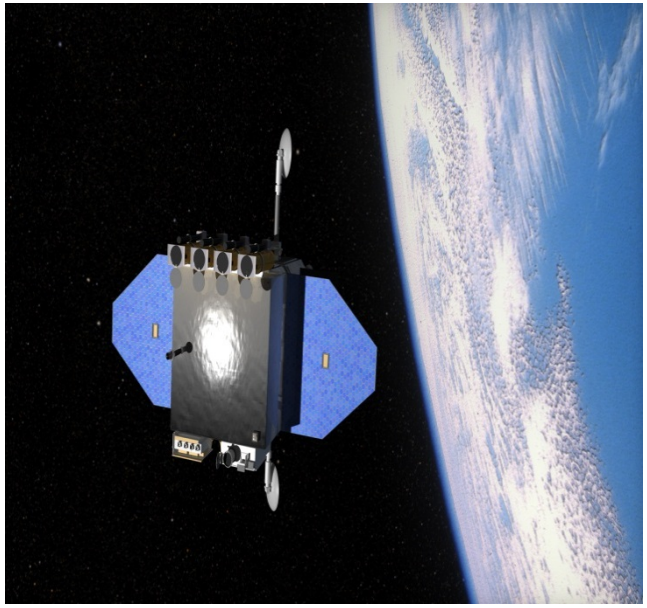
Heliophysics Event Knowledge database: solar events catalogs using VOEvent format

Véronique Delouille

ESCAPE WP4 Provenance Workshop, 8 September 2020

Solar Dynamics Observatory (SDO)



Artist's concept image of the SDO satellite orbiting Earth. Credit: NASA

- SDO is a NASA mission, launched in **2010**. Its EUV telescope, AIA, observes the Sun in 10 wavelengths and returns the equivalent of one 4k x 4k image every second (1TB/day).
- Automated feature-detection methods are essential to helps researchers find data sets relevant for their topics of interest.

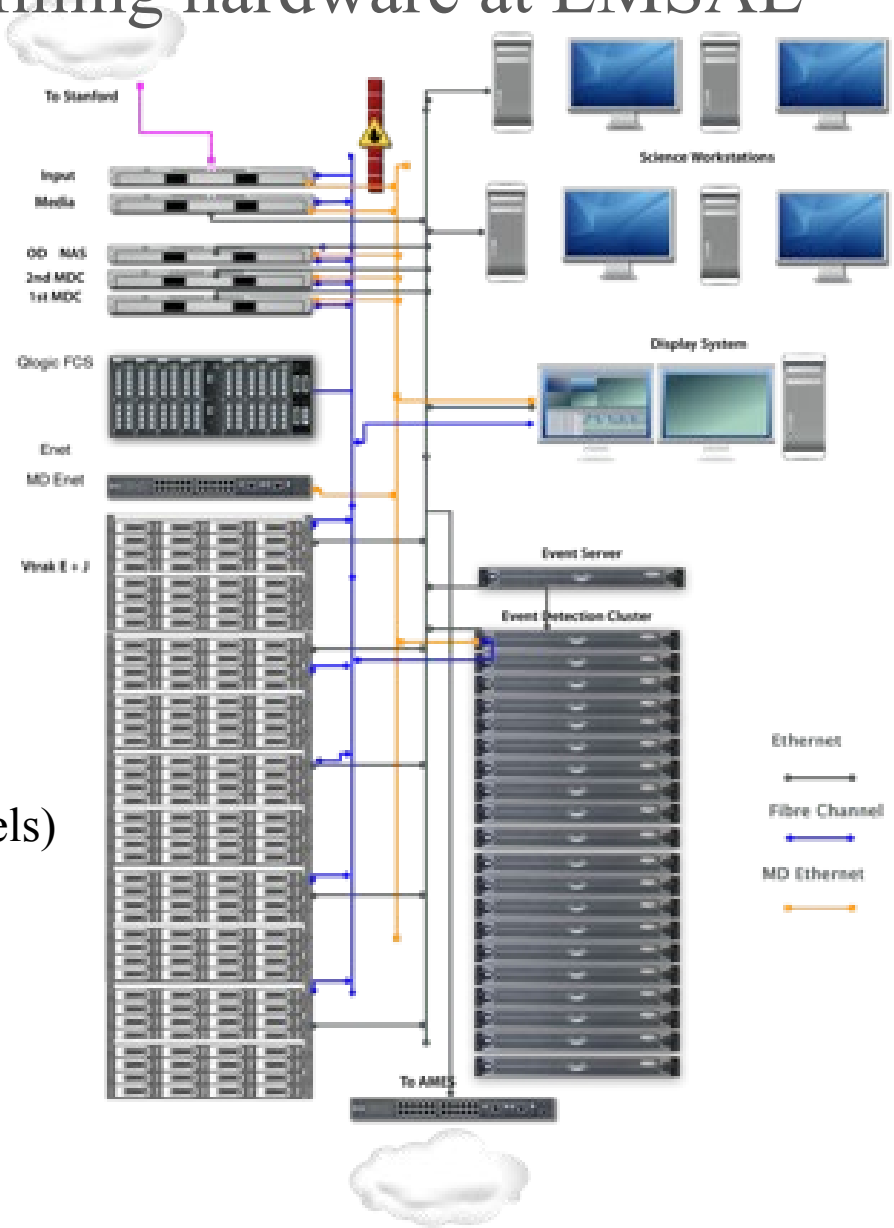
HOW?

