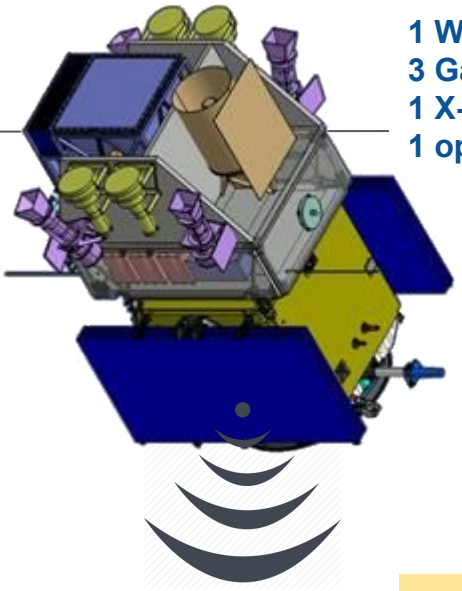

Provenance DM Use Case

The SVOM Ground Segment



- 1 Wide field Gamma Ray camera
- 3 Gamma Ray monitors
- 1 X-ray camera
- 1 optical telescope

X-Band data
Downlink 6 times a day

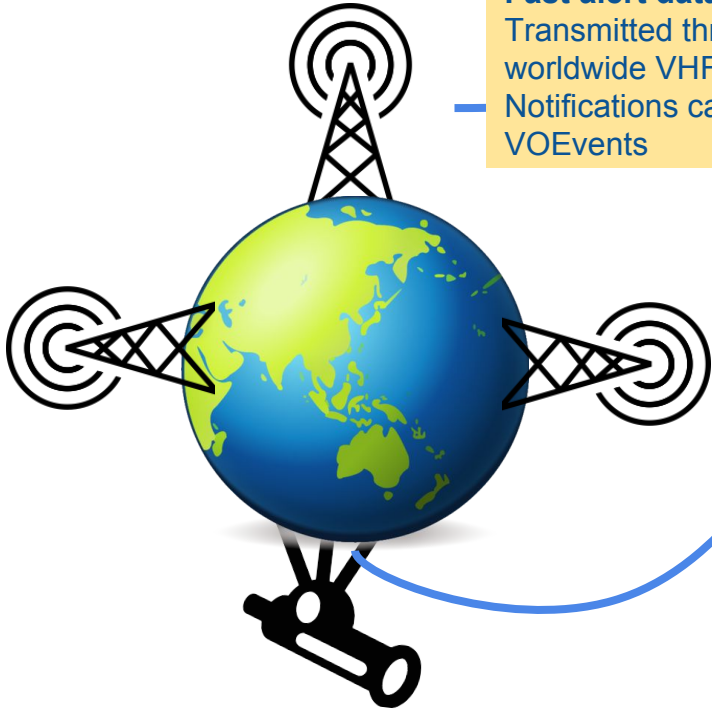
Fast alert data
Transmitted through a worldwide VHF network
Notifications carried out by VOEvents

Follow-up Data
2 ground IR and visible telescopes
China and Mexico
Transmitted through a dedicated link



SUOMI

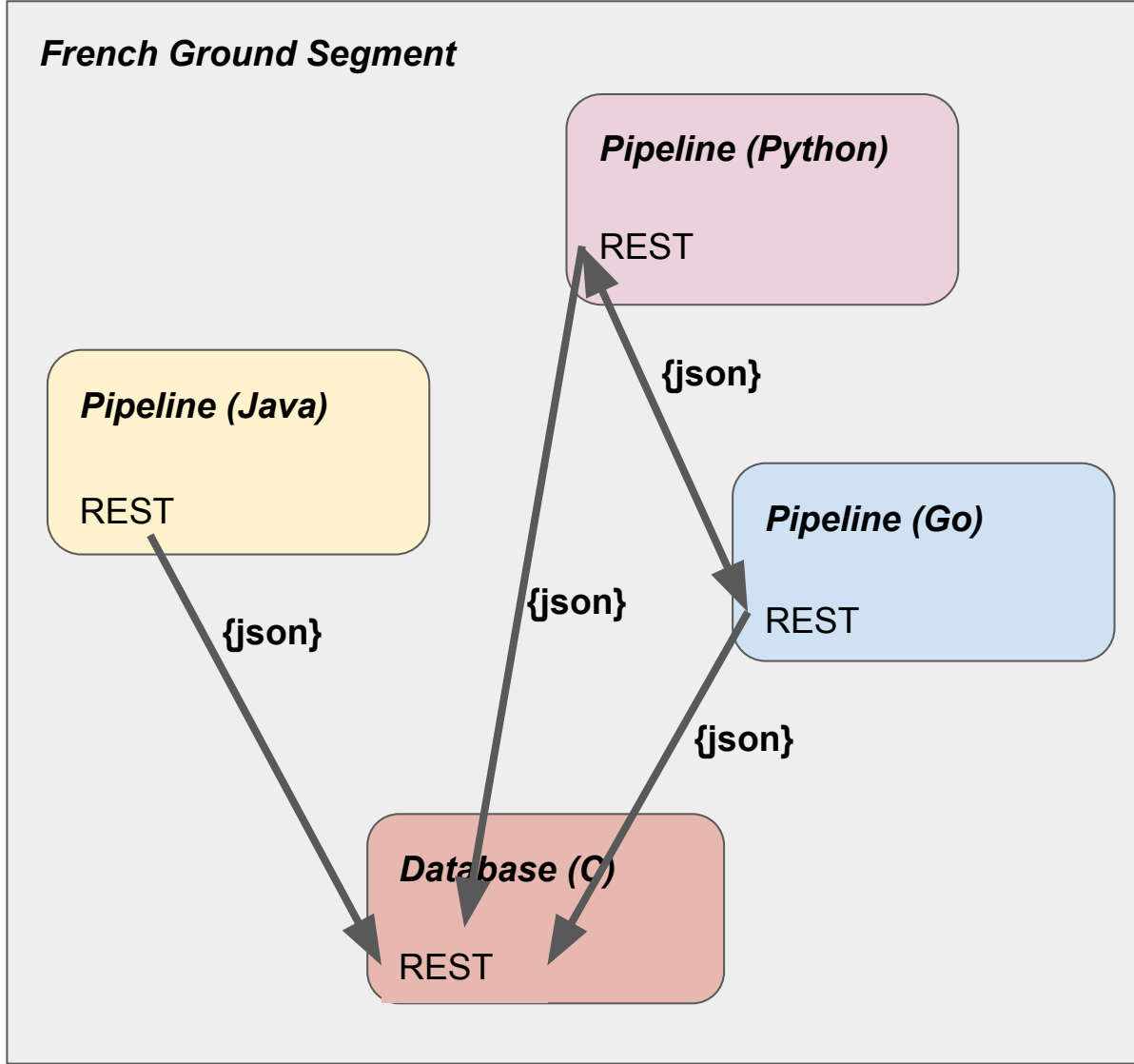
Ground segment
China and France (FSC)



French Ground Segment: Micro Service Arch.

