding: Example 3

- A project generates daily means of total cloud fraction of AIRX2RET collection.
- Another project uses the daily means of cloud fraction to do a seasonal comparison with the Community Atmosphere Model (CAM) 5.1.1 model data.
- prov:bundle
  - "A bundle is a named set of provenance descriptions, and is itself an entity, so allowing provenance of provenance to be expressed."

```
bundle jpl:measures_wvcc
  entity(airx2ret, [prov:type="eos:DerivedDataProduct", eos:collection="AIRX2RET",
eos:version="6", doi="..."] )
  entity(airx2ret_dailymeans, [prov:type="eos:DerivedDataProduct",
eos:data_level="3"] )
  activity(daily_averaging, [prov:type="eos:Averaging", period="daily",
variables="CldFrcTot"] )
  wasGeneratedBy(airx2ret_dailymeans, daily_averaging)
  used(daily_averaging, airx2ret)
  wasStartedBy(daily_averaging, airx2ret_dailymeans, -, 2013-08-29T10:00:00)
  agent(project, [prov:label="MEaSUREs", prov:type='project'])
  wasAttributedTo(airx2ret_dailymeans, project, [prov:role="authorship"])
  wasDerivedFrom(airx2ret_dailymeans, airx2ret, daily_averaging)
endBundle
```

PROV-ES Encoding: Example 3

- A project generates daily means of total cloud fraction of AIRX2RET collection.
- Another project uses the daily means of cloud fraction to do a seasonal comparison with the Community Atmosphere Model (CAM) 5.1.1 model data.
- prov:bundle
  - "A bundle is a named set of provenance descriptions, and is itself an entity, so allowing provenance of provenance to be expressed."

```
bundle ncar:user
  entity(airx2ret_dailymeans, [prov:type="eos:DerivedDataProduct",
eos:data_level="3"] )
  entity(cam5, [prov:type="eos:DerivedDataProduct", version="5.1.1"] )
  entity(statistics_results, [prov:type="eos:ScienceDataProduct"] )
  activity(comparison, [prov:type="eos:DataComparison", start_time="2010-06-
01T00:00:00", end_time="2010-09-01T00:00:00"] )
  used(comparison, airx2ret_dailymeans)
  used(comparison, cam5)
  wasGeneratedBy(statistics_results, comparison)
  wasStartedBy(comparison, statistics_results, -, 2013-09-29T10:00:00)
  agent(project, [prov:label="ACCESS", prov:type='project'])
  wasAttributedTo(statistics_results, project, [prov:role="authorship"])
  wasDerivedFrom(statistics_results, airx2ret_dailymeans, comparison)
  wasDerivedFrom(statistics_results, cam5, comparison)
endBundle
```

PROV-ES Implementation

- Leveraged open source PROV Python API
  - A library for W3C Provenance Data Model supporting PROV-JSON import/export
  - http://prov.readthedocs.org/
  - Follows PROV-JSON
    - http://www.w3.org/Submission/prov-json/
- Define additional attributes to W3C PROV to encode PROV-ES constructs
- PROV Earth Science mapping to existing W3C PROV terms
  - W3C PROV's Table 2 : Mapping of PROV core concepts to types and relations
  - Temporarily using GCIS namespace

- Develop instrumentation Python API for PROV-ES
- Infuse into production system for near real-time InSAR processing science data system