- Data mining
  - Automated recognition module for solar events
  - Event Detection System (EDS) to control and manage mining modules
- Data markup
  - Recording and annotating events for later data recall, extracting sample images and movies
  - Capturing mission metadata associated with data, e.g. Planning logs, instrument settings...
- Guided data searches
  - Web and Java clients

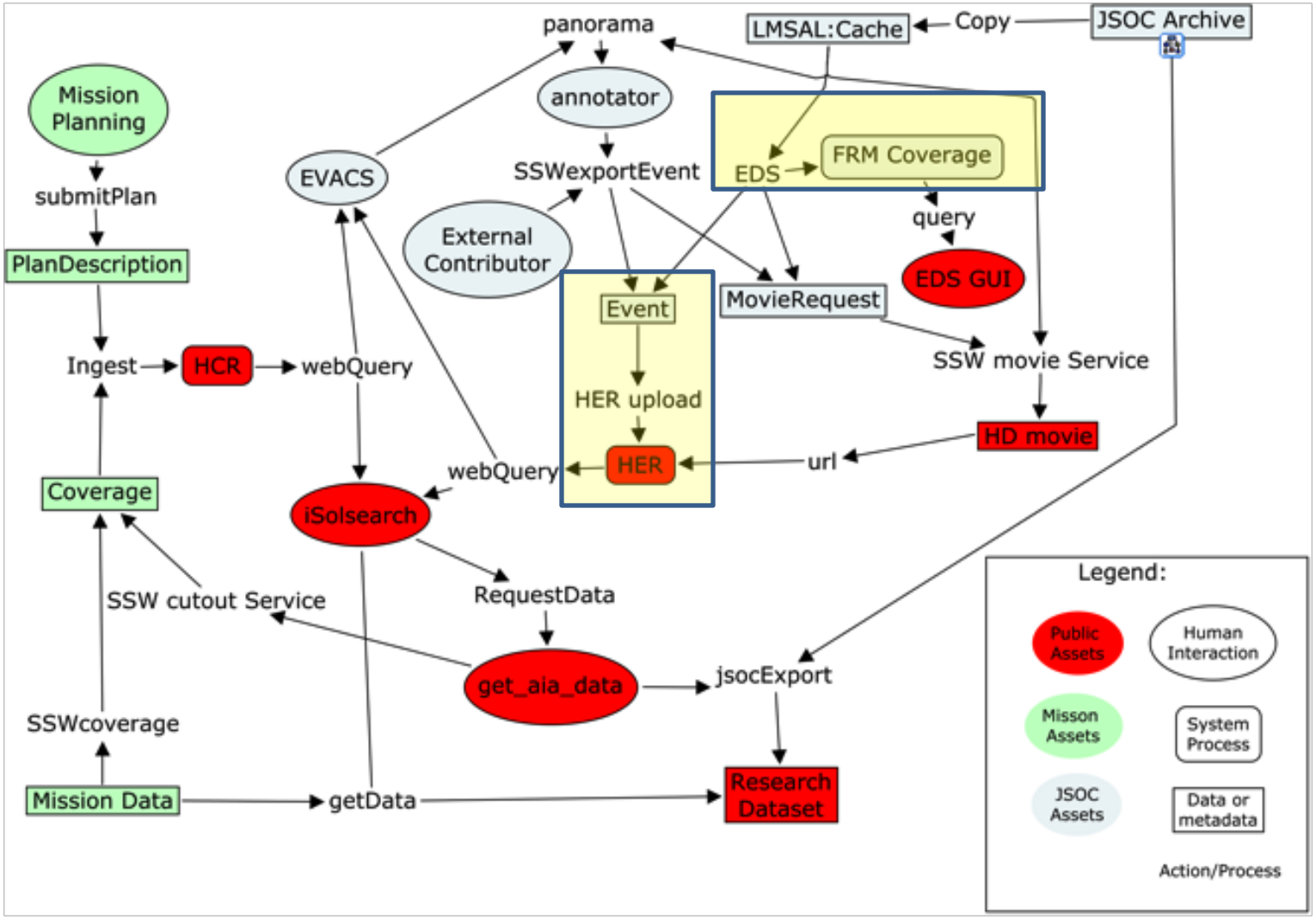
Complex system running at Lockheed Martin