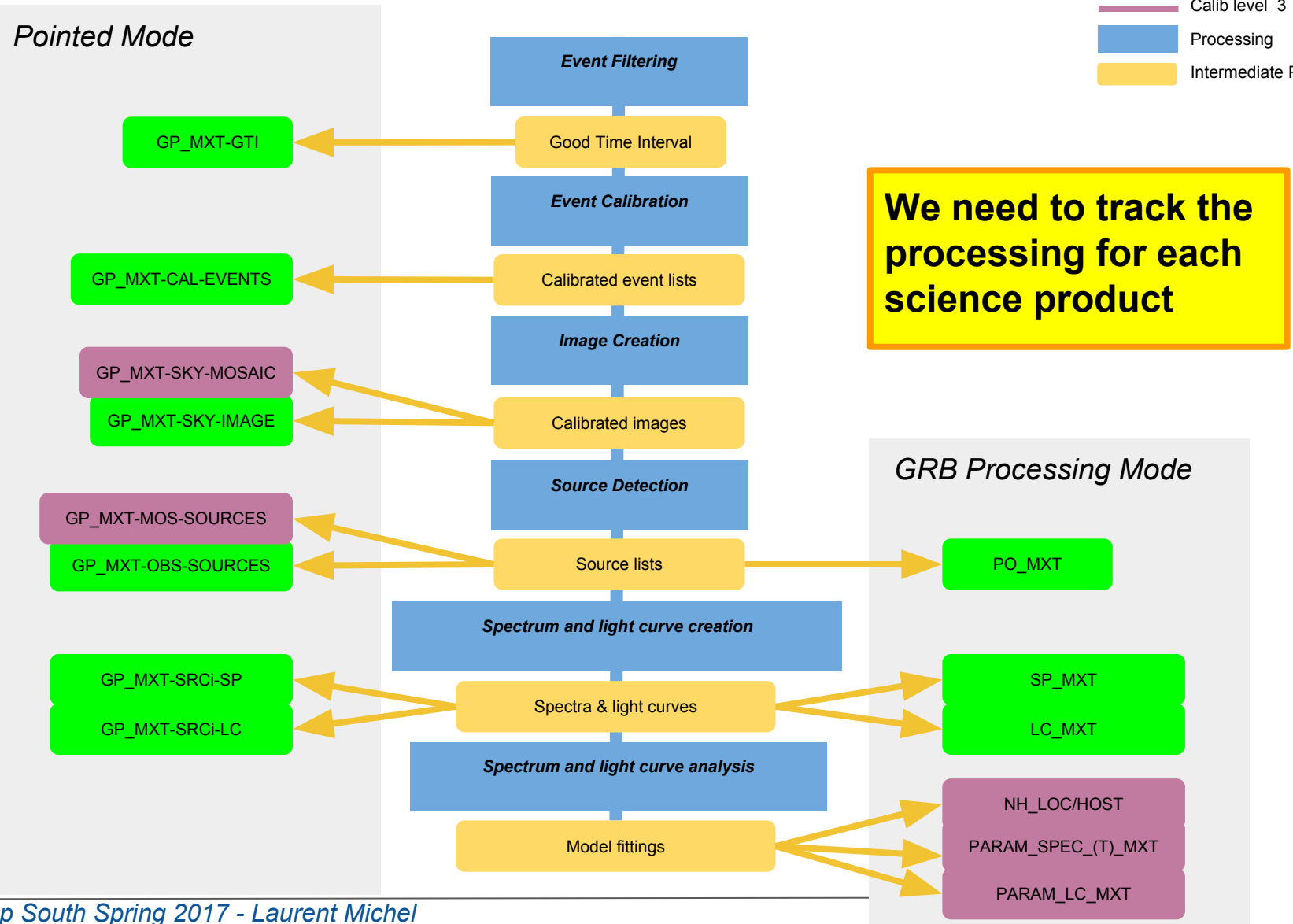


- **Heterogeneous system**
 - Multiple Agents
 - Multiple institutes
 - Multiple languages
- **All module facets under the responsibility of the development team**



Pipeline Workflow

- █ Calib Level 2
- █ Calib level 3
- █ Processing
- █ Intermediate Prds



Why Using Provenance

- **Why tracking the processings**
 - High level requirement
 - Start- point for further re-processings
 - Pipeline tuning

Why Using Provenance

- **Why tracking the processings**

- High level requirement
- Start- point for further re-processings
- Pipeline tuning

