Provenance for MuseWISE

Ole Streicher

ole@aip.de

Victoria, 2018-05-31

MUSE

- Multi Object Spectrographic Explorer (1st light 2014)
- 24 Integral Field Units (Spectrographs)
- Spectral range: 4600-9300 Angstrom, 1.25 A binning, 4000 bins
- Spacial: 300 x 300 pixel
  - Wide Field Mode: 1 x 1 arcmin FoV, 0.2 arcsec sampling
  - Narrow Field Mode (in commissioning): 7.5 arcsec FoV, 0.025 arcsec sampling
- sampling not uniform
- intermediate product: pixel tables
- final science product: rebinned data cubes

MuseWISE

- Based on astroWISE, adopted forMuseWISE
- Standard data reduction framework for MUSE GTO
- Distributed; Python based, Oracle DB
- stores processing Provenance
- integration of instrument specific pipeline

Provenance in MuseWISE

- Data files: generating process, time, input files
- Parameters *partially*
- *QC/QA results*
- **No** project specific (science) processing (outside of MuseWISE)
- using Python `prov` package
- Access:
  - provSAP (ex. *provDAL*): W3C formats (prov-xml, prov-json, prov-n)
  - no provTAP

Provenance volume

- One step back (`depth` = 1)
  - one activity (`muse_scipost`)
  - 55 entities "used", 7 entities "generated"
- "Full" provenance of one science data cube:
  - 269 activities (recipes)
  - 2694 entities (623 generated)
  - `depth` $\approx$ 10
- 24 IFU - 24 individual files and activities in many steps

Provenance experience

- Full provenance useless (no known use case)
- Specific use cases
  - Interactive traversal ("debugging")
  - "What are the input files for the the LSF used in sky subtraction?"
  - Ensure homogenity of a dataset (selected parameters, processing)
- Provenance query "language"? GraphQL?
- Problem: Structuration of data (Hierarchy)
  - data model (W3C, voprov)
  - Framework (MuseWISE) support