Figure for Data Mining hardware at LMSAL

- **400+TB Apple Xsan**
  - –5TB SAS for database
  - –100TB Cache
  - –300TB user/archive
  - –10Gb link to Stanford
  - –10Gb link to LM science network
  - – Attached Apple and SGI servers
- **SGI Compute servers**
  - –22 node SGI XE Cluster for EDS
  - –SGI UV100 w/72 core, 750GB memory
- **HiPerspace Datawall:**
  - –Quad HD display (3840x2160 pixels)
  - –9-panel 30” display (7680x4800 pixels)
  - –Seven-node CGLX cluster
- **MacPro science workstations**



Heliphysics Event Knowledge base

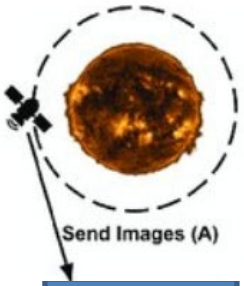


Event classes

- AR = ActiveRegion
- CE = CME
- CD = CoronalDimming
- CH = CoronalHole
- CW = CoronalWave
- FI = Filament
- FE = FilamentEruption
- FA = FilamentActivation
- FL = Flare
- LP = Loop
- OS = Oscillation
- SS = Sunspot
- EF = EmergingFlux
- CJ = CoronalJet
- PG = Plage
- OT = Other
- NR = NothingReported
- SG = Sigmoid
- SP = SpraySurge
- CR = CoronalRain
- CC = CoronalCavity
- ER = Eruption
- TO = TopologicalObject
- HY = Hypothesis
- BU = UVBurst
- EE = ExplosiveEvent
- PB = ProminenceBubble
- …

SPoCA-AR and SPoCA-CH modules

- The two modules for recognition of active regions and coronal holes on the Sun were developed at Royal Observatory of Belgium
- As its core, SPoCA is a fuzzy clustering algorithm that allows to decompose a UV image into regions of **similar intensity**, typically active regions, coronal holes, and quiet sun.
- Different steps: classification of pixels, determination of feature maps (AR, CH), tracking of features
- Output formatted as **VOEvent** *(IVOA standard of 2006),* and used within search tools, image browser, etc via API.



