Provenance workshop questionnaire

Markus Demleitner msdemlei@ari.uni-heidelberg.de **- 20/07/2020**

1. Name of the project? (it could be a large project, a subproject, or a simple task)

> **GAVO** – and I'm speaking as DaCHS author, mostly.

2. (Data) products provided by the project?

> Rougly: Everything.

3. Is the capture of provenance a requirement for your project?

> Let's say: I'd like to do it better than it's done at this point. My pet project would be to machine-readably link photometric points with the pixels that generated them.

4. Do you already provide provenance information? Do you use the IVOA Provenance data model ?

> Well, there's obviously a lot of human-readable docs on how things came to be, and we're using datalink #progenitor and #derivation links here and there. And then there is wild, sometimes really extensive, semi-structured provenance info in lots of the FITS files we serve.
> It's painful and non-standard throughout.

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
1 other: help end-users figure out flukes in reduced data
2 other: machine-readably declare observation (or other
  experiment) conditions.
3 acknowledge people
4 keep the traceability of the data products
5 provide more information to the final user
6 find contact information
7 reproduce or reprocess the data
8 ensure quality and reliability
9 debug a pipeline
```

> (the pipeline debugging is at the last position because I'd say that's done by the people writing the pipeline, and thus there's little need for *interoperable* provenance in that case).

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

> When I muck around with FITSes, I generally add a HISTORY card.
> Beyond that: README files.

Jutta Schnabel jutta.schnabel@fau.de - 21/07/2020

À Paschal, Damien, Kay, Tamas, Mathieu

Dear Mathieu,
Find below the answers from KM3NeT on your questionnaire. We are looking forward to the workshop in September!
Cheers,
Jutta

Questions:

1. Name of the project? (it could be a large project, a subproject, or a simple task)

> **KM3NeT**

2. (Data) products provided by the project?

> tables of particle measurements
> tables of simulated events, derived histograms and functions

3. Is the capture of provenance a requirement for your project?

> Considering the requirement for FAIR data processing, yes.

4. Do you already provide provenance information? Do you use the IVOA Provenance data model ?

> Working on the implementation. We are drawing on the IVOA Provenance data model without, currently, strictly implementing it.

5. Which goals should the addition of provenance information serve (if possible by order of priority)? (scale 1 to 5 from low priority to high, 0 for not considered a goal)

```
3 provide more information to the final user
5 keep the traceability of the data products
4 ensure quality and reliability
0 find contact information
2 acknowledge people
4 reproduce or reprocess the data
3 debug a pipeline
1 other: facilitate a visual representation of processing pipelines
```

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

- for high-level data (tier-2)
  * capturing program configuration and runtime environment
  * storage inside processed file
- for tier-0 and tier-1: to be implemented

Mireille LOUYS mireille.louys@unistra.fr - 22/07/2020

1. Name of the project? (it could be a large project, a subproject, or a simple task)

PROV-Hips
describes the activities , image dataproducts and agents involved in the elaboration of HST HiPS survey, as delivered by CADC / cf D. Durand

2. (Data) products provided by the project?

HiPS image collection , access from one tile to image dataproducts distributed at HST CADC archive

3. Is the capture of provenance a requirement for your project?

This is an improvement on top of an existing archive to trace data transformation
this project is a provenance add-on for HiPS survey .
It is built after all data have been generated and as such is a 'a posteriori' provenance representation

4. Do you already provide provenance information? Do you use the IVOA Provenance data model ?

yes we provide Activities with their description , image datasets in Hips and access to progenitors

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
2 provide more information to the final user
1 keep the traceability of the data products
3 ensure quality and reliability
_ find contact information
5 acknowledge people
4 reproduce or reprocess the data
6 debug a pipeline
_ other:
```

NB: contact and credits are given at the archive level and not realy traceable in detail for each operation inside HST pipeline
credits and contact for Hips creation was created from Prov DM agents and appropriate relations

6. How do you consider or already do the capture, storage and/or

> distribution of provenance information?
> Provenance info was extracted from the image fits headers.
> Activity properties also with guidance from D.Durand

Mark Kettenis kettenis@jive.eu - 27/07/2020

Here are my answers to the questionnaire. My collegue Harro Verkouter might have a slightly different view on things, but he is currently on holiday. He might send you an updated version when he is back in early august.

