



HAL
open science

ENVRI-FAIR provenance: AnaEE use case

Amélie Fiocca, Christian Pichot

► **To cite this version:**

| Amélie Fiocca, Christian Pichot. ENVRI-FAIR provenance: AnaEE use case. 2021. hal-03277338

HAL Id: hal-03277338

<https://hal.inrae.fr/hal-03277338>

Preprint submitted on 3 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



ENVRI-FAIR provenance



Amélie Fiocca et Christian Pichot

INRAE

February 18th, 2021



ENVRI-FAIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068



AnaEE sites = 'platforms' for experimentation on ecosystems

- In vitro**



- Analytical**



Managing data/metada for:

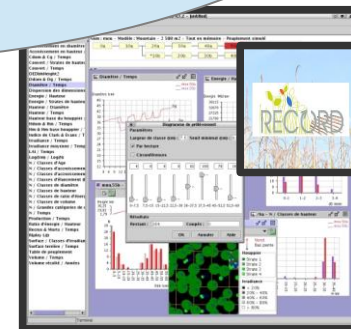
- discovery
- access to resources : platforms, datasets...

Including PROVENANCE!



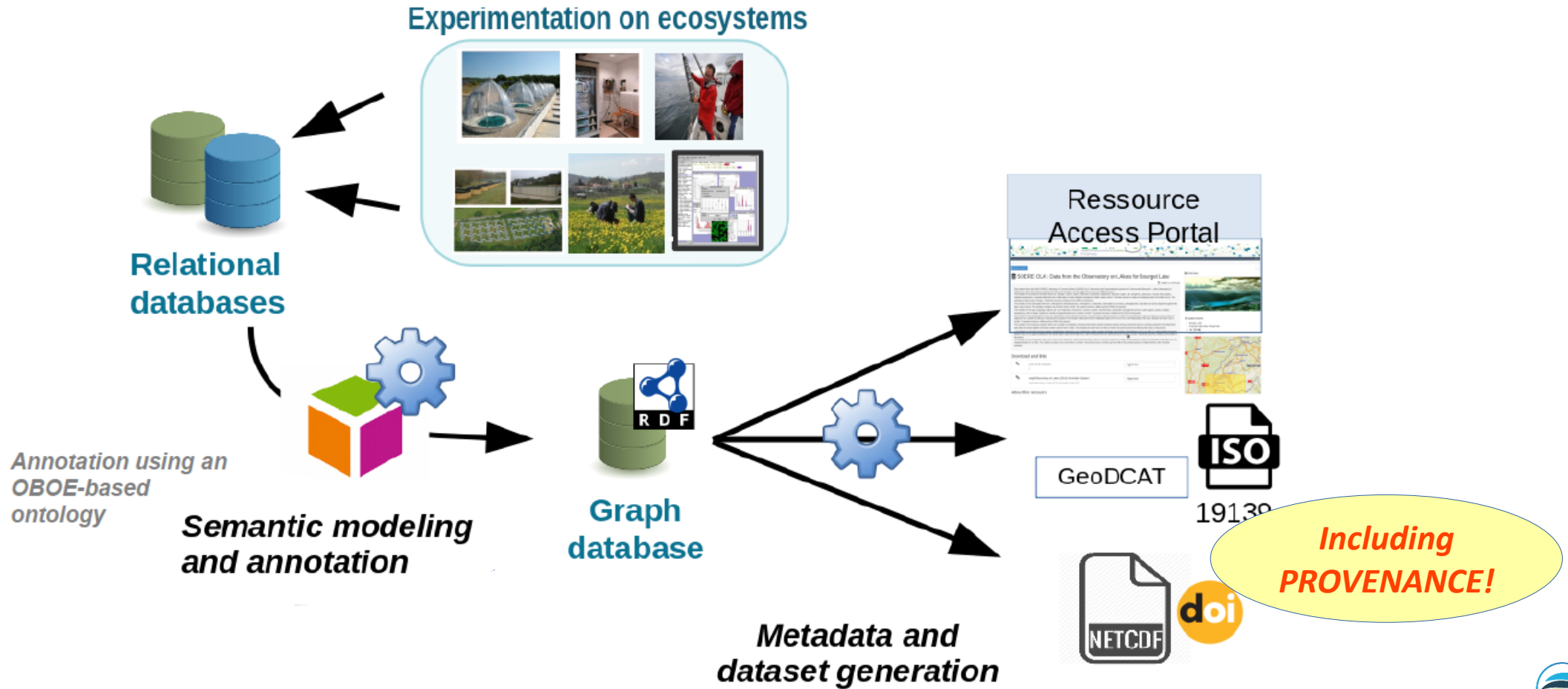
- Data & modelling**

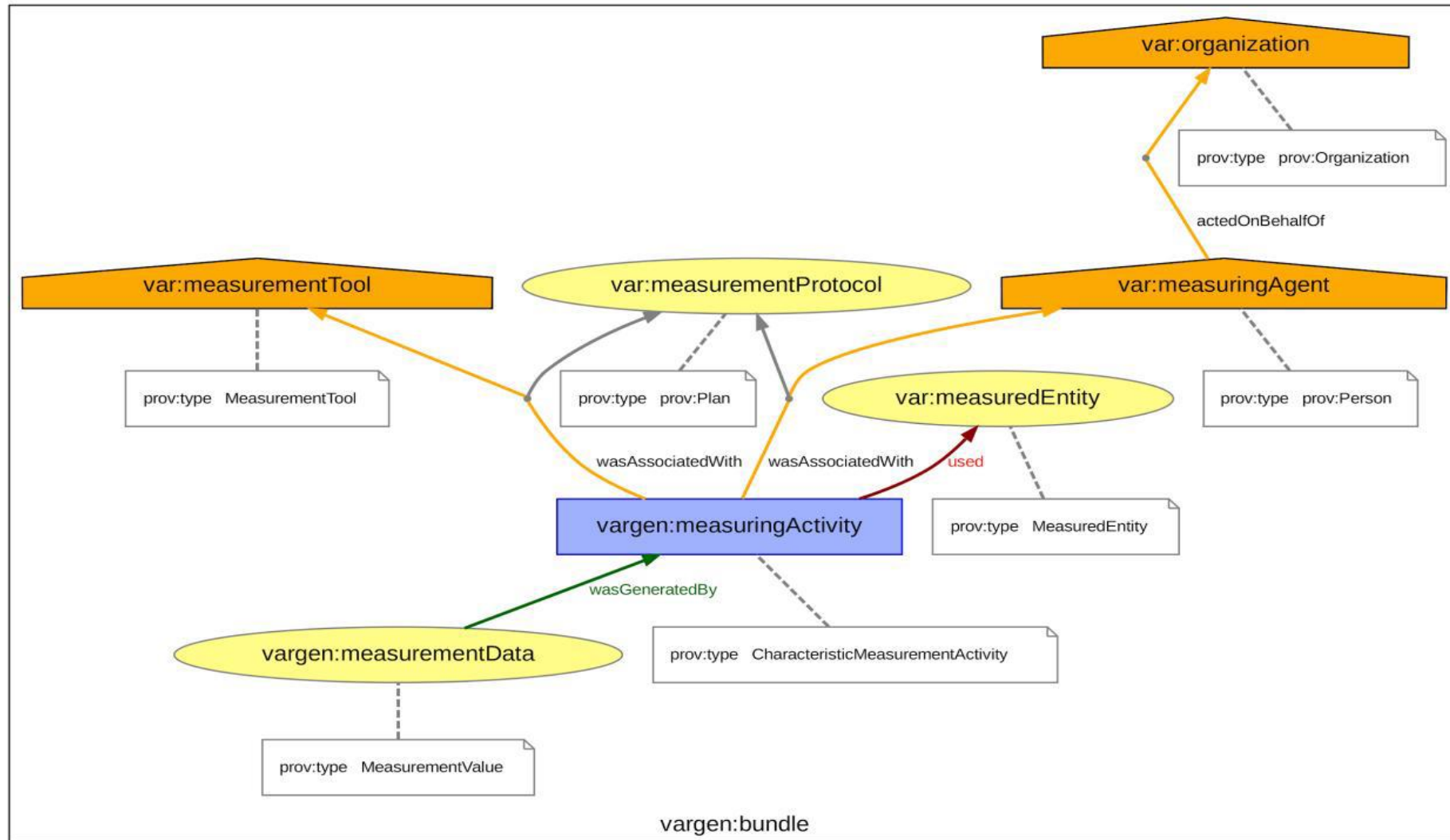
AnaEE provenance use case





AnaEE workflows