JSOC

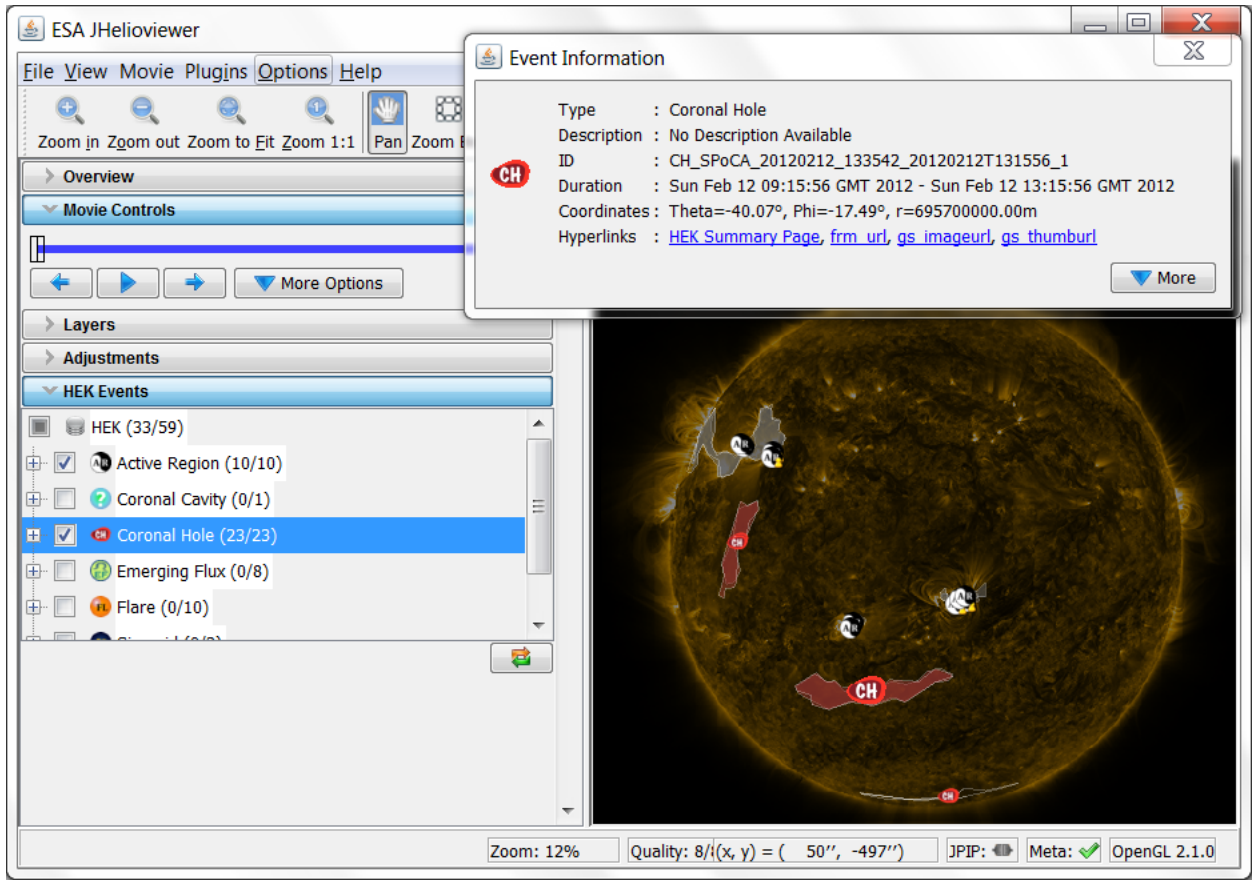
FRM
SPoCA-AR
SPoCA-CH

HEK

Helioviewer,
JHelioviewer

Python
script,
etc…

Screenshot from the ESA JHelioviewer tool.



SPOCA FRM

- Written in C++
- Wrapper for IDL (used in HEK) and for Python
- 3 main activities (execution):
  - Pixel-wise Classification - AIA.CH.classification.config
  - Building of AR and CH maps – AIA.get_CH_map.config
  - Tracking from one map to the next map – AIA.tracking.config
- Main entities:
  - AIA original files
  - Pixel-wise Classification map
  - Region map
- Agents: responsabilities shared between LMSAL (who keeps the system running) and ROB (is responsible for the scientific software)
- Two slightly different versions of SPoCA are also running at ROB.

Success so far and going forward

- Uninterrupted detection of AR and CH for 10years
- Catalog of AR and CH generated
- Used as benchmark e.g. for newer supervised algorithm based on deep learning
- Recurent requests for ‘more information about intermediary products’, such as classification maps, region maps, etc,…