Regards,
Mark

Questions:

1. Name of the project? (it could be a large project, a subproject, or a simple task)

> **"EVN Archive"**

2. (Data) products provided by the project?

> FITS-IDI files containing visibility data
> Calibration tables to be used with these FITS-IDI files
> Pipeline plots/images

3. Is the capture of provenance a requirement for your project?

> To some extent.

4. Do you already provide provenance information? Do you use the IVOA Provenance data model ?

> Not really.

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
X keep the traceability of the data products
X ensure quality and reliability
X provide more information to the final user
X reproduce or reprocess the data
```

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

> Part of the process that produces FITS-IDI files is tracked by keeping information in intermediate files and databases. This includes the version of the correlator software used to process the data, corelation parameters, a list of the correlation jobs that were included in the final data product.
> It is anticipated that our new processing pipeline that generates the final data product will also produce data records that can be translated into the IVOA provenance model.

```
Harro Verkouter <verkouter@jive.eu>
3 août 2020 12:24
À Mark, Servillat

Hi Mathieu, Mark,

Mark has captured our current situation quite well.

If anything could be said different it would be that, according to my
interpretation of our ESCAPE WP5 activities, provenance is more than
"to some extent".

Within WP5 our goal is to be able to re-run a pipeline and reproduce
results, and the approach we are investigating is one where the pipeline
script /is/ the provenance - e.g. through the use of versioned Jupyter
notebooks kept in a public repository. The data processing steps
and their parameters are both stored in such a notebook.

At this point I'm not sure that that would be a 1:1 match with the
provenance data model or where the boundary between "provenance DM"
and "pipeline script" actually lies. As such Mark has said it well
when he mentions that it might be possible in the future to translate
our data records into the Provenance DM.

Cheers,
h
```

Yan Grange grange@astron.nl & Mattia Mancini mancini@astron.nl - 27/07/2020

Pour : François Bonnarel francois.bonnarel@astro.unistra.fr, Mattia Mancini mancini@astron.nl, Mark Kettenis kettenis@jive.eu, Harro Verkouter verkouter@jive.eu, Marco Iacobelli iacobelli@astron.nl

Hi François,

Mattia and myself (well mostly Mattia) answered the questionnaire:

1. Name of the project? (it could be a large project, a subproject, or a simple task)

LOFAR
APERTIF
several WSRT surveys

2. (Data) products provided by the project?

MeasurementSets (interferometric visibilities)
Calibration solutions (in hdf5)
image/cubes fits
beam formed observations (in hdf5)
pulsar profiles (‘timer’ format)

3. Is the capture of provenance a requirement for your project?

yes

4. Do you already provide provenance information? Do you use the IVOA Provenance data model ?

yes but without using the ivoa provenance data model

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
X provide more information to the final user
X keep the traceability of the data products
X ensure quality and reliability
_ find contact information
_ acknowledge people
X reproduce or reprocess the data
X debug a pipeline
```

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

In the LTA archive and in some of the data (e.g. MeasurementSet) it is possible to retrieve some of the provenance information, which is stored in the history of the MeasurementSet and in the LTA database.

Jose Enrique Ruiz - 30/07/2020

1. Name of the project? (it could be a large project, a subproject, or a simple task)

a. **LST1 Collaboration**
b. **Gammapy**

2. (Data) products provided by the project?

a. The data involved in the LST1 reduction pipeline are of different nature, going from raw data (R0) coming directly from the data acquisition pipeline through the final reduced dataframes of events lists (DL3) usually in FITS format. The size of the data files are reduced along the pipeline from several GB to ~ 10^2 MB, in this process data format changes from raw protozfits, intermediate hdf5 containers, and the final compressed fits, in a chained process as the following R0->R1->DL0->DL1->DL2->DL3. The final data products that will be provided to final users are DL3 event lists, planned to be in compressed fits format. All the other intermediate data are provided to the internal LST1 Collaboration for very specific detailed and custom analysis.

b. Gammapy uses event lists and different parametrized source emission models to build spectra, images and light curves.