- **Why a DM instead of plain text**

- CONTINUE keywords not user/machine friendly

```
CONTINUE '/P0125310101M2U009IMAGE_8000.FIT product/P0125310101M1U009IMAGE_8000.FIT'
CONTINUE '0.FIT' exposuresets=expmap.ds withexposure=no mosaicedset=product/g/
CONTINUE 'P0125310101EPX0000IMAGE8000.FIT sampling-point forceuniformkwds=yes&
CONTINUE '# (mosaic-1.13.2) [xmmsas 20061026 1802-6.6.0]'
XPROC1 = 'addattribute set-product/P0125310101EPX0000IMAGE8000.FIT withattributen&
CONTINUE 'ame='INSTRUME BAND DATAMODE TELESCOP OBS ID DATE-OBS DATE-END OBJE&
CONTINUE 'CT OBSERVER' attributetype='string integer string string string s&
CONTINUE 'tring string string string' withattributecomment='no no no no no &
CONTINUE 'no no no no' attributecomment=''' realvalue=''' stringvalue='EP&
CONTINUE 'IC IMAGING_XMM 0125310101 2000-06-01T10:34:52 2000-06-01T18:58:21 "G&
CONTINUE 'Abell 2690" "Dr Fred Jansen"' integervalue=8 booleanvalue=''' # (&
CONTINUE 'addattribute-1.9.2) [xmmsas 20061026 1802-6.6.0]'
XPROC2 = 'addattribute set-product/P0125310101EPX0000IMAGE8000.FIT attributen&
CONTINUE 'ame="SEQ ID REVOLUT PROCDATE PROCREV PPSVERS SASVERS OSVSER ORIGIN&
CONTINUE 'ODFVER CONTENT"' attributetype='string string string string string&
CONTINUE 'g string string string string string' withattributecomment='yes y&
CONTINUE 'es yes yes yes yes yes yes yes' attributecomment='Pipeline s&
CONTINUE 'equence' "Satellite Revolution Number" "Processing date" "Processin&
CONTINUE 'g revision" "PPS configuration" "SAS version" "ODS version" "Origin&
CONTINUE 'of FITS file" "ODF VERSION" "Contents of file"' realvalue=''' strig&
CONTINUE 'ngvalue="032162 0088 2006-11-08T07:20:50 1 06000518 20061025 152516'
CONTINUE '4 xmmsas 20061026 1802-6.6.0 11.0.1 Leicester/SSC 000 "EPIC OBSERVA&
CONTINUE 'TION IMAGE"' integervalue=''' booleanvalue=''' # (addattribute-1&
CONTINUE '.9.2) [xmmsas 20061026 1802-6.6.0]'
XPROC3 = 'addattribute set-product/P0125310101EPX0000IMAGE8000.FIT attributen&
CONTINUE 'ame="SEQ ID REVOLUT PROCDATE PROCREV PPSVERS SASVERS OSVSER ORIGIN&
CONTINUE 'ODFVER CONTENT"' attributetype='string string string string string&
CONTINUE 'g string string string string string' withattributecomment='yes y&
CONTINUE 'es yes yes yes yes yes yes yes' attributecomment='Pipeline s&
CONTINUE 'equence' "Satellite Revolution Number" "Processing date" "Processin&
CONTINUE 'g revision" "PPS configuration" "SAS version" "ODS version" "Origin&
CONTINUE 'of FITS file" "ODF VERSION" "Contents of file"' realvalue=''' strig&
CONTINUE 'ngvalue="032162 0088 2006-11-08T07:20:50 1 06000518 20061025 152516'
CONTINUE '4 xmmsas 20061026 1802-6.6.0 11.0.1 Leicester/SSC 004 "EPIC OBSERVA&
CONTINUE 'TION IMAGE"' integervalue=''' booleanvalue=''' # (addattribute-1&
CONTINUE '.9.2) [xmmsas 20070301 1824-6.6.0]'
XDAL0 = 'product/P0125310101EPX0000IMAGE8000.FIT 2007-03-03T13:59:25.000 Mod&
CONTINUE 'ify addattribute (addattribute-1.9.2) [xmmsas 20070301 1824-6.6.0] &
CONTINUE 'High SAS_MEMORY_MODEL-high SAS_ROWS= SAS_ZERO_ROWS= SAS_COLUMN_WISE&
CONTINUE '='
XDAL1 = 'product/P0125310101EPX0000IMAGE8000.FIT 2006-11-08T10:40:44.000 Mod&
CONTINUE 'ify addattribute (addattribute-1.9.2) [xmmsas 20061026 1802-6.6.0] &
CONTINUE 'High SAS_MEMORY_MODEL-high SAS_ROWS= SAS_ZERO_ROWS= SAS_COLUMN_WISE=
XDAL2 = 'product/P0125310101EPX0000IMAGE8000.FIT 2006-11-08T08:18:28.000 Mod&
CONTINUE 'ify addattribute (addattribute-1.9.2) [xmmsas 20061026 1802-6.6.0] &
CONTINUE 'High SAS_MEMORY_MODEL-high SAS_ROWS= SAS_ZERO_ROWS= SAS_COLUMN_WISE=
XDAL3 = 'product/P0125310101EPX0000IMAGE8000.FIT 2006-11-08T08:18:20.000 Cre&
CONTINUE 'ate emosaic (emosaic-1.13.2) [xmmsas 20061026 1802-6.6.0] High SAS &
CONTINUE 'MEMORY_MODEL-high SAS_ROWS= SAS_ZERO_ROWS= SAS_COLUMN_WISE='
```

Why Using Provenance

- Why tracking the processings

- High level requirement
- Start- point for further re-processings
- Pipeline tuning