Output of FRM: VOEvent

```
−<voe:VOEvent ivorn="ivo://helio-informatics.org/CH_SPoCA_20190226_020627_20190226T014841_0" role="observation" version="1.1" xsi:schemaLocation="http://www.ivoa.net/xml/VOEvent/v1.1
 http://www.lmsal.com/helio-informatics/VOEvent-v1.1.xsd">
  −<Who>
     <!--Data pertaining to curation-->
    −<AuthorIVORN>
       ivo://helio-informatics.org/CH_SPoCA_20190226_020627_20190226T014841_0
     </AuthorIVORN>
    −<Author>
       <contactName>veronique.delouille@sidc.be</contactName>
     </Author>
     <Date>2019-02-26T02:06:27.741</Date>
   </Who>
  −<What>
     <!--Data about what was measured/observed.-->
     <Description/>
   +<Group name="CoronalHole_optional"></Group>
   </What>
 +<WhereWhen></WhereWhen>
  −<How>
    +<!---->
    +<lmsal:data></lmsal:data>
    +<lmsal:method></lmsal:method>
    +<Group name="CoronalHole_optional"></Group>
   </How>
  −<Why>
     <Inference probability="0.882927"/>
     <Concept>CoronalHole</Concept>
     <lmsal:EVENT_TYPE>CH: CoronalHole</lmsal:EVENT_TYPE>
    +<Group name="CoronalHole_optional"></Group>
   </Why>
   <Reference name="FRM_URL" uri="http://sdoatsidc.oma.be/web/sdoatsidc/SoftwareSPoCA"/>
   <Reference name="OBS_DATAPREPURL" uri="http://sdoatsidc.oma.be/web/sdoatsidc/SoftwareSPoCA"/>
   <Reference name="Edge" type="follows" uri="ivo://helio-informatics.org/CH_SPoCA_20190225_220540_20190225T214841_0"/>
 </voe:VOEvent>
```

Unique ID for the VOEVent

Encoding of tracking of information

‘What’ was measured/observed

```
−<What>
   <!--Data about what was measured/observed.-->
   <Description/>
  −<Group name="CoronalHole_optional">
      <Param name="INTENSMIN" value="3.00058"/>
      <Param name="INTENSMAX" value="17.5034"/>
      <Param name="INTENSMEAN" value="12.3582"/>
      <Param name="INTENSMEDIAN" value="12.5024"/>
      <Param name="INTENSVAR" value="6.62766"/>
      <Param name="INTENSSKEW" value="-0.0578303"/>
      <Param name="INTENSKURT" value="-0.580218"/>
      <Param name="INTENSTOTAL" value="1.38835e+06"/>
      <Param name="INTENSUNIT" value="DN/s"/>
      <Param name="AREA_ATDISKCENTER" value="108293."/>
      <Param name="AREA_ATDISKCENTERUNCERT" value="9963.24"/>
      <Param name="AREA_RAW" value="31411.6"/>
      <Param name="AREA_UNCERT" value="3596.68"/>
      <Param name="AREA_UNIT" value="Mm2"/>
      <Param name="EVENT_NPIXELS" value="169263"/>
      <Param name="EVENT_PIXELUNIT" value="DN/s"/>
   </Group>
 </What>
```

Global measured quantities for CH

Note: Associated UCD would be good
To have

Where/when: Data pertaining to when(time) and where(location on the Sun) something occured

```
    Data pertaining to when and where something cooured
  -->
-<ObsDataLocation>
 -<ObservatoryLocation>
    <AstroCoordSystem/>
    <AstroCoords id="UTC-HPC-TOPO" coord_system_id="UTC-HPC-TOPO"/>
  </ObservatoryLocation>
 -<ObservationLocation id="SDO">
    <AstroCoordSystem/>
   -<AstroCoords coord_system_id="UTC-HPC-TOPO">
     -<Time>
       -<TimeInstant>
          <ISOTime>2019-02-25T21:48:41.840</ISOTime>
        </TimeInstant>
      </Time>
     -<Position2D unit="arcsec,arcsec">
       -<Value2>
          <C1>-36.9470</C1>
          <C2>-861.919</C2>
        </Value2>
       -<Error2>
          <C1>1.22704</C1>
          <C2>15.0041</C2>
        </Error2>
      </Position2D>
    </AstroCoords>
   -<AstroCoordArea coord_system_id="UTC-HPC-TOPO">
     -<TimeInterval>
       -<StartTime>
          <ISOTime>2019-02-25T21:48:41.840</ISOTime>
        </StartTime>
       -<StopTime>
          <ISOTime>2019-02-26T01:48:41.000</ISOTime>
        </StopTime>
      </TimeInterval>
     -<Box2>
       -<Center>
          <C1>-43.5000</C1>
          <C2>-850.500</C2>
        </Center>
       -<Size>
          <C1>812.400</C1>
          <C2>234.000</C2>
        </Size>
      </Box2>
    </AstroCoordArea>
  </ObservationLocation>
</ObsDataLocation>
-<Group name="CoronalHole_optional">
    <Param name="EVENT_CLIPPEDSPATIAL" value="T"/>
    <Param name="BOUND_CCNSTEPS" value="31"/>
    <Param name="BOUND_CCSTARTC1" value="-449.700"/>
    <Param name="BOUND_CCSTARTC2" value="-854.700"/>
    <Param name="BOUND_CHAINCODE" value="-449.700,-854.700,-449.100,-854.100,-409.500,-852.900,-318.900,-894.300,-287.100,-873.900,-291.900,-838.500,-268.500,-
    770.100,-235.500,-761.700,-207.300,-790.500,-174.300,-795.300,-154.500,-825.900,-74.700,-830.100,-6.900,-767.700,89.700,-734.700,159.900,-733.500,246.30[illegible]
    804.900,175.500,-786.300,123.900,-823.500,159.900,-852.900,220.500,-851.100,240.300,-876.900,170.700,-873.300,152.100,-906.900,198.300,-912.300,228.90[illegible]
    903.900,362.700,-896.700,147.900,-956.100,-57.300,-965.700,-258.300,-932.100"/>
    <Param name="CHAINCODETYPE" value="ordered list of points in HPC"/>
  </Group>
</WhereWhen>
```