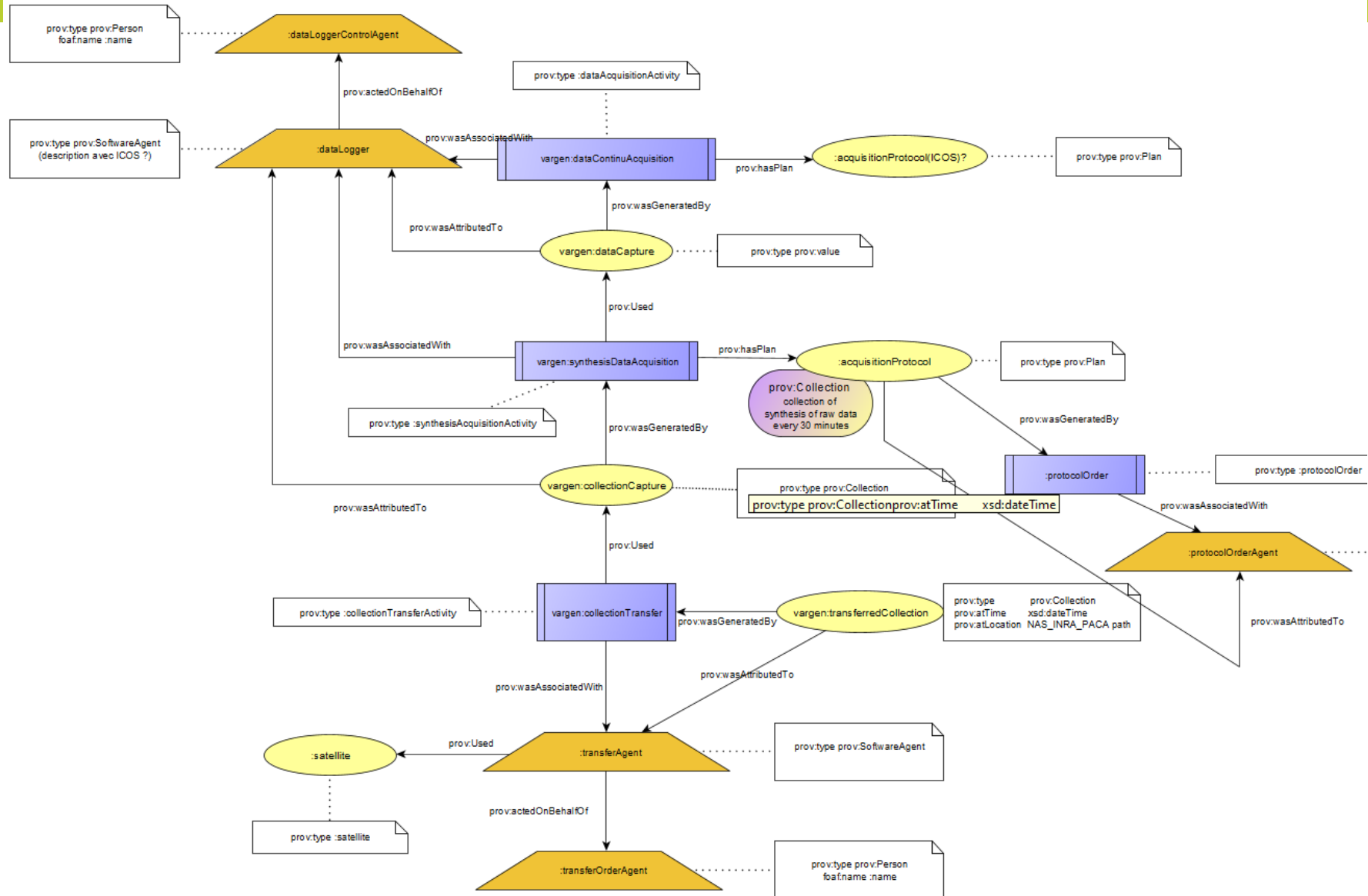


Objective and methodology

- Our objective :
 - Provide provenance informations for datas and metadatas users
 - Generate automatically this provenance informations as metadata for datasets
- Methodology :
 - Provenance element identification and flow modeling
 - Test provenance demonstrator with Jupyter Notebook
 - Provenance implementation test with AnaEE semantic pipelines and generation of rdf triples

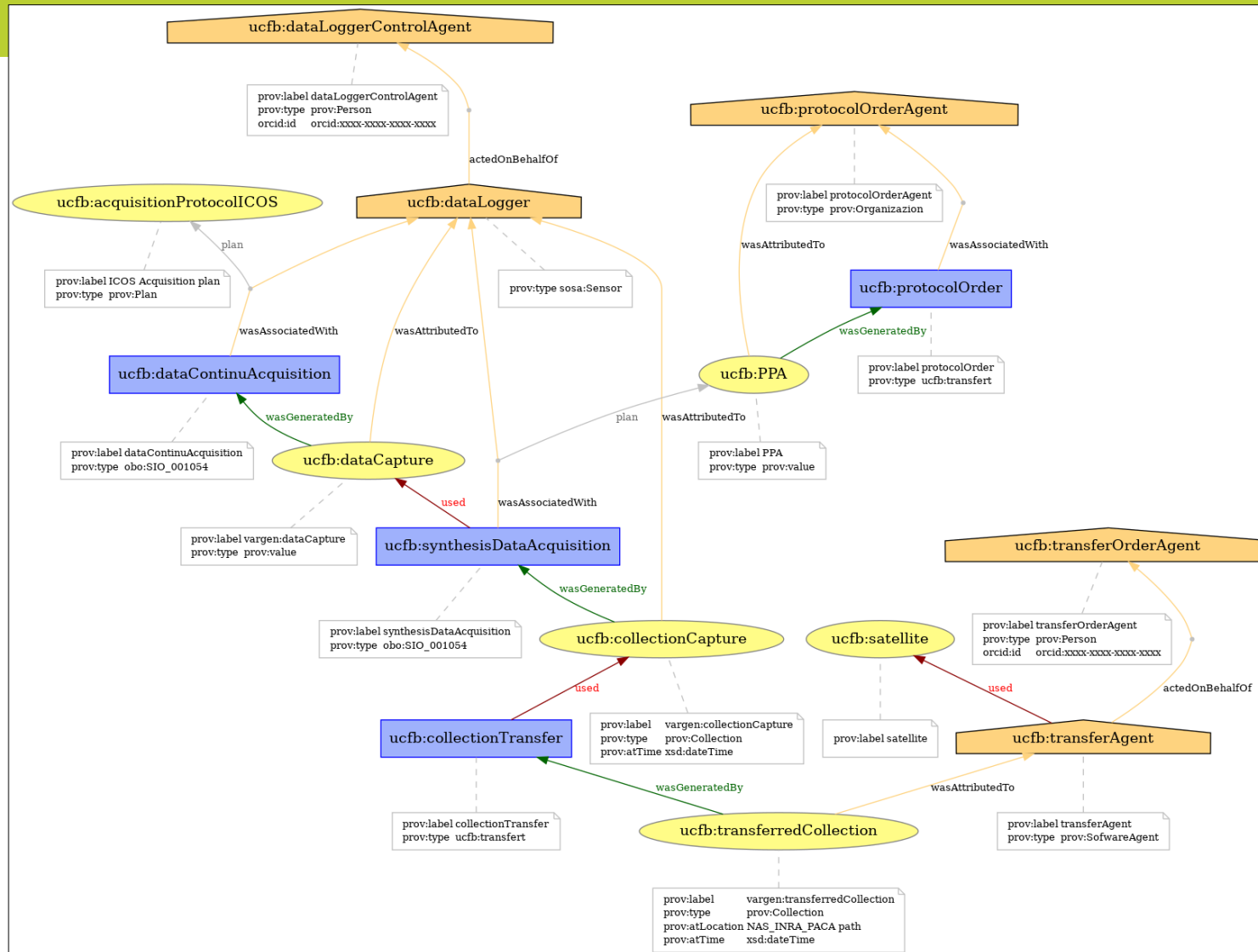


Flux modelisation «Font-Blanche» use case





Provenance demonstrator with Jupyter Notebook

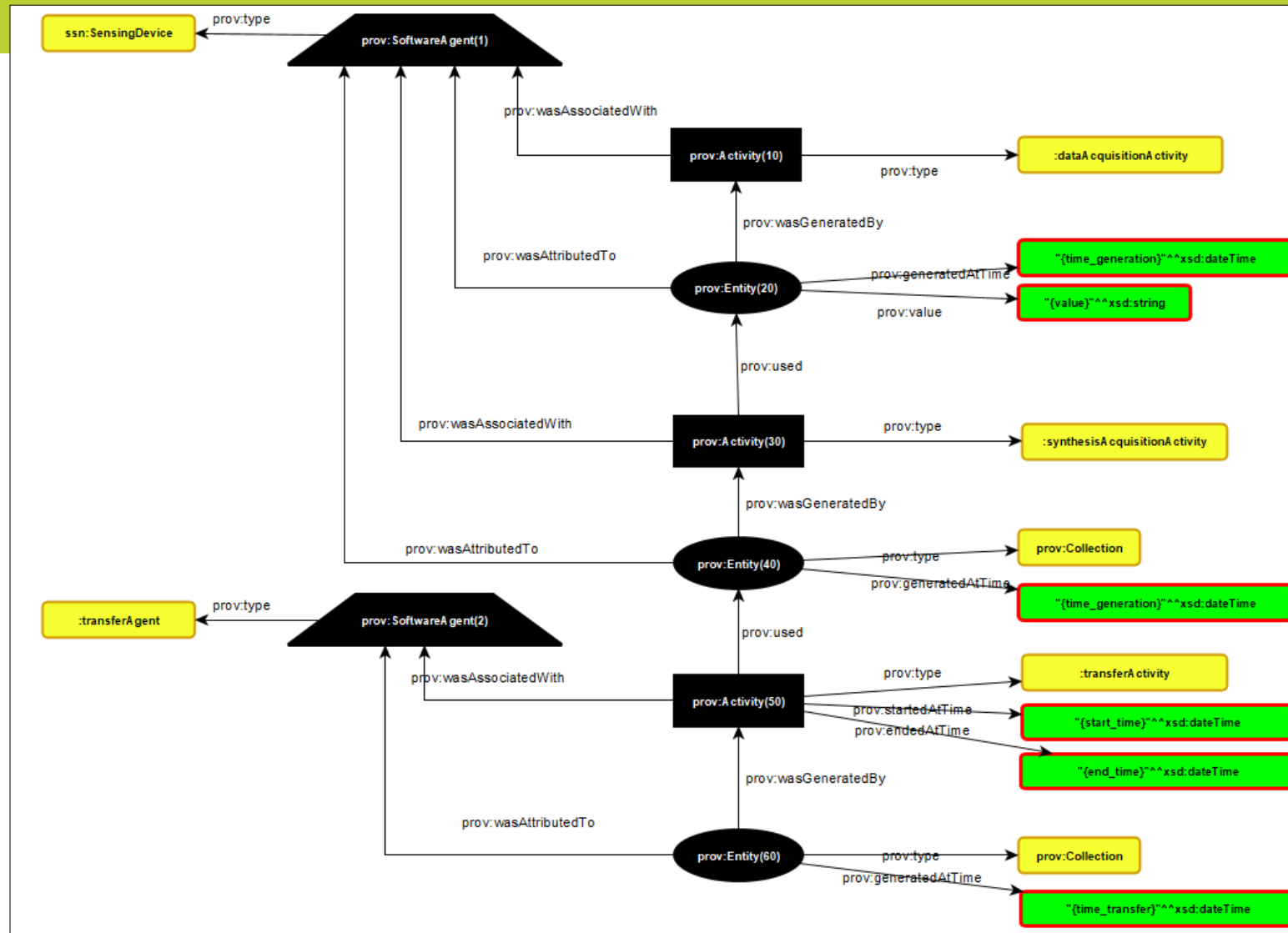


AnaEE provenance use case



Provenance implementation with AnaEE semantic pipelines

Graph syntax compliant with existing pipelines



AnaEE provenance use case



Generation of triples for Provenance

```
<http://semanticscience.org/resource/SIO_001054> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/prov#Activity> .
<http://semanticscience.org/resource/SIO_001054> <http://www.w3.org/ns/prov#wasAssociatedWith> <http://prov_agent/FontBlancheDataLogger1/1> .
<http://semanticscience.org/resource/SIO_001054> <http://www.w3.org/ns/prov#type> <http://www.anaee-france.fr/ontology/anaee-france_ontology#dataAcquisitionActivity> .
<http://semanticscience.org/resource/SIO_001054> <http://www.w3.org/ns/prov#wasAssociatedWith> <http://prov_agent/FontBlancheDataLogger1/3> .
<http://prov_agent/FontBlancheDataLogger1/1> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/prov#SoftwareAgent> .
<http://prov_agent/FontBlancheDataLogger1/1> <http://www.w3.org/ns/prov#type> <http://purl.oclc.org/NET/ssnx/ssn#SensingDevice> .
<http://prov_agent/FontBlancheDataLogger1/3> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/prov#SoftwareAgent> .
<http://prov_agent/FontBlancheDataLogger1/3> <http://www.w3.org/ns/prov#type> <http://purl.oclc.org/NET/ssnx/ssn#SensingDevice> .
<http://prov_entity/sensor4CO2/2020-10-19%2012:00:00> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/prov#Entity> .
<http://prov_entity/sensor4CO2/2020-10-19%2012:00:00> <http://www.w3.org/ns/prov#value> "440.20100000000022"^^<http://www.w3.org/2001/XMLSchema#string> .
<http://prov_entity/sensor4CO2/2020-10-19%2012:00:00> <http://www.w3.org/ns/prov#wasGeneratedBy> <http://semanticscience.org/resource/SIO_001054> .
<http://prov_entity/sensor4CO2/2020-10-19%2012:00:00> <http://www.w3.org/ns/prov#wasAttributedTo> <http://prov_agent/FontBlancheDataLogger1/1> .
<http://prov_entity/sensor4CO2/2020-10-19%2012:00:00> <http://www.w3.org/ns/prov#generatedAtTime> "2020-10-19 12:00:00"^^<http://www.w3.org/2001/XMLSchema#dateTime> .
<http://prov_entity/sensor4CO2/2020-10-19%2012:30:00> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/prov#Entity> .
<http://prov_entity/sensor4CO2/2020-10-19%2012:30:00> <http://www.w3.org/ns/prov#value> "440.218999999999994"^^<http://www.w3.org/2001/XMLSchema#string> .
<http://prov_entity/sensor4CO2/2020-10-19%2012:30:00> <http://www.w3.org/ns/prov#wasGeneratedBy> <http://semanticscience.org/resource/SIO_001054> .
<http://prov_entity/sensor4CO2/2020-10-19%2012:30:00> <http://www.w3.org/ns/prov#wasAttributedTo> <http://prov_agent/FontBlancheDataLogger1/1> .
<http://prov_entity/sensor4CO2/2020-10-19%2012:30:00> <http://www.w3.org/ns/prov#generatedAtTime> "2020-10-19 12:30:00"^^<http://www.w3.org/2001/XMLSchema#dateTime> .
<http://prov_entity/sensor4CO2/2020-10-19%2013:00:00> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/prov#Entity> .
<http://prov_entity/sensor4CO2/2020-10-19%2013:00:00> <http://www.w3.org/ns/prov#value> "440.220000000000027"^^<http://www.w3.org/2001/XMLSchema#string> .
<http://prov_entity/sensor4CO2/2020-10-19%2013:00:00> <http://www.w3.org/ns/prov#wasGeneratedBy> <http://semanticscience.org/resource/SIO_001054> .
<http://prov_entity/sensor4CO2/2020-10-19%2013:00:00> <http://www.w3.org/ns/prov#wasAttributedTo> <http://prov_agent/FontBlancheDataLogger1/1> .
<http://prov_entity/sensor4CO2/2020-10-19%2013:00:00> <http://www.w3.org/ns/prov#generatedAtTime> "2020-10-19 13:00:00"^^<http://www.w3.org/2001/XMLSchema#dateTime> .
<http://prov_entity/sensor4CO2/2020-10-19%2013:30:00> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/prov#Entity> .
<http://prov_entity/sensor4CO2/2020-10-19%2013:30:00> <http://www.w3.org/ns/prov#value> "440.221999999999998"^^<http://www.w3.org/2001/XMLSchema#string> .
<http://prov_entity/sensor4CO2/2020-10-19%2013:30:00> <http://www.w3.org/ns/prov#wasGeneratedBy> <http://semanticscience.org/resource/SIO_001054> .
<http://prov_entity/sensor4CO2/2020-10-19%2013:30:00> <http://www.w3.org/ns/prov#wasAttributedTo> <http://prov_agent/FontBlancheDataLogger1/1> .
<http://prov_entity/sensor4CO2/2020-10-19%2013:30:00> <http://www.w3.org/ns/prov#generatedAtTime> "2020-10-19 13:30:00"^^<http://www.w3.org/2001/XMLSchema#dateTime> .
<http://prov_entity/sensor4Temp/2020-10-19%2012:00:00> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/prov#Entity> .
<http://prov_entity/sensor4Temp/2020-10-19%2012:00:00> <http://www.w3.org/ns/prov#value> "16.3999999999999986"^^<http://www.w3.org/2001/XMLSchema#string> .
<http://prov_entity/sensor4Temp/2020-10-19%2012:00:00> <http://www.w3.org/ns/prov#wasGeneratedBy> <http://semanticscience.org/resource/SIO_001054> . <http://prov_entity/sensor4Temp/2020-10-19%2012:00:00>
<http://www.w3.org/ns/prov#wasAttributedTo> <http://prov_agent/FontBlancheDataLogger1/3> . <http://prov_entity/sensor4Temp/2020-10-19%2012:00:00> <http://www.w3.org/ns/prov#generatedAtTime> "2020-10-19
12:00:00"^^<http://www.w3.org/2001/XMLSchema#dateTime> .
```



State of advancement and issues

- Identification of really useful provenance elements for dataset users
- Evaluation of current provenance elements availability or the possibility of making them available in AnaEE platforms
- Provenance instantiation (=>rdf triples)



Identification of really useful provenance elements for dataset users (correspondance with D8.6 specifications)

Acquisition

- **Measuring instrument (AnaEE-R5)**
 - Name
 - Type
 - Brand
 - Model
 - Installation date
 - Installation protocole
 - Production date
 - Calibration informations
 - [PIDINST / B2INST ...](#)
- **Acquisition method**
 - Text description
- **Acquisition context (AnaEE-R4)**
 - Date
 - Position
- **Technical responsible (AnaEE-R6)**
 - Name
 - Contact
 - [Orchid ID](#)
- **Scientific responsible (AnaEE-R6)**
 - Name
 - Contact
 - [Orchid ID](#)
- **Project description**
 - Text description

Curation

- **Curation agent (AnaEE-R7)**
 - Name
 - Type
 - Author -> name, contact, [Orchid ID](#)
 - [\(Repository link\)](#)
- **Responsible for data curation (AnaEE-R6)**
 - Name
 - Contact
 - [Orchid ID](#)

Annotation / Production

- **Annotation / Production agent (AnaEE-R8-10)**
 - Name
 - Type
 - Author -> name, contact, [Orchid ID](#)
 - [\(Repository link\)](#)
- **Responsible for annotation / production (AnaEE-R6)**
 - Name
 - Contact
 - [Orchid ID](#)

Publication

- **Publication agent (AnaEE-R10)**
 - Name
 - Type
 - Author -> name, contact, [Orchid ID](#)
 - [\(Repository link\)](#)
- **Responsible for data Publication (AnaEE-R6)**
 - Name
 - Contact
 - [Orchid ID](#)