- Why a DM instead of plain text

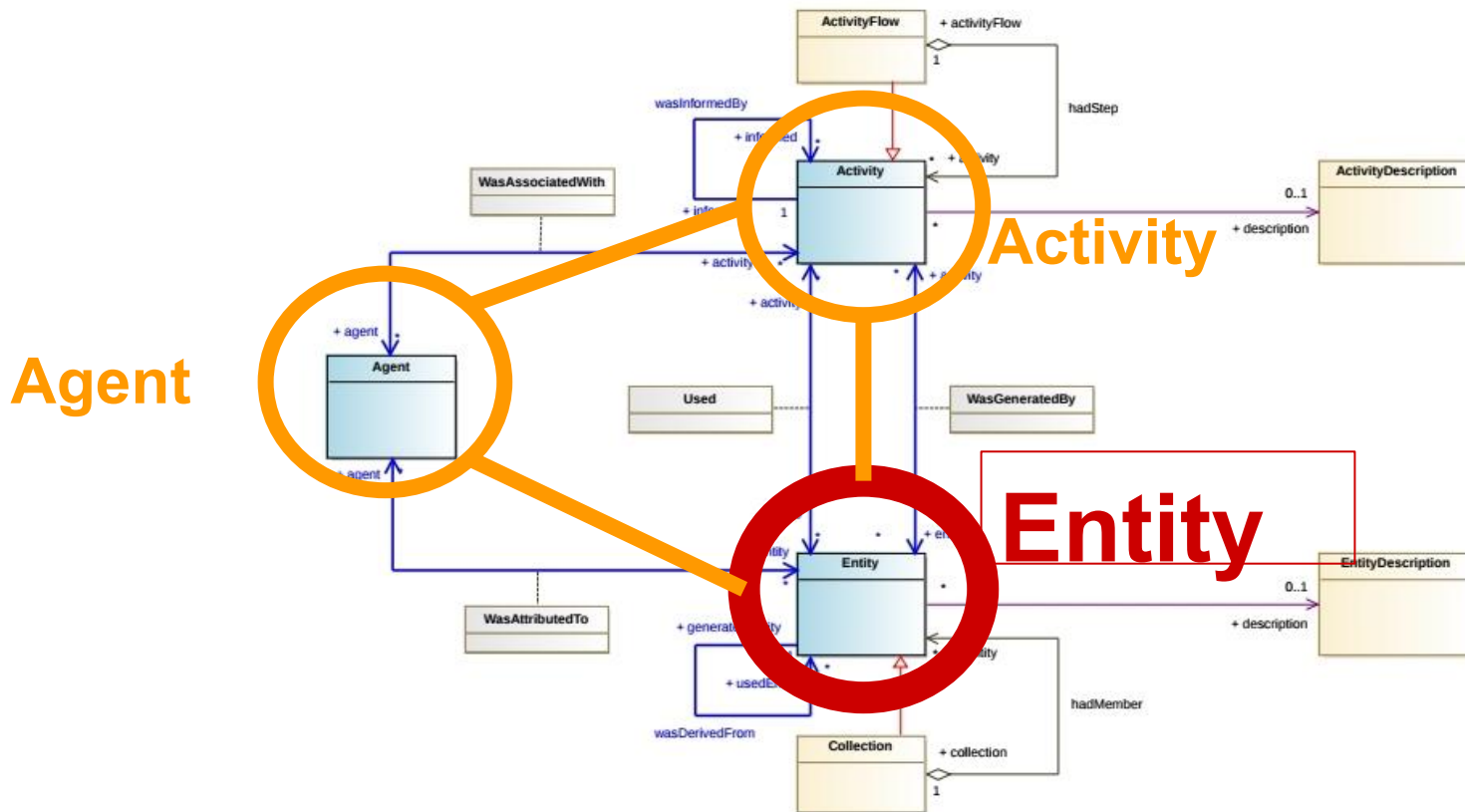
- CONTINUE keywords not user/machine friendly
- A structured representation is helpful
 - Clean layout
 - Machine readable
- Easier to make searchable

```
CONTINUE /P012531010I2M009IMAGE.8000.FIT product/P012531010I1M1009IMAGE
CONTINUE '0.FIT' exposuresets=expmap.ds withexposure=no mosaicdes=prod
CONTINUE 'P012531010I1EPX0000IMAGE8000.FIT sampling=point forceniformkws=
CONTINUE '# (mosaic-1.13.2) [xmsas 20061026 1802-6.6.0]'
CONTINUE 'xproc1 = addattribute set-product/P012531010I1EPX0000IMAGE8000.FIT attrib
CONTINUE 'ame='INSTRUME BAND DATAMODE TELESCOP OBS ID DATE-OBS DATE-END C
CONTINUE 'CT OBSERVER' attributetype='string integer string string stri
CONTINUE 'tring string string string' withattributecoment='no no no no
CONTINUE 'no no no no' attributecoment=''' realsvalue=''' stringvalue=
CONTINUE 'IC IMAGING XMM 0125310101 2000-06-01T10:34:52 2000-06-01T18:58:
CONTINUE 'Abell 2690' 'Dr Fred Jansen'' integervalue=8 booleanvalue='''
CONTINUE 'addattribute-1.9.2) [xmsas 20061026 1802-6.6.0]'
CONTINUE XPROC2 = addattribute set-product/P012531010I1EPX0000IMAGE8000.FIT attrib
CONTINUE 'ame='SEQ ID REVOLUT PROCDATE PROCREV PPSVRS SASVRS ODSVER OR
CONTINUE 'ODFVER CONTENT' attributetype='string string string string s
CONTINUE 'g string string string string' withattributecoment='ye
CONTINUE 'es yes yes yes yes yes yes yes' attributecoment=''' Pipelin
CONTINUE 'equence' 'Satellite Revolution Number' 'Processing date' 'Proce
CONTINUE 'g revision' 'PPS configuration' 'SAS version' 'ODS version' 'Or
CONTINUE 'of FITS file' 'ODF VERSION' 'Contents of file'' realsvalue='''
CONTINUE 'ngvalue="032162 0088 2006-11-08T07:20:50 1 06000518 20061025.1
CONTINUE '4 xmsas 20061026 1802-6.6.0 11.0.1 Leicester/SSC 000 "EPIC OBS
CONTINUE 'TION IMAGE'' integervalue=''' booleanvalue=''' # (addattribu
CONTINUE '.9.2) [xmsas 20061026 1802-6.6.0]'
CONTINUE XPROC3 = addattribute set-product/P012531010I1EPX0000IMAGE8000.FIT attrib
CONTINUE 'ame='SEQ ID REVOLUT PROCDATE PROCREV PPSVRS SASVRS ODSVER OR
CONTINUE 'ODFVER CONTENT' attributetype='string string string string s
CONTINUE 'g string string string string' withattributecoment='ye
CONTINUE 'es yes yes yes yes yes yes yes' attributecoment=''' Pipelin
CONTINUE 'equence' 'Satellite Revolution Number' 'Processing date' 'Proce
CONTINUE 'g revision' 'PPS configuration' 'SAS version' 'ODS version' 'Or
CONTINUE 'of FITS file' 'ODF VERSION' 'Contents of file'' realsvalue='''
CONTINUE 'ngvalue="032162 0088 2006-11-08T07:20:50 1 06000518 20061025.1
CONTINUE '4 xmsas 20061026 1802-6.6.0 11.0.1 Leicester/SSC 004 "EPIC OBS
CONTINUE 'TION IMAGE'' integervalue=''' booleanvalue=''' # (addattribu
CONTINUE '.9.2) [xmsas 20070301 1824-6.6.0]'
XDAL0 = /product/P012531010I1EPX0000IMAGE8000.FIT 2007-03-03T13:59:25.000
CONTINUE 'if addattribute (addattribute-1.9.2) [xmsas 20070301 1824-6.6.
CONTINUE 'High SAS_MEMORY_MODEL=high SAS_ROWS= SAS_ZERO_ROWS= SAS_COLUMN_V
CONTINUE '='
XDAL1 = /product/P012531010I1EPX0000IMAGE8000.FIT 2006-11-08T10:40:44.000
CONTINUE 'if addattribute (addattribute-1.9.2) [xmsas 20061026 1802-6.6.
CONTINUE 'High SAS_MEMORY_MODEL=high SAS_ROWS= SAS_ZERO_ROWS= SAS_COLUMN_V
XDAL2 = /product/P012531010I1EPX0000IMAGE8000.FIT 2006-11-08T08:18:28.000
CONTINUE 'if addattribute (addattribute-1.9.2) [xmsas 20061026 1802-6.6.
CONTINUE 'High SAS_MEMORY_MODEL=high SAS_ROWS= SAS_ZERO_ROWS= SAS_COLUMN_V
XDAL3 = /product/P012531010I1EPX0000IMAGE8000.FIT 2006-11-08T08:18:20.000
CONTINUE 'ate mosaic (mosaic-1.13.2) [xmsas 20061026 1802-6.6.0] High S
CONTINUE 'MEMORY_MODEL=high SAS_ROWS= SAS_ZERO_ROWS= SAS_COLUMN_WISE=
```