Trick to encode tracking

<StartTime>: you need to say when did the event occur, but for a long lived feature you do not know, hence you report the previous time when events were reported

<StopTime = >T_OBS in the FITS file in which CH was detected, at time *t*

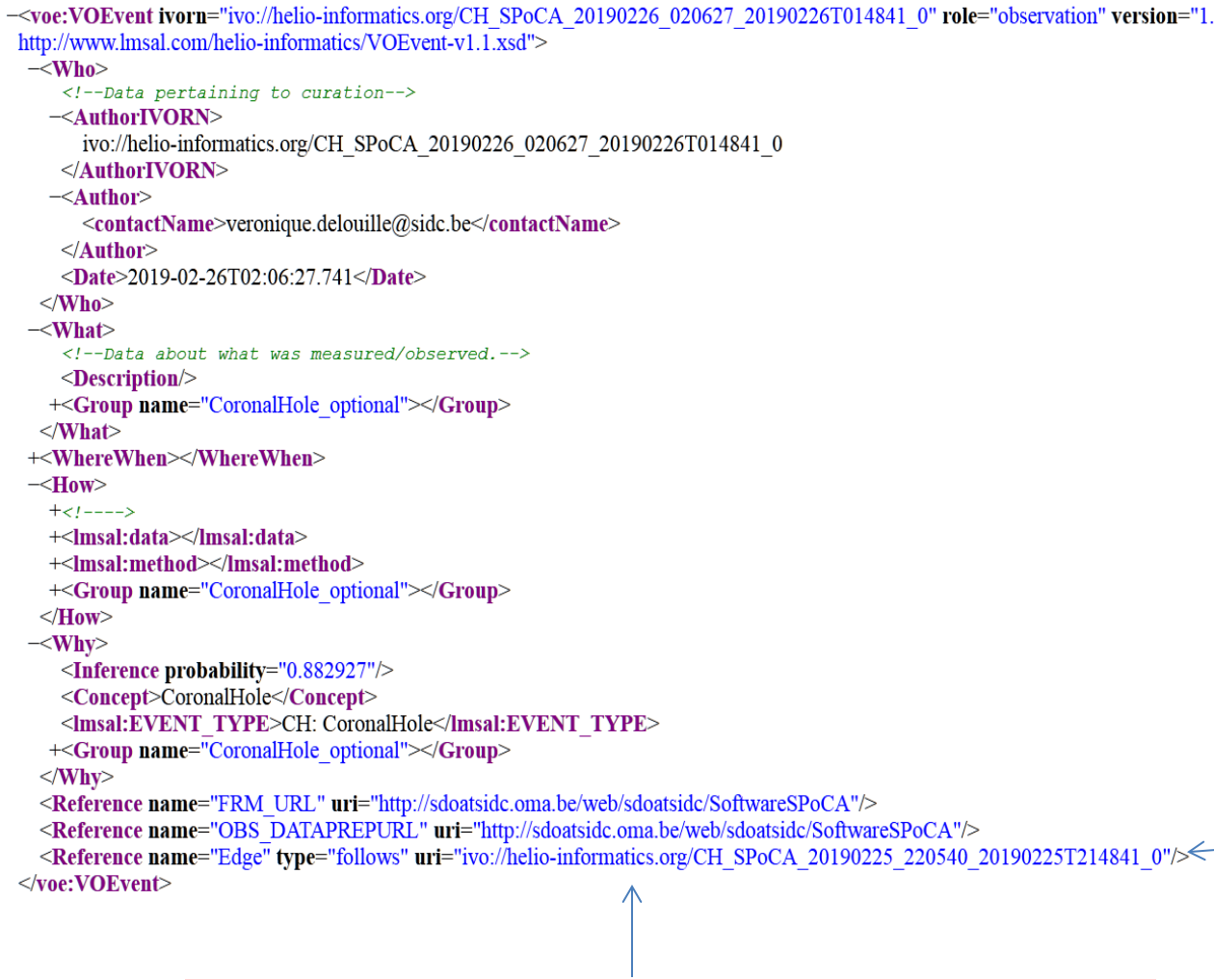
From this time, it may be possible to recover initial AIA Filename, here : AIA.20190226_014840.0193.image_lev1.fits