3. Is the capture of provenance a requirement for your project?

a. Yes
b. No

4. Do you already provide provenance information? Do you use the IVOA Provenance data model ?

a1. Yes
b1. No

a2. No
b2. No

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
_ provide more information to the final user
_ keep the traceability of the data products
_ ensure quality and reliability
_ find contact information
_ acknowledge people
_ reproduce or reprocess the data
_ debug a pipeline
_ other:
```

a.

- keep the traceability of the data products
- reproduce or reprocess the data
- debug a pipeline
- ensure quality and reliability
- provide more information to the final user

b.

- keep the traceability of the data products
- refined results from detailed analysis and filtering of the provenance produced
- provide more information to the final user

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

At this moment, in the astronomy community final users may find most of the provenance stored in headers of FITS files, while those related with pipelines processes are in plain text logs with little detailed structure and information. There is much way to improve related to storage and distribution using common models to collect much more info in a structured way, and once this is done final users will need tools to do accurate provenance analysis to refine their previously produced scientific results.

a.
Capture is done by initially logging information from the actions of the Python pipeline in text files, following a structured format based on W3C/IVOA Prov model. The internal logging machinery needs a descriptive model file, mapping actions and variables used in the pipeline to provenance entities and vocabulary, so to produce a structured-format logging file. Because of the big size of the log file produced, these are reduced in a second step to produce smaller log files per observation-run, as well as JSON files and PDF graphs. We do consider storing this reduced per observation-run provenance in JSON format in a no-SQL Mongo database, which will also help in building a provenance custom analysis tool. Also to provide SVG graphs that rendered in a browser could provide access to the different datasets present in the graph when clicking in the graph boxes. Finally, it has been considered to store some provenance information within the intermediate HDF5 data containers, as additional metadata.

b.
Capture may be done by initially logging information from the actions of the Gammapy high-level interface in text files, following a structured format based on W3C/IVOA Prov model. The internal logging machinery needs a descriptive model file, mapping actions and variables used in the Gammapy high-level interface to provenance entities and vocabulary, so to produce a structured-format logging file. Each provenance file produced could be related to specific analysis sessions producing high-level datasets (spectra, light-curves, images) from event-lists observations, capturing all parameters/values and intermediate datasets and actions involved in one analysis session. The provenance files produced are plain text files, that may be converted easily to other formats i.e. JSON, and PNG, PDF, SVG graphs, with a command from a provenance inspection API provided, which would also do simple filtering of the session provenance info according to a set of predefined pairs of keywords/values.

Andrea Bignamini andrea.bignamini@inaf.it - 02/08/2020

Molinaro, Marco marco.molinaro@inaf.it
2 août 2020 12:52
À Mathieu, François, Alessandra, Vincenzo, Andrea, Cristina, Martina

Dear Mathieu,
please find attached 3 responses to the survey from
INAF colleagues (in CC).

1. Name of the project? (it could be a large project, a subproject, or a simple task)

> **GAPS Time Series**. GAPS (Global Architecture of Planetary Systems) is a long-term program for the comprehensive characterization of the architectural properties of planetary systems as a function of the hosts' characteristics (mass, metallicity, environment). To reach these scientific goals, GAPS mainly uses the radial velocity method applied to high resolution spectra acquired with HARPS-N@TNG, and the data reduction pipelines installed at the IA2 Data Center.

2. (Data) products provided by the project?

> Time Series (TS) for host's Radial Velocity (RV) out of HARPS-N@TNG high resolution spectra are one of the main data product of GAPS.

3. Is the capture of provenance a requirement for your project?

> Yes. The determination of RV for each point in the TS depends crucially not only, obviously, on the starting spectrum, but also on the data reduction pipeline, correlation mask used, other custom input parameters such as step and width of the cross-correlation function, flux correction, recipes adopted for data reduction, and so on. A detailed provenance description can handle all these aspects.

4. Do you already provide provenance information? Do you use the IVOA Provenance data model?

> We do store the key elements (entities, activities and actors) of the TS datasets generation, even if not in an easily machine readable format, only for human consumption. We are not yet using the Provenance DM.

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
4 provide more information to the final user
2 keep the traceability of the data products
3 ensure quality and reliability
6 find contact information
7 acknowledge people
1 reproduce or reprocess the data
5 debug a pipeline
_ other:
```

