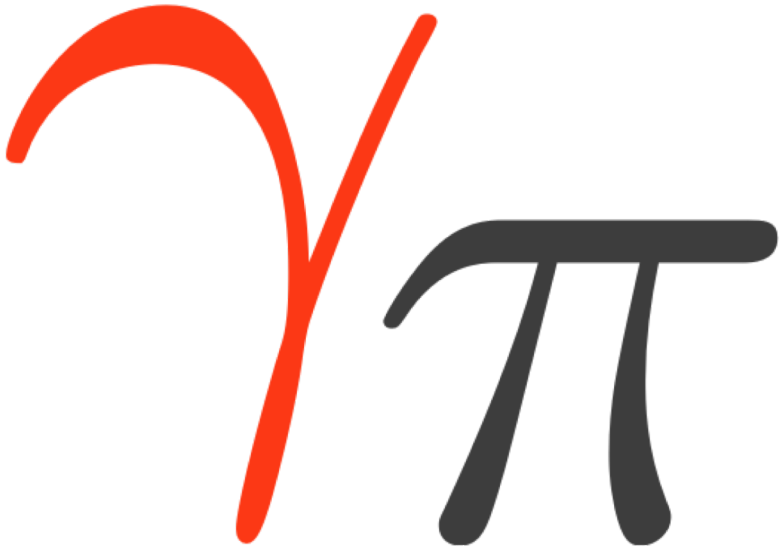

Provenance capture prototype

Granada 18th -22nd November 2019

J. Enrique Ruiz (IAA – CSIC), Mathieu Servillat (Obs. Paris-Meudon)

Context

- PIG 17 – Provenance capture PR #2458

- GOAL – Automatically record a structured provenance of the data analysis processes undertaken with the high-level interface in IPython working sessions (shell and notebooks) or in Python scripts

- Non-intrusive implementation
- Use IVOA Provenance Model
- Applied to Gammapy and LSTChain

Implementation

- Start from re-use of actual state of `ctapipe` provenance package
- All code is in `gammapy.utils.provenance`
- Add `@provenance` decorator in `Analysis` class
- Declare workflow dependencies in a `definition.yaml` file
- Record provenance in a log `prov.log` file

definition.yaml

```
activities:

    get_observations:
        description:
            "Fetch observations from the data store according
            to criteria defined in the configuration"
        parameters:
            - name: datastore_path
              description: "DataStore path as string"
              value: settings.observations.datastore
            - name: filters
              description: "Filter criteria to select observations"
              value: settings.observations.filters
        usage:
            - role: datastore
              description: "DataStore object file"
              entityName: DataStore
              value: settings.observations.datastore
        generation:
            - role: observations_selected
              description: "Observations selected"
              entityName: Observations
              value: observations
              has_members:
                  list: observations.list
                  entityName: Observation
                  id: obs_id
                  location: location(hdu_type="events").path().__str__()
                  namespace: gamma-events

    get_datasets:
        description: "Produce reduced datasets"
        parameters:
            - name: stack-datasets
              description: "Stack datasets flag"
              value: settings.datasets.stack-datasets
            - name: dataset-type
              description: "Datasets type"
              value: settings.datasets.dataset-type
              name: geom
```