Chain code to report the boundaries of CH = Approximation of information available in region maps !

How was the file produced : information about SPoCA method

```
−<How>
  +<!---->
  −<lmsal:data>
     <lmsal:OBS_ChannelID>AIA 193</lmsal:OBS_ChannelID>
     <lmsal:OBS_Instrument>AIA</lmsal:OBS_Instrument>
     <lmsal:OBS_MeanWavel>193.000</lmsal:OBS_MeanWavel>
     <lmsal:OBS_WavelUnit>Angstroms</lmsal:OBS_WavelUnit>
   </lmsal:data>
  −<lmsal:method>
     <lmsal:FRM_Contact>veronique.delouille@sidc.be</lmsal:FRM_Contact>
     <lmsal:FRM_DateRun>2019-02-26T02:06:27.714</lmsal:FRM_DateRun>
     <lmsal:FRM_HumanFlag>F</lmsal:FRM_HumanFlag>
     <lmsal:FRM_Identifier>vdelouille</lmsal:FRM_Identifier>
     <lmsal:FRM_Institute>ROB</lmsal:FRM_Institute>
     <lmsal:FRM_Name>SPoCA</lmsal:FRM_Name>
    −<lmsal:FRM_ParamSet>
       image195 : calibrated image 193/195 A; spocaPreprocessing=DivExpTime,ALC,ThrMax80,TakeSqrt; spocaClassifierType=HFCM; spocaNumberclasses=4; spocaChannels=
       [AIA_193]; spocaPrecision=0.00150000; spocaRadiusRatio=1.20; spocaBinsize=0.0100000; spocaSegmentationType=max; spocaVersion=2.00;
       intensitiesStatsPreprocessing=NAR,DivExpTime; intensitiesStatsRadiusRatio=0.95; trackingDeltat=36000; trackingOverlap=2; trackingNumberImages=6; minLifeTime=259200
       minDeathTime=28800; spocaCenters=(2.9581),(5.4468),(7.2253),(8.9175)
     </lmsal:FRM_ParamSet>
   </lmsal:method>
  −<Group name="CoronalHole_optional">
     <Param name="FRM_VERSIONNUMBER" value="1.00000"/>
     <Param name="FRM_SPECIFICID" value="SPoCA_v1.0_CH_0000029662"/>
     <Param name="OBS_DATAPREPURL" value="http://sdoatsidc.oma.be/web/sdoatsidc/SoftwareSPoCA"/>
     <Param name="OBS_LASTPROCESSINGDATE" value="2019-02-26T01:59:39"/>
     <Param name="OBS_LEVELNUM" value="1.50000"/>
     <Param name="OBS_INCLUDESNRT" value="T"/>
   </Group>
 </How>
```

Description of activities (but all mixed up, ie parameters for classification, getting AR maps, and tracking all together)



Encoding of tracking of information

In a second pass, these information are linked together to form a meta-event, e.g. indicating a CH which may last several days on the solar disk.

Questions

- Existing software, combination of C++, IDL, Python. One version at LMSAL (little control on it), one version at ROB (more control)
- How to structure better provenance information within VOEvent files?
- How to get a provenance DM with an 'on top' approach ?
- How to proceed if one wants to save provenance info during the processing ('inside' approach)

Concerns

- Scientific developers have a lot on their plate
- Hierarchy will balance the benefit of having a provenance DM implemented vs the time it costs
- Need to clearly see the benefits, and what it brings in addition to a well documented workflow
- Learning curve.