

Minutes - M. Allen, March 27, 2007

OBJECTIVES OF THE MEETING

- First face-to-face meeting of Internal Science Team
- Discuss IST input for DCA June workshop
- Discuss IST input on Data Centre Census

LOCATION

ESAC, Spain - following the 'Astronomical Spectroscopy and the Virtual Observatory' workshop.

PARTICIPANTS

Mark ALLEN (MA)
Matteo GUAINAZZI (MG)
Jonathan TEDDS (JT)
Piero ROSATI (PR) [via Video Skype]

Appologies: Eduardo MARTIN, Gijs VERDOES, Santi CASSISI, Niv DRORY

MINUTES

1. MA - Welcome everyone to first face-to-face meeting of the DCA-IST. All agreed associating IST meetings with scientific/VO meetings is a good idea, although the IST members who are less familiar with VO activities are unfortunately not present. Effort will be made to get those members to next face-to-face meeting.

2. MA - Reports that the idea of including science examples, or overall scientific motivation in the workshop tutorials has been discussed with C.Arviset and MG.

MG - explained that workshop participants are expected to bring own examples and science cases to work on during workshop, so perhaps not necessary to develop new science cases for tutorials

All - Over-all scientific motivation for VO uptake and use of VO tools could be provided by IST at the opening of the workshop, or perhaps in each tutorial

JT - Is a member of the workshop organising committee, and will discuss IST input in the WOC up-coming WOC meeting

3. Data Centre Census

Action BOARD2-WP2-A3 asks the IST to comment on the process for gathering and presenting the information in the data centre census.

MA - Brief overview of the census goals. We understand that the census has two main roles in the DCA project. Firstly as a major deliverable to EC in a form that clearly sets out the number, size, distribution and diversity etc. of data centres in Europe. We imagine this is to be used for future detailed assessment of resourcing VO up-take in Europe

Secondly, and more immediately relevant to IST, is that the census be used within the DCA project to focus the topics for the DCA workshops, tutorials and data centre visits. In commenting on the process for gathering and presenting the census information we have concentrated on this aspect of the census. In particular we consider what information should be considered in the census in order for the DCA project to best help the data centres in take-up of VO standards, and to help DCA prioritize which data centres to help.

All - Review of initial census constructed from partner input.

- Initial census is simplistic, and has not been presented in a uniform way. For example LEDAS (Leicester Database and Archive Service) which has some 3000+ catalogues plus several X-ray data archive, is presented as a single row entry, whereas each ESO instrument is presented on a separate row. Clearly there is some hierarchy in the information which could better be presented by grouping rows, or by using some kind of tree structure.
- Overall the information in the census could be more clearly presented with the aid of graphics like graphs and pi-charts etc.

We agreed that the 'raw' data of the census is still best kept as a table or in a simple database. Starting with the initial census columns, we discussed the information that should be collected in the census. A short list of the items discussed is included below, and the IST undertakes to put this into a more coherent form by April 10.

Action: IST-M1-1 MA - circulate draft set of recommendations from IST for census

Action: IST-M1-2 MA - compile comments from IST and provide to DCA board by April 10 as the formal response to BOARD2-WP2-A3

4. List of items discussed for census, expanding on the set of columns used in the initial census.

- 'Data Archive / Science ready data'

Separating information on the data/products from the information on the archive. For the data/products the following information should be considered:

Volume of data:

- total volume of data centre (GB)
- catalogues: number of objects/rows/columns

Type of data:

- energy regime (Radio, IR, Optical, UV, X-ray, Gamma-ray)
- images / spectra / multi-d data/ catalogues / other
- Level of products as: 1-raw, 2-previews 3-science ready
- Fraction of products in each level

For Archives the following information should be considered

- Number of users
- Human resources supporting the archive
- Contact point and Scientific contact point
- History and expected lifetime of the archive
- Uniqueness - is the same data hosted elsewhere?
- Original data - mirror services
- Public/Proprietary

Level of VO awareness

- current compliance with VO standards
- Intent to 'put data on the VO'
- Is VO compliance required by your governing bodies?