Example PROV-ES (JSON)


```
{
  "used": {
    "_:id5": {
      "prov:time": "2014-12-15T06:25:21.982294+00:00",
      "prov:entity": "hysds:network_selector",
      "prov:activity": "hysds:job"
    },
    "_:id4": {
      "prov:time": "2014-12-15T06:25:21.982294+00:00",
      "prov:entity": "hysds:dem_file",
      "prov:activity": "hysds:job"
    },
    ...
  },
  "agent": {
    "eos:JPL": {
      "prov:type": {
        "type": "prov:QualifiedName",
        "$": "prov:Organization"
      }
    },
 "eos:ASI": {
      "prov:type": {
        "type": "prov:QualifiedName",
        "$": "prov:Organization"
      }
    }
  },
  "entity": {
    "hysds:interferogram_output": {
      "gcis:sourceInstrument": "eos:CSKS4-SAR",
      "eos:level": "L1",
      "prov:type": "eos:dataset",
      "eos:version": "v0.4",
      "gcis:downloadURL": ".../interferogram_T37_F347-354_CSKS4_20140406-
CSKS4_20140116_2014-12-15T062516.922740"
    },
    "eos:ISCE": {
      "gcis:implements": "eos:interferogram_generation",
      "eos:version": "2.0.0_201410",
      "prov:type": "eos:software"
    },
    "eos:SAR": {
      "gcis:inInstrument": "eos:CSKS4-SAR",
      "prov:type": "eos:sensor"
    },
    "eos:CSKS4-SAR": {
      "gcis:hasSensor": "eos:SAR",
      "gcis:inPlatform": "eos:CSKS4",
      "prov:type": "eos:instrument",
      "gcis:hasGoverningOrganization": "eos:ASI"
    },
    ...
  },
  "activity": {
    "hysds:job": {
      "eos:runTimeContext": "hysds:project",
      "prov:wasAssociatedWith": "hysds:rossini-vm-5.jpl.nasa.gov/7174",
      "prov:startTime": "2014-12-15T05:59:25.903682+00:00",
      "prov:type": "eos:processStep",
      "eos:usesSoftware": "eos:ISCE",
      "hysds:job_id": "hysds:create_interferogram-
CSKS4_RAW_HI_14_HH_RA_20140116134252_20140116134259.interferogram.json_1-
20141215T052408.367084Z",
      "prov:endTime": "2014-12-15T06:25:22.380207+00:00"
    }
  },
  "wasGeneratedBy": {
    "_:id7": {
      "prov:time": "2014-12-15T06:25:21.982294+00:00",
      "prov:entity": "hysds:interferogram_output",
      "prov:activity": "hysds:job"
    }
  }
}
```