prov.log

```
INFO provLogger _PROV_2019-11-18T10:48:20.428982_PROV_{'entity_id': 519318473740404787, 'name': 'Observations', 'type': 'PythonObject'}
INFO provLogger _PROV_2019-11-18T10:48:20.429298_PROV_{'activity_id': '722a28', 'generated_id': 519318473740404787, 'generated_role':
'observations_selected'}
INFO provLogger _PROV_2019-11-18T10:48:20.430010_PROV_{'entity_id': 'gamma-events:23523', 'name': 'Observation', 'location': '/Users/test/DATA/
gammapy-datasets/hess-dl3-dr1/data/hess_dl3_dr1_obs_id_023523.fits.gz', 'type': 'File', 'contentType': 'application/fits'}
INFO provLogger _PROV_2019-11-18T10:48:20.430217_PROV_{'entity_id': 519318473740404787, 'member_id': 'gamma-events:23523'}
INFO provLogger _PROV_2019-11-18T10:48:20.431010_PROV_{'entity_id': 'gamma-events:23526', 'name': 'Observation', 'location': '/Users/test/DATA/
gammapy-datasets/hess-dl3-dr1/data/hess_dl3_dr1_obs_id_023526.fits.gz', 'type': 'File', 'contentType': 'application/fits'}
INFO provLogger _PROV_2019-11-18T10:48:20.431250_PROV_{'entity_id': 519318473740404787, 'member_id': 'gamma-events:23526'}
INFO provLogger _PROV_2019-11-18T10:48:20.432081_PROV_{'entity_id': 'gamma-events:23559', 'name': 'Observation', 'location': '/Users/test/DATA/
gammapy-datasets/hess-dl3-dr1/data/hess_dl3_dr1_obs_id_023559.fits.gz', 'type': 'File', 'contentType': 'application/fits'}
INFO provLogger _PROV_2019-11-18T10:48:20.432216_PROV_{'entity_id': 519318473740404787, 'member_id': 'gamma-events:23559'}
INFO provLogger _PROV_2019-11-18T10:48:20.433059_PROV_{'entity_id': 'gamma-events:23592', 'name': 'Observation', 'location': '/Users/test/DATA/
gammapy-datasets/hess-dl3-dr1/data/hess_dl3_dr1_obs_id_023592.fits.gz', 'type': 'File', 'contentType': 'application/fits'}
INFO provLogger _PROV_2019-11-18T10:48:20.433252_PROV_{'entity_id': 519318473740404787, 'member_id': 'gamma-events:23592'}
INFO provLogger _PROV_2019-11-18T10:48:20.433418_PROV_{'activity_id': '722a28', 'endTime': '2019-11-18T10:48:18.995966'}
INFO provLogger _PROV_2019-11-18T10:48:25.940142_PROV_{'activity_id': '665a0c', 'name': 'get_datasets', 'startTime':
'2019-11-18T10:48:21.012240', 'in_session': 9223372036581382375, 'agent_name': 'test'}
INFO provLogger _PROV_2019-11-18T10:48:25.940709_PROV_{'activity_id': '665a0c', 'parameters': {'stack-datasets': True, 'dataset-type':
'MapDataset', 'geom': {'skydir': [83.633, 22.014], 'width': [5, 5], 'binsz': 0.04, 'coordsys': 'CEL', 'proj': 'TAN', 'axes': [{'name': 'energy',
'hi_bnd': 10, 'lo_bnd': 1, 'nbin': 4, 'interp': 'log', 'node_type': 'edges', 'unit': 'TeV'}]}, 'offset-max': '2.5 deg', 'psf-kernel-radius': '0.3
deg'}}
INFO provLogger _PROV_2019-11-18T10:48:25.941046_PROV_{'entity_id': 519318473740404787, 'name': 'Observations', 'type': 'PythonObject'}
INFO provLogger _PROV_2019-11-18T10:48:25.941174_PROV_{'activity_id': '665a0c', 'used_id': 519318473740404787, 'used_role':
'observations_selected'}
INFO provLogger _PROV_2019-11-18T10:48:25.941341_PROV_{'entity_id': 15337330674484208277, 'name': 'Datasets', 'type': 'PythonObject'}
INFO provLogger _PROV_2019-11-18T10:48:25.941449_PROV_{'activity_id': '665a0c', 'generated_id': 15337330674484208277, 'generated_role':
'reduced_datasets'}
INFO provLogger _PROV_2019-11-18T10:48:25.941574_PROV_{'entity_id': 'stacked', 'name': 'Dataset', 'type': 'PythonObject'}
INFO provLogger _PROV_2019-11-18T10:48:25.941668_PROV_{'entity_id': 15337330674484208277, 'member_id': 'stacked'}
INFO provLogger _PROV_2019-11-18T10:48:25.941835_PROV_{'activity_id': '665a0c', 'endTime': '2019-11-18T10:48:25.940069'}
INFO provLogger _PROV_2019-11-18T10:48:26.495235_PROV_{'activity_id': '1a1ace', 'name': 'set_model', 'startTime': '2019-11-18T10:48:26.245433',
'in_session': 9223372036581382375, 'agent_name': 'test'}
INFO provLogger _PROV_2019-11-18T10:48:26.495513_PROV_{'activity_id': '1a1ace', 'parameters': {'filename': 'model.yaml'}}
INFO provLogger _PROV_2019-11-18T10:48:26.495759_PROV_{'entity_id': 'b5fe131e1d320d9c44adf492a7b14f1d', 'name': 'YAMLFile', 'location':
'model.yaml', 'hash': 'b5fe131e1d320d9c44adf492a7b14f1d', 'hash_type': 'md5', 'type': 'File', 'contentType': 'application/x-yaml'}
INFO provLogger _PROV_2019-11-18T10:48:26.495967_PROV_{'activity_id': '1a1ace', 'used_id': 'b5fe131e1d320d9c44adf492a7b14f1d', 'used_role':
'model_file'}
```

standard files and graph

https://openprovenance.org/store/documents/1852

https://openprovenance.org/store/documents/1852.png

code

https://github.com/Bultako/gammapy/tree/prov/gammapy/utils/provenance