Progress Report on the AVO Project

Markus Dolensky (ESO) on behalf of the AVO Project Team

When the Astrophysical Virtual Observatory (AVO) was first mentioned in the July 2001 issue of this Newsletter it was merely a paper concept based on experience with ASTROVIRTEL and other archive activities. Since then a geographically dispersed project team in France, Germany and the UK has started its work, