PROVENANCE IN FUSION

Applying the W3C-PROV model to fusion data

Nathan Cummings (CCFE)
Frédéric Imbeaux, Jorge Morales and Jean-Françios Artaud (CEA)

Introduction

- Overview of provenance in fusion

- The W3C-PROV standard for provenance

- fusionprov – a demonstration and exploration of applying PROV to fusion data

Provenance in fusion

- Lack of FAIR-ness in fusion makes it difficult to assess the extent of provenance description

- Looking at MAST/-U – provenance information available from log files

- For IDS - data dictionary in development, including recent additions for provenance, but much more to be done

Provenance in fusion – why does it matter?

- It is often hard to know where a signal came from.

- Can you reproduce an analysed signal with just the raw data?

- Do you always know what raw data your signal was derived from?

- What can you tell me (if anything) about the calibration of the device that recorded the raw signal?

W3C-PROV

- Standard for recording and serializing provenance of data (or anything other 'thing')

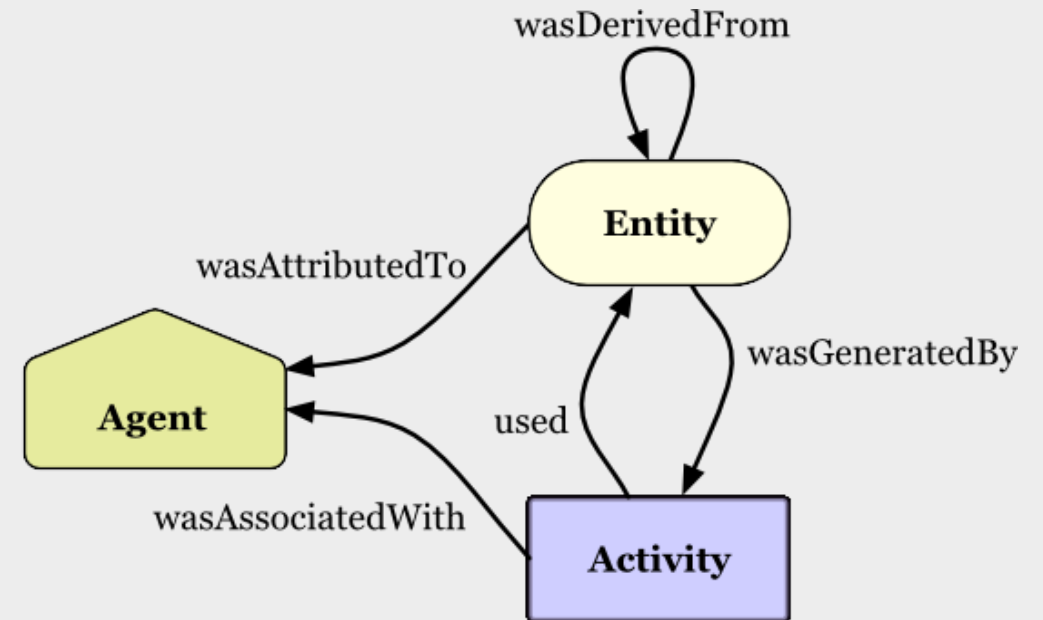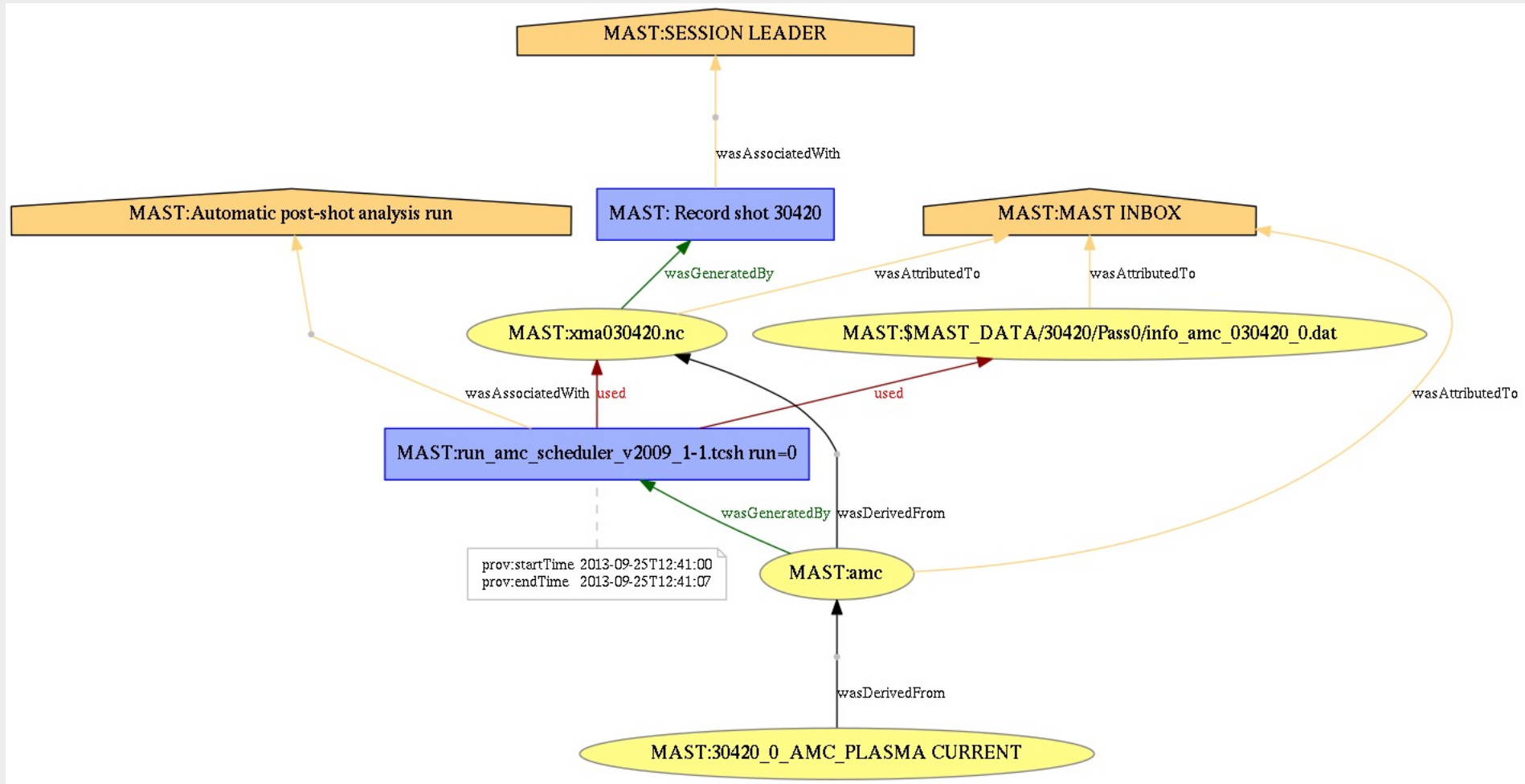- Describes provenance through 'Entity', 'Activity' and 'Agent' nodes, and their relationships
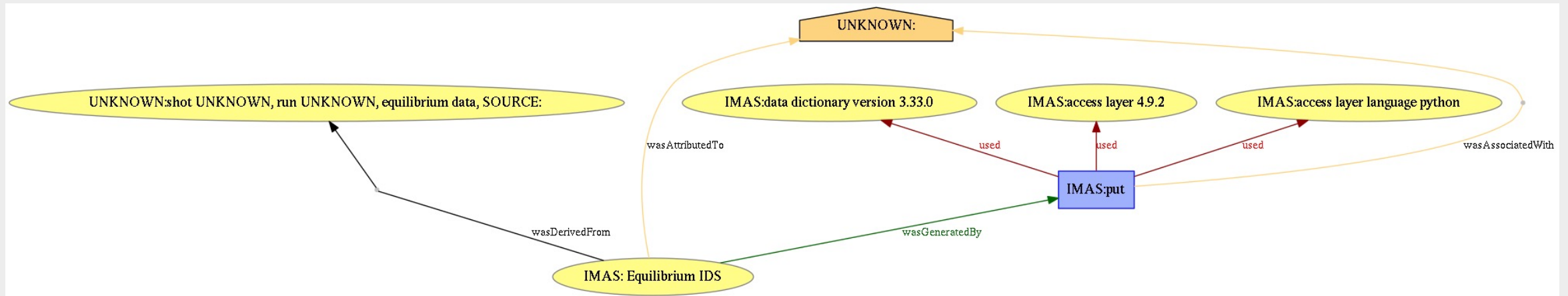
fusionprov

- A python package that generates PROV reports

- Currently supports MAST/-U data and IDS

- Open-source, licensed under Apache 2.0

- GitLab repository – (https://gitlab.com/fair-for-fusion/fusionprov)

fusionprov

fusionprov

Thank you for listening

Fair4Fusion - open access for fusion data in Europe