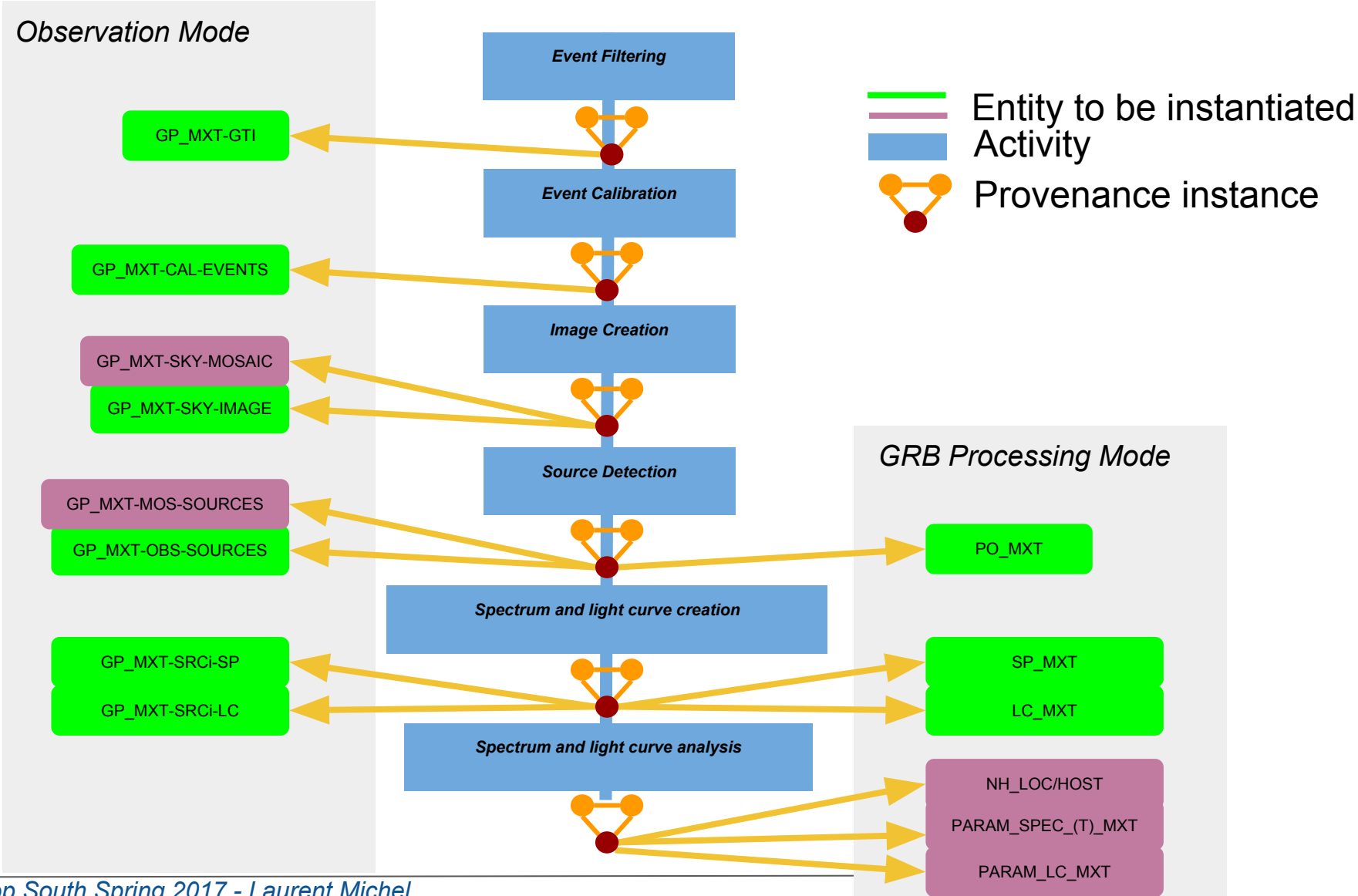
```
{
  "creationTime": "2017-09-08 09:44:30.879485",
  "entityDescription": {
    "annotation": "What a nice Entity",
    "category": "SCIENCE",
    "name": "niceEntity"
  },
  "id": "123",
  "name": "entity1",
  "wasAttributedTo": {
    "agent": [
      {
        "name": "MXTPipeline",
        "type": {
          "individual": "Laurent Michel",
          "organization": "SVOM FSC"
        }
      }
    ]
  },
  "wasGeneratedBy": {
    "activity": [
      {
        "activityDescription": {
          "annotation": "Automatic calibration",
          "name": "CalibTask",
          "subtype": "calibration",
          "type": "processing"
        },
        "annotation": "First Activity",
        "endTime": "2017-09-08 09:44:30.879469",
        "name": "Activity1",
        "parameter": [
          "param1",
          "param2"
        ],
        "startTime": "2017-09-08 09:44:30.879457",
        "wasAssociatedWith": [
          {
            "agent": [
              {
                "name": "MXTPipeline",
                "type": {
                  "individual": "Laurent Michel",
                  "organization": "SVOM FSC"
                }
              }
            ]
          }
        ]
      },
      {
        "activityDescription": {
          "annotation": "Source Detection",
          "name": "DetecTask",
          "subtype": "detection",
          "type": "processing"
        },
        "annotation": "Second Activity",
        "endTime": "2017-09-08 09:44:30.879909",
        "name": "Activity2",
        "parameter": [
          "paramA",
          "paramB"
        ],
        "startTime": "2017-09-08 09:44:30.879895",
        "wasAssociatedWith": [
          {
            "agent": [
              {
                "name": "MXTPipeline",
                "type": {
                  "individual": "Laurent Michel",
                  "organization": "SVOM FSC"
                }
              }
            ]
          }
        ]
      }
    ]
  },
  "time": "2017-09-08 09:44:30.880344"
}
```