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

> Considering only TS with RV obtained with HARPS-N@TNG, the capture, storage and distribution of provenance information is straightforward, since starting entities (the spectra), activities (the pipelines), and final entities (the TS) are all stored and managed within the same data center (IA2). The main effort is to adopt the Provenance DM. However, in exoplanet science using RV obtained with a single instrument sometimes can't be enough to build a scientific robust TS. Just to highlight a couple of use case. If you are studying a candidate exoplanet with a long orbital period, let say greater than one year, you may need a huge time base for your TS to cover several orbital epoch, even larger than 10 years, that cannot be afforded by a single instrument. As another example, if you are facing a weak orbital signals, you need some hundreds of RV points, both private and public, also acquired with different instruments and in different wavelengths to remove host stellar activity and other source of noise. Therefore is a crucial point having a detailed provenance description of all the RV points, with regard to both source of the data and reduction recipes adopted in order to correctly interpret the whole TS.

Alessandra Zanichelli alessandra.zanichelli@inaf.it - 02/08/2020

Molinaro, Marco marco.molinaro@inaf.it
2 août 2020 12:52
À Mathieu, François, Alessandra, Vincenzo, Andrea, Cristina, Martina

Dear Mathieu,
please find attached 3 responses to the survey from
INAF colleagues (in CC).

1. Name of the project? (it could be a large project, a subproject, or a simple task)

> Italian Radio Data Archive (Coordinator: Dr. A. Zanichelli, INAF-IRA). The Italian Radio Data Archive hosts data taken with the three Italian radio telescopes during observing projects approved by the INAF Time Allocation Committee. The Archive is a project in collaboration with the INAF Italian Center for Astronomical Archives (Responsible: Dr. C. Knapic, INAF-OATs).

2. (Data) products provided by the project?

> The Archive contains continuum and spectropolarimetric raw data from single-dish, pulsar and VLBI observations. Storage of processed data is planned once the Science Gateway and the User Space will be finalized. At that point, processing pipelines, calibration and processing information as well as some level of quality metrics will be provided. The Archive is publicly accessible, while data are subject to the INAF one-year proprietary period policy.

3. Is the capture of provenance a requirement for your project?

Yes, it is. The variety of observing projects and the heterogeneity of the data hosted in the Radio Archive is such that an accurate characterisation of the dataset is mandatory. In this way a “generic” Archive user (not the PI) will be able to address 1) if the data are suitable for her/his own research; 2) if all the necessary information for data processing is available (e.g. calibration observations). Once the Archive hosts also processed data (through the Science Gateway and User Space), it will be even more important to document all the phases of data reduction and assess the data products quality.

4. Do you already provide provenance information? Do you use the IVOA Provenance data model?

Information like the unique project ID of the scientific program and the actor performing the observation (Observer name), as well as the produced data (entity/ies) are stored in the archive’s raw data files. Also, information on weather parameters and receiver performance (e.g. wind conditions and system temperature), i.e. activityDescriptions, are saved as ancillary information as well. Therefore, while we don’t used the Provenance data model directly (but we intend to), the needed provenance information is already taken care for.