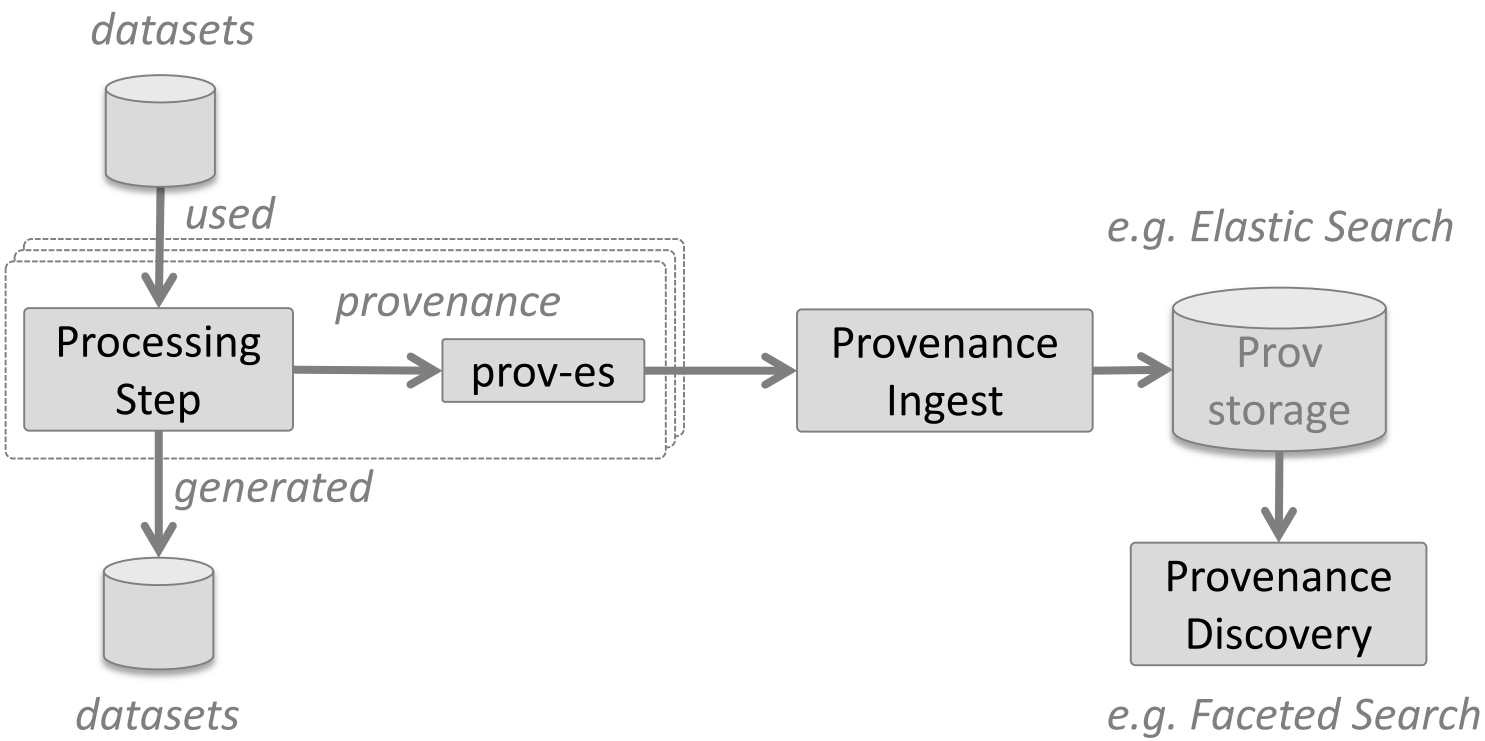
Example PROV-ES (JSON)


```
"activity": {
  "hysds:create_interferogram-EL20130625_673969_2940232.6.2": {
    "eos:runTimeContext": "hysds:runtime_context",
    "prov:wasAssociatedWith": "hysds:ariamh-worker-32.jpl.nasa.gov/12353",
    "prov:startTime": "2014-09-23T18:43:52.654935+00:00",
    "prov:type": "eos:processStep",
    "eos:usesSoftware": "eos:ISCE",
    "prov:endTime": "2014-09-23T22:07:45.654935+00:00"
  }
},
"wasGeneratedBy": {
  "_:id3": {
    "prov:time": "2014-09-23T22:07:45.654935+00:00",
    "prov:entity": "hysds:create_interferogram-EL20130625_673969_2940232.6.2",
    "prov:activity": "hysds:interferogram__T22_F314-330_CSKS1_20130828-CSKS1_20130609"
  }
}

"used": {
  "_:id1": {
    "prov:time": "2014-09-23T18:43:52.654935+00:00",
    "prov:entity": "hysds:CSKS2_RAW_HI_06_HH_RA_20140922062622_20140922062629",
    "prov:activity": "hysds:create_interferogram-EL20130625_673969_2940232.6.2"
  },
  "_:id2": {
    "prov:time": "2014-09-23T18:43:52.654935+00:00",
    "prov:entity": "hysds:srtm/version2_1/SRTM1/Region_01/N31W114",
    "prov:activity": "hysds:create_interferogram-EL20130625_673969_2940232.6.2"
  }
},
"agent": {
  "hysds:ariamh-worker-32.jpl.nasa.gov/12353": {
    "hysds:pid": "12353",
    "prov:type": {
      "type": "prov:QualifiedName",
      "$": "prov:SoftwareAgent"
    },
```

FACETED PROVENANCE

PROV-ES Generation and Consumption



PROV-ES Faceted Search

- Enables drilling down across end-to-end provenance concepts such as datasets, processing steps, algorithms, documentation, instruments, sensors, platforms, etc.
- Distributed prov-es traces crawled and ingested into Elastic Search datastore
- Faceted view of aggregate provenance
- Lineage
- Provenance graph links assembled via search links
- Supports real-world aspect of independent provenance bundles

PROV-ES Faceted Search

PROV-ES Facet Search

https://provenance.jpl.nasa.gov:5000/facetview_prov_es?source={"query":{"bool":{"must":[{"term":{"prov:concept.raw":

PROV-ES Facet Search Home

PROV-ES Facet Search faceted search interface for earth science provenance

FDT

? 10 search term

prov:type
? 10 count OR range
eos:dataset (814)

prov:concept
? 10 count OR range
prov:Entity (814)

− gcis:sourceInstrument
? 10 count OR range
eos:CSKS4-SAR (389)
eos:CSKS1-SAR (309)
eos:CSKS3-SAR (8)

− eos:level
? 10 count OR range
L0 (679)
L1 (27)

eos:version
? 10 count OR range
v0.4 (706)

gcis:hasInstrument

prov:Entity ✖ eos:dataset ✖

.. 1 – 10 of 814 next »

http://rossini-vm-5.jpl.nasa.gov:8085/jobs/2014/12/15/create_interferogram-CSKS4_RAW_HI_14_HH_RA_20140116134252_20140116134259.interferogram.json_0-20141215T052408.362130Z/interferogram_T37_F347-354_CSKS4_20140116-CSKS4_20131215_2014-12-15T055914.094460

prov:type: eos:dataset
prov:concept: prov:Entity
gcis:downloadURL: http://rossini-vm-5.jpl.nasa.gov:8085/jobs/2014/12/15/create_interferogram-CSKS4_RAW_HI_14_HH_RA_20140116134252_20140116134259.interferogram.json_0-20141215T052408.362130Z/interferogram_T37_F347-354_CSKS4_20140116-CSKS4_20131215_2014-12-15T055914.094460
gcis:sourceInstrument: eos:CSKS4-SAR
eos:level: L1
eos:version: v0.4

http://rossini-vm-5.jpl.nasa.gov:8085/jobs/2014/12/15/create_interferogram-CSKS4_RAW_HI_14_HH_RA_20140116134252_20140116134259.interferogram.json_0-20141215T052408.362130Z/demLat_N34_N36_Lon_W118_W117.dem.xml

prov:type: eos:dataset
prov:concept: prov:Entity
gcis:downloadURL: http://rossini-vm-5.jpl.nasa.gov:8085/jobs/2014/12/15/create_interferogram-CSKS4_RAW_HI_14_HH_RA_20140116134252_20140116134259.interferogram.json_0-20141215T052408.362130Z/demLat_N34_N36_Lon_W118_W117.dem.xml

http://rossini-vm-5.jpl.nasa.gov:8085/jobs/2014/12/15/create_interferogram-CSKS4_RAW_HI_14_HH_RA_20140116134252_20140116134259.interferogram.json_0-20141215T052408.362130Z/interferogram_T37_F347-354_CSKS4_20140116-CSKS4_20131215_2014-12-15T055914.094460/job_description.json_0

HARMONIZATION?