Provenance DM at a Glance

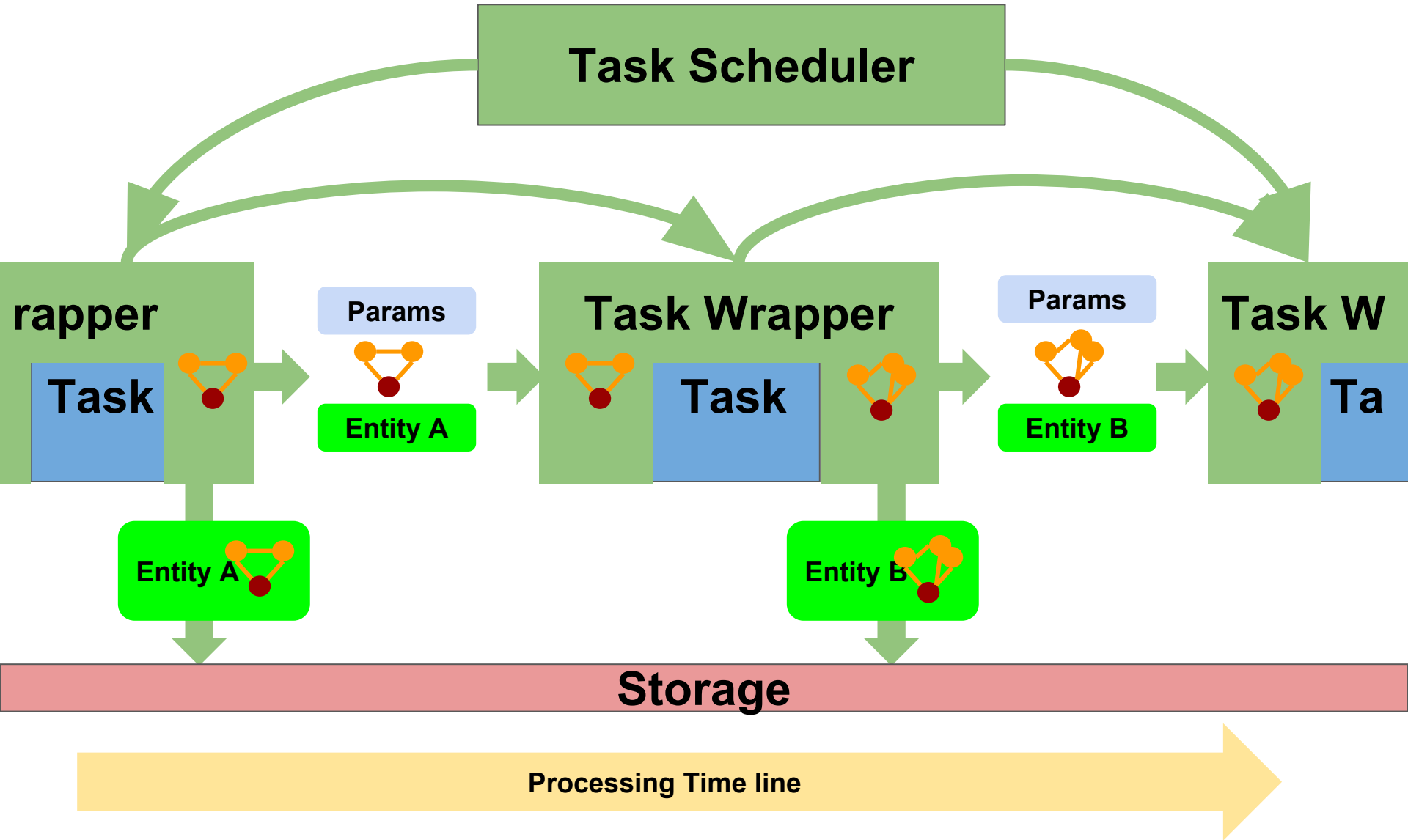
- **Designed around 3 poles**
 - Activity, Agent and Entity
- **Prov Speaking: We want to describe the activities leading to our entities**
 - Entities are final science data files