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
3 provide more information to the final user
2 keep the traceability of the data products
1 ensure quality and reliability
4 reproduce or reprocess the data
5 debug a pipeline
7 acknowledge people
6 find contact information
_ other:
```

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

Provenance information is stored in metadata keywords inside the header of raw data files, which are written in different FITS flavours. PSRFITS and UVFITS well-known formats are adopted for pulsar and interferometric correlated visibilities, respectively. Single dish data are written in a custom FITS flavour following the requirements of the FITS standard. A thorough analysis of the necessary information to be stored in the header has been conducted prior to the definition of the FITS metadata content for single dish data. Additional provenance information for a specific dataset can be retrieved also in the telescope log files and schedules, which are stored in the Archive as well. Other information to describe e.g. the UV coverage of a specific interferometric dataset can be added in graphical form but currently is still to be implemented.

Provenance information for processed data, once stored in the Archive, will be added in different ways. Additional keywords will be used to describe the data reduction process with particular focus on the calibration steps (including RFI or atmospheric opacity removal). Also, we plan to use accompanying graphical plots to describe some specific characteristics of the data processing that benefit from a more visual approach, for instance the stability in the counts-to-Jansky transformation during prolonged observing session. The Archive web interface already include provenance information among the search parameters, for instance, raw data can be searched by means of the project ID or telescope name descriptors. A similar approach will be applied to processed data, for which the distribution of provenance details will include all the ancillary information (including pipeline scripts) necessary to fully describe the reduction history of the original dataset.

Vincenzo Galluzzi vincenzo.galluzzi@inaf.it - 02/08/2020

Molinaro, Marco marco.molinaro@inaf.it
2 août 2020 12:52
À Mathieu, François, Alessandra, Vincenzo, Andrea, Cristina, Martina

Dear Mathieu,
please find attached 3 responses to the survey from
INAF colleagues (in CC).

1. Name of the project? (it could be a large project, a subproject, or a simple task)

Multi-frequency polarimetry of a complete sample of extragalactic radio sources (107 objects in total), a PhD project exploiting the observations performed with the Australia Telescope Compact Array (ATCA, proposal C2922, P.I.: Marcella Massardi, ≈ 34 h observed) and the Atacama Large Millimeter/sub-millimeter Array (ALMA, proposal 2015.1.01522.S, P.I.: Vincenzo Galluzzi, ≈ 10 h observed).

2. (Data) products provided by the project?

The project provided the following data products (in listing them, we group according to the calibration level, as defined by the IVOA ObsCore DM):

- Level 0:
  a. raw visibilities (ATCA observations available through the Australia Telescope Online
  Archive, ATOA, in the MIRIAD RPFITS format; ALMA observations available through the
  ALMA Science Archive, ASA, in the ASDM format);
- Level 2:
  b. calibrated visibilities (available on request for ATCA observations in the MIRIAD UVFITS
  format, while for ALMA observations they can be directly downloaded from ASA in the MS
  format);
- Level 3:
  c. Stokes' I, Q and U maps for each source in the sample (FITS files available on request;
  for ALMA observations, thumbnails are published as supplementary material of a
  dedicated paper, while IQU CASA image cubes can be downloaded from ASA);
- Level 4:
  d. various catalogues reporting full-Stokes' flux densities and parameters resulting from SED
  fitting (published as supplementary material of papers or available through VizieR);
  e. various plots and tables reporting statistical estimators (e.g. mean and quartiles) of
  spectral indexes, polarization fractions and rotation measures (the code used for
  generating them has been mainly written in IDL and is available on request);
  f. normalized source counts in total intensity and polarization at 20 and 100 GHz (tables are
  in the papers; the code is publicly available on gitlab);
  g. contamination forecasts due to extragalactic radio sources on CMB angular power spectra
  at 20 and 100 GHz (plots are reported in dedicated papers; the code is publicly available
  on gitlab).

- Additionally, ancillary products [1] are:
  h. ATCA observation schedules (available on request or directly accessible to all registered
  ATCA observers).
  i. calibration scripts (exemplar MIRIAD scripts for ATCA data are also published in the
  Galluzzi's PhD thesis) and imaging scripts;
  j. diagnostics plots produced during the execution of scripts (image files in png or jpeg);
  k. observations and processing log files (human-readable .txt files).

[1] Similarly to what presented for data products, ancillary products related to ATCA observations are available on request, while for ALMA observations such products can be directly downloaded from ASA.

3. Is the capture of provenance a requirement for your project?

Yes, it is. The main reason is represented by the fact that radio polarimetry involves more calibration steps and presents more sources of systematic issues with respect to radio observations encompassing only the total intensity. Thus, additional diagnostics is usually addressed to provide indicators about data quality (e.g. statistics of leakage terms solutions and of Stokes' parameters flux densities). Moreover, the variety of data products ranging from observations to more advanced ones (e.g., catalogues and source counts) particularly require traceability information.

4. Do you already provide provenance information? Do you use the IVOA Provenance data model?

We have not used the IVOA Provenance model so far, and the main source of provenance information we provide (e.g. the proposal id, a brief description of the calibration recipe, some parameters for data quality, dedicated sections to each type of advanced data products generated from observations, contact information and acknowledgements) is represented by published papers and attached materials. In a sense, we do have the information needed to fill in entities, activities, actors and their parameters and descriptions, but these information are currently understandable by humans only and are not stored in a specific or easily retrievable format.

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
3 provide more information to the final user
2 keep the traceability of the data products
1 ensure quality and reliability
6 find contact information
7 acknowledge people
4 reproduce or reprocess the data
5 debug a pipeline
_ other:
```