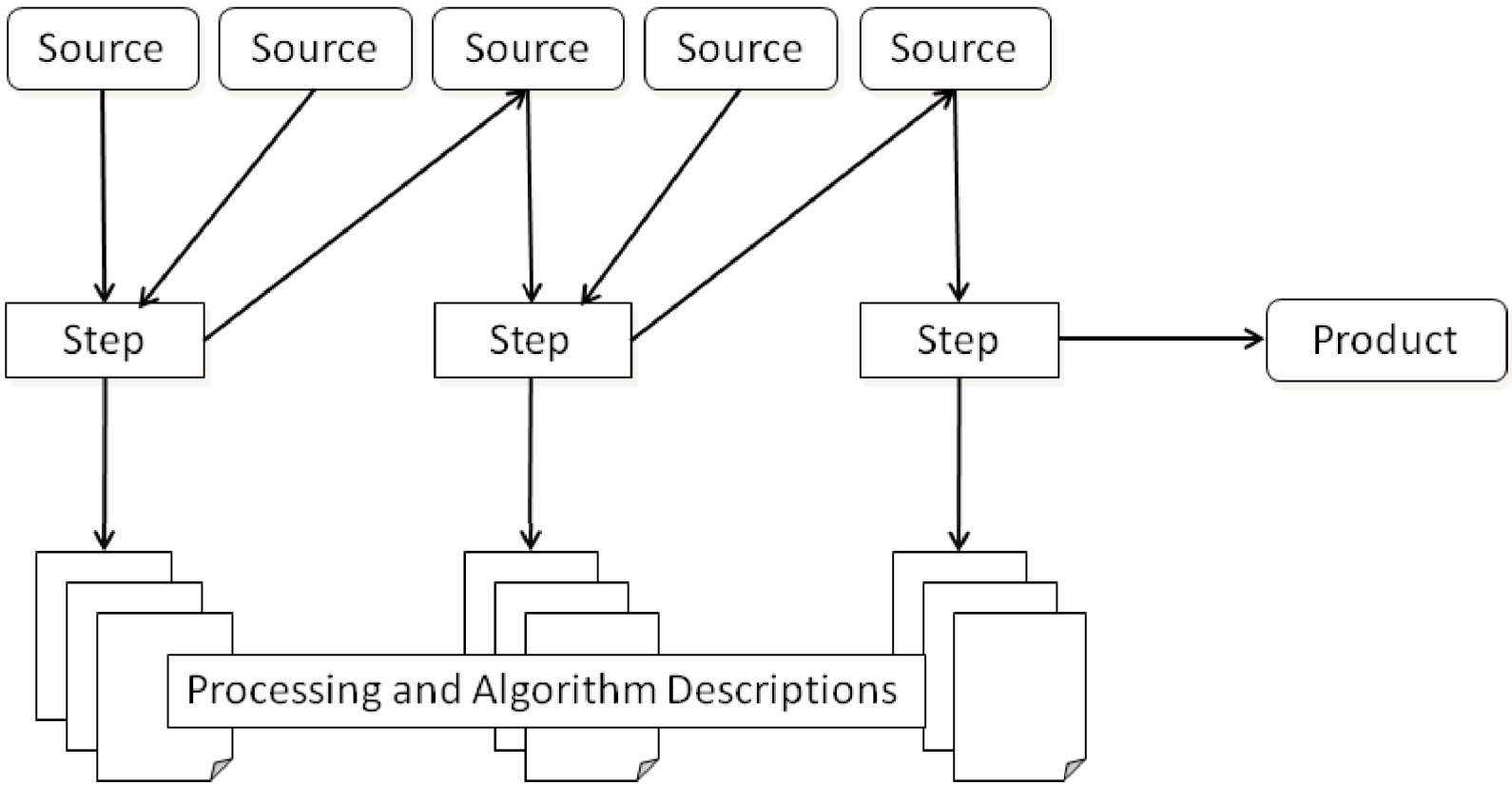
Harmonization

- Harmonizing
  - W3C PROV-O
    - http://www.w3.org/TR/2013/REC-prov-o-20130430/
  - PROV-ES
    - https://docs.google.com/spreadsheets/d/1gsyuKY-xD1yQaHD6QVB6vJ3YmDWL9WKPiSopW8BtuB8/edit#gid=0
  - GCIS OWL
    - http://data.globalchange.gov/gcis.owl
- PROV Earth Science concepts should map to existing W3C PROV terms
  - W3C PROV's Table 2 : Mapping of PROV core concepts to types and relations

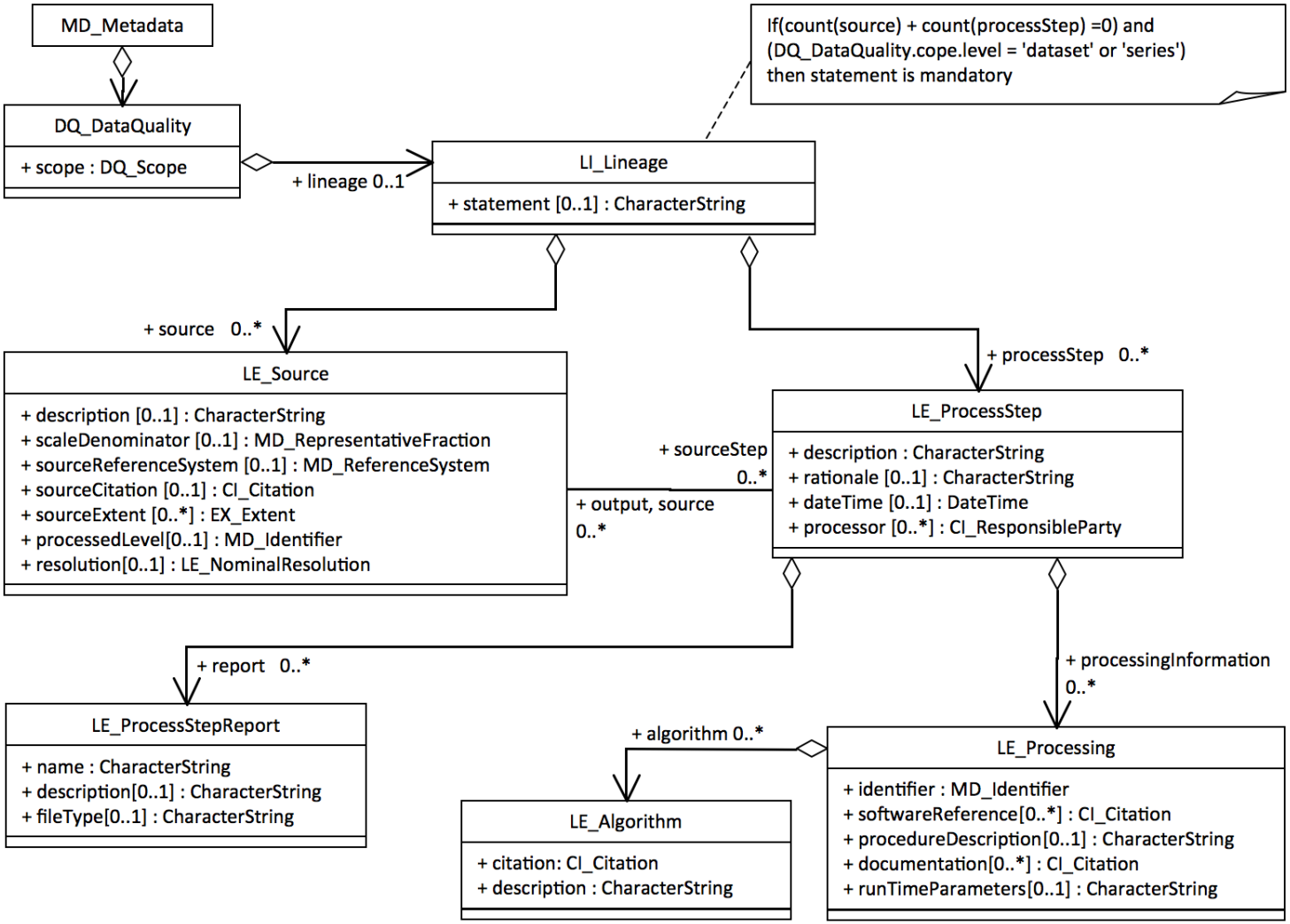
Can We Harmonize the Models ?

- ISO Lineage
  - Provides high-level linage records
  - Can embed more detailed provenance traces via mechanisms such as Record/RecordTypes
- W3C PROV
  - A generic model for provenance
  - Bundles for “provenance of provenance”
- NASA/ESIP PCS
  - Captures mission data life-cycle concepts
- PROV-ES
  - Earth-science extensions to W3C PROV

ISO 19115 Lineage



ISO 19115-2 Lineage



DQ_Lineage (19115-2)

Resources

- ESDSWG PROV-ES Working Group
  - Wiki
    - https://wiki.earthdata.nasa.gov/display/ESDSWG/Provenance+for+Earth+Science+(PROV-ES)+Working+Group (login required)
  - Mailing list
    - esdswg-prov-es@lists.nasa.gov
    - Subscribe at https://lists.nasa.gov/mailman/listinfo/esdswg-prov-es
  - Telecon notes
    - https://docs.google.com/document/d/1mw0ROVcW4-7W3IfZk2pizM1k6sHM2hgmRTvChrYtpO8/edit?pli=1
  - Use Case Mapping
    - https://docs.google.com/spreadsheet/ccc?key=0AlQ95ca89UmYdEJXX05NVDRhRGRWV09hcmNENnFROHc#gid=0
  - PROV-ES Mapping to W3C PROV
    - https://docs.google.com/spreadsheets/d/1gsyuKY-xD1yQaHD6QVB6vJ3YmDWL9WKPiSopW8BtuB8/edit#gid=0
- NASA Earth Science Data Preservation Content Specification (NASA ESD PCS)
  - http://earthdata.nasa.gov/sites/default/files/field/document/423-SPEC-001_NASA%20ESD_Preservation_Spec_OriginalCh01_0.pdf
- Provenance and Context Content Standard
  - http://wiki.esipfed.org/index.php/Provenance_and_Context_Content_Standard