Provenance View



Incremental Provenance Construction




Development Chain

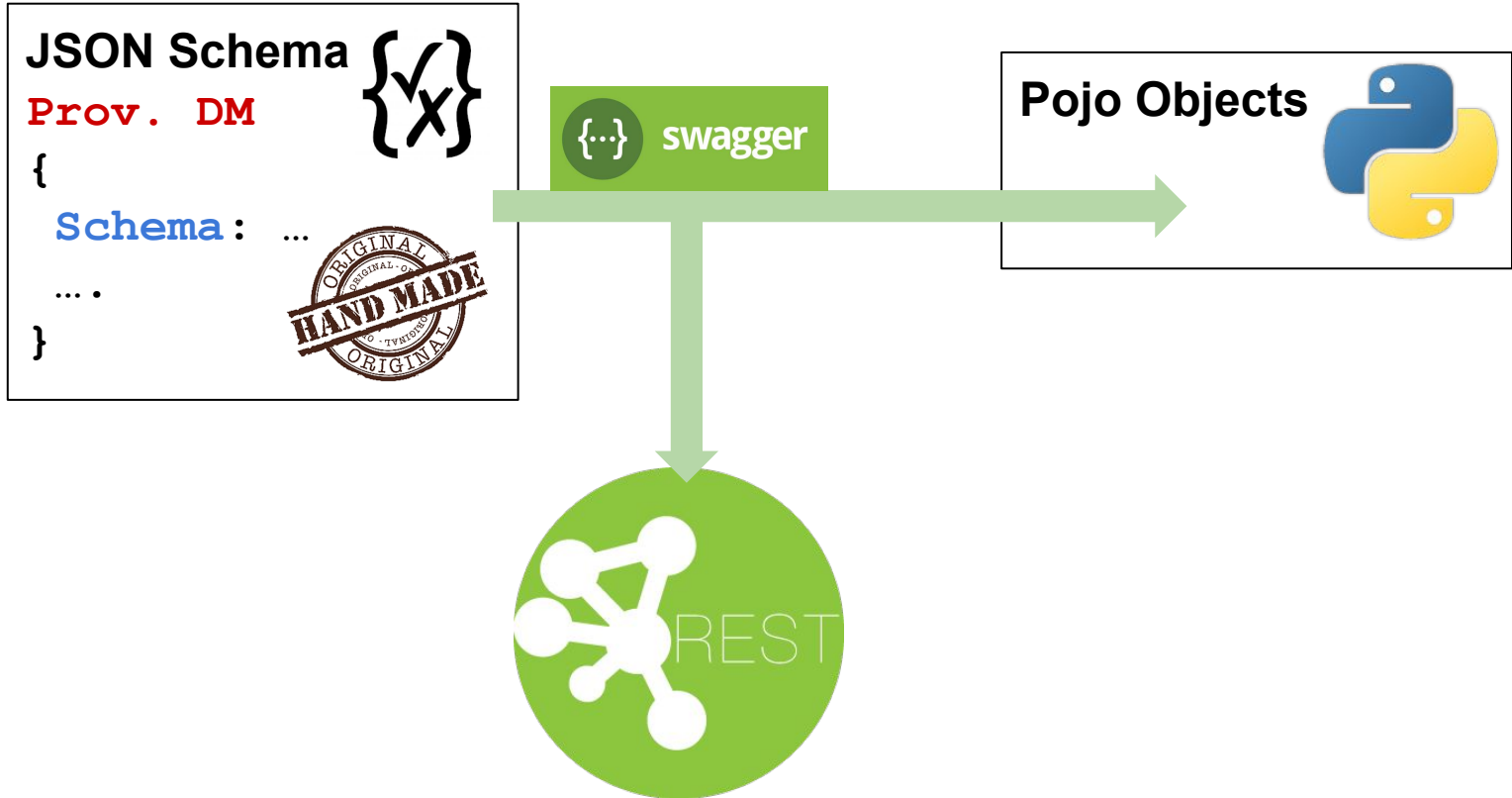
JSON Schema 

Prov. DM

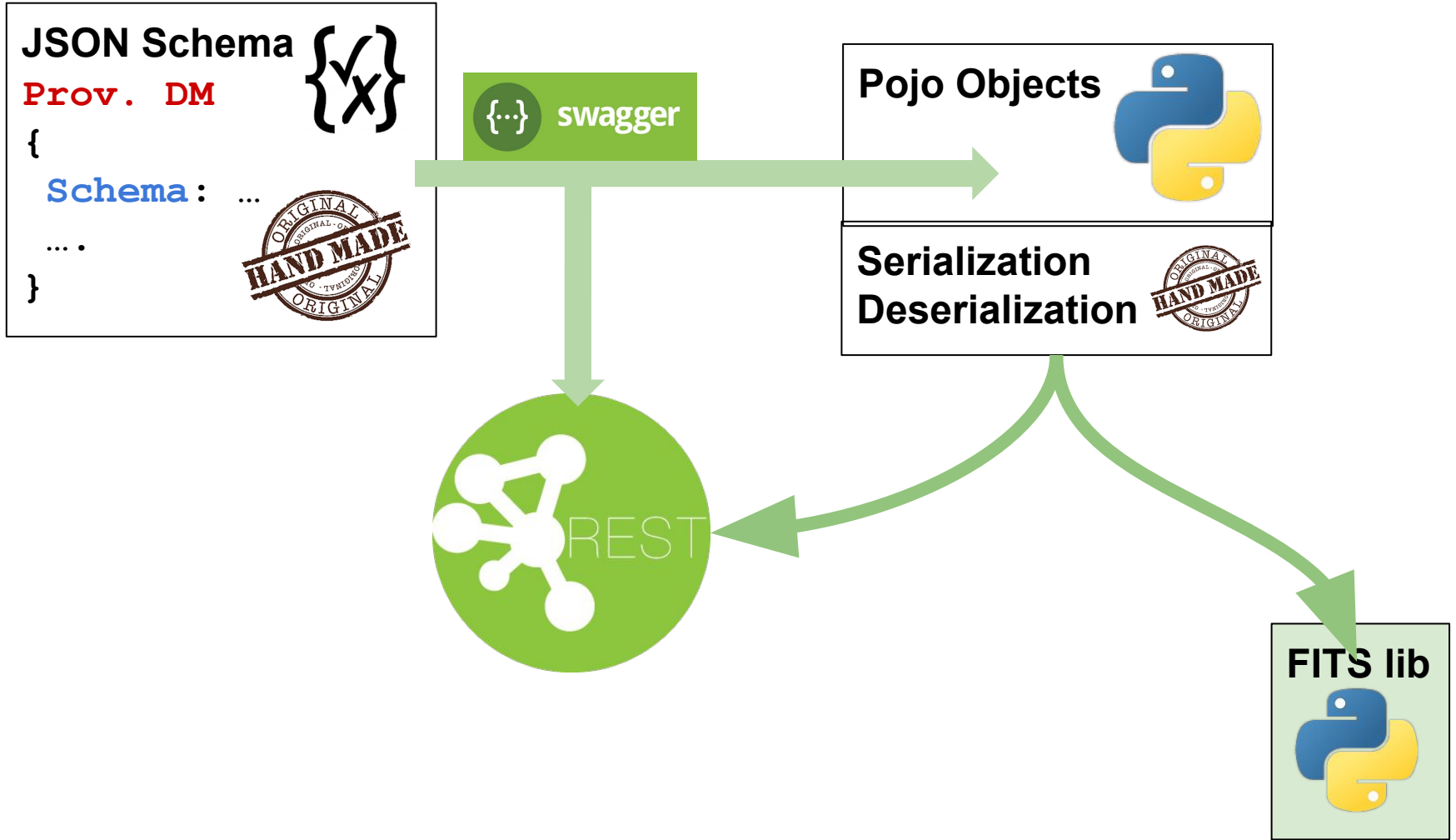
```
{  
  Schema: ...  
  ...  
}
```



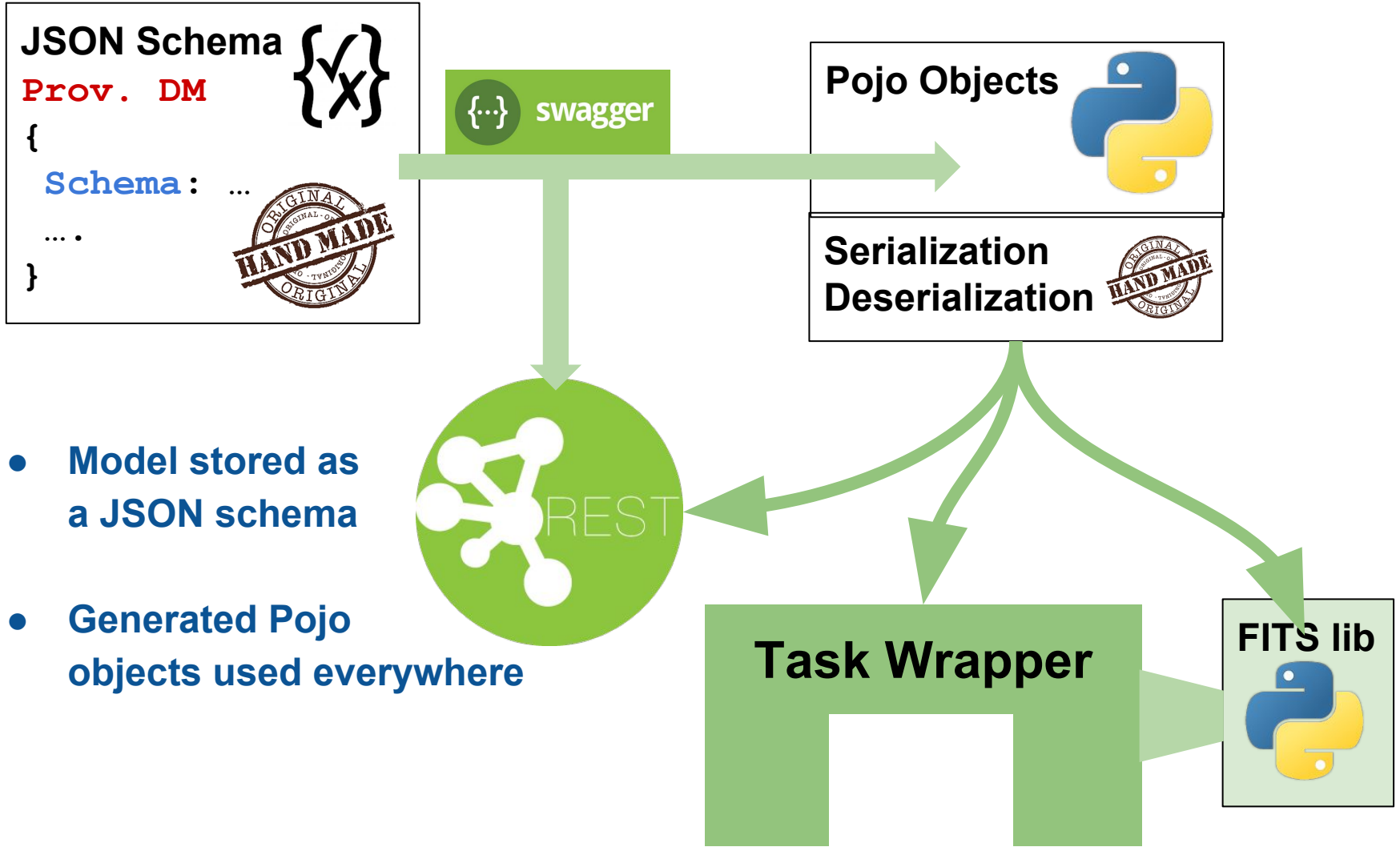
Development Chain



Development Chain



Development Chain



- Model stored as a JSON schema
- Generated Pojo objects used everywhere

Status

- Mechanism validated by the FSC
- Still in Design Phase
 - Modeling tasks as Provenance Activities
 - JSON API
 - FITS API
- On Test
 - Provenance DM expressed as a JSON schema
 - REST endpoint
 - Client stub
 - Server stub
 - Test suite

```
{
  "creationTime": "2017-09-08 09:44:30.879485",
  "entityDescription": {
    "annotation": "What a nice ENTITY",
    "category": "SCIENCE",
    "name": "niceEntity"
  },
  "id": "123",
  "name": "entity1",
  "wasAttributedTo": {
    "agent": [
      {
        "name": "MXTPipeline",
        "type": {
          "individual": "Laurent Michel",
          "organization": "SVOM FSC"
        }
      }
    ]
  },
  "wasGeneratedBy": {
    "activity": [
      {
        "activityDescription": {
          "annotation": "Automatic calibration",
          "name": "CalibTask",
          "subtype": "calibration",
          "type": "processing"
        },
        "annotation": "First Activity",
        "endTime": "2017-09-08 09:44:30.879469",
        "name": "Activity1",
        "parameter": [
          "param1",
          "param2"
        ],
        "startTime": "2017-09-08 09:44:30.879457",
        "wasAssociatedWith": [
          {
            "agent": [
              {
                "name": "MXTPipeline",
                "type": {
                  "individual": "Laurent Michel",
                  "organization": "SVOM FSC"
                }
              }
            ]
          }
        ]
      }
    ]
  },
  "activityDescription": {
    "annotation": "Source Detection",
    "name": "DetectTask",
    "subtype": "detection",
    "type": "processing"
  },
  "annotation": "Second Activity",
  "endTime": "2017-09-08 09:44:30.879909",
  "name": "Activity2",
  "parameter": [
    "paramA",
    "paramB"
  ],
  "startTime": "2017-09-08 09:44:30.879895",
  "wasAssociatedWith": [
    {
      "agent": [
        {
          "name": "MXTPipeline",
          "type": {
            "individual": "Laurent Michel",
            "organization": "SVOM FSC"
          }
        }
      ]
    }
  ]
},
  "time": "2017-09-08 09:44:30.880344"
}
```


Incremental Provenance Construction (followed)

- **Tasks run by wrappers**

- Parameter checking
- Logging setup
- Timestamp management
- Provenance management

- **Provenance instances are part of the workflow**

- A global Prov. instance follows the processing chain
- Each task generates its own local Prov. instance
- Local instances are merged with the global one by the wrappers
- The global instance is transmitted from task to task

- **Provenance meta-data are inserted into archived Entities**

- JSON in FIT Extensions