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

The main source of provenance information for our project are the published papers, as already mentioned at the point 4 of this questionnaire. However, some amount of information to reproduce our results and some quality metrics indicators can be typically extracted from headers of data products or can be found in ancillary products, such as human-readable .txt files or images in standard formats (cf. the list of products given at the point 2): we highlight that in some cases, e.g. for MIRIAD calibrated visibilities or images, any executed tasks leave its trace (through input parameters and outcome of procedures) as HISTORY keywords, hence the full workflow could be, in principle, reconstructed. Again, in terms of data calibration activities, a collection of calibration scripts are already published (in the ASA or in a publicly accessible PhD thesis); in terms of enhanced or analysis data products generation, in particular source counts and forecasts for CMB observations, we have already publicly delivered documented code on gitlab. The rest of code for imaging (e.g. scripts in bash for MIRIAD), machine readable catalogues generation, SED fitting, statistical analysis of particular quantities or plotting (mainly IDL programs) is typically available on request, and has been already shared to interested users through limited access folders.

Veronique Delouille V.Delouille@oma.be - 07/08/2020

First project: Automatic detection of AR and CH (SPoCA)

1. Name of the project? (it could be a large project, a subproject, or a simple task)

SPoCA

2. (Data) products provided by the project?

Catalogs of Active regions and Coronal holes

3. Is the capture of provenance a requirement for your project?

No

4. Do you already provide provenance information? Do you use the IVOA Provenance data model ?

Yes, via parameters name

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
_ provide more information to the final user
X keep the traceability of the data products
X ensure quality and reliability
_ find contact information
_ acknowledge people
X reproduce or reprocess the data
X debug a pipeline
_ other:
```

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

> Via parameters in a VOEvent file.

Second project: ROB Sunspot Database, done from Solar drawings
Questions:

1. Name of the project? (it could be a large project, a subproject, or a simple task)

> **USET**

2. (Data) products provided by the project?

> Database of solar drawings and of sunspots

3. Is the capture of provenance a requirement for your project?

> For new project using this database, yes.

4. Do you already provide provenance information? Do you use the IVOA Provenance data model ?

> Provenance information such as who made the drawing, who verified it, are given as column variable in the DB
> Do you use the IVOA Provenance data model ? No

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
X provide more information to the final user
X keep the traceability of the data products
X ensure quality and reliability
X find contact information
_ acknowledge people
_ reproduce or reprocess the data
_ debug a pipeline
_ other:
```

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

> There is no clear structure about provenance information that is being stored/distributed.

Gilles Landais <> - /08/2020

Questions:

1. Name of the project? (it could be a large project, a subproject, or a simple task)

> **VizieR** (CDS) : catalogue service published in astronomy.

2. (Data) products provided by the project?

> mainly tables (~40K tables from ~20K articles) , but also spectra, images the data comes from article published in the major astronomical journals (AAS, A&A, MNRAS) and science-ready data from space agencies (Nasa, ESA, ESO, ...)

3. Is the capture of provenance a requirement for your project?

> VizieR is a trusted respository certified by the CTS : workflows are specified in CTS, and the provenance is one of the topic who evloves.
> Provenance (as W3C) is not today required - today , it is a perspective ...

4. Do you already provide provenance information? Do you use the IVOA Provenance data model ?

> in test today

5. Which goals should the addition of provenance information serve (if possible by order of priority)?

```
X provide more information to the final user
_ keep the traceability of the data products
X ensure quality and reliability
X find contact information
X acknowledge people
_ reproduce or reprocess the data
_ debug a pipeline
_ other:
```

> in order of priority : (vor the VizieR cotnext)

```
1) provide more information to the final user
2) ensure quality and reliability
3) find contact information
==> DOI/ORCID do already the job
4) acknowledge people
==> are DOI suficient ?
5) keep the traceability of the data products
```

6. How do you consider or already do the capture, storage and/or distribution of provenance information?

We limit the provenance to a simple view - the information comes from metadata availablme in the database
Provenance is more a standard way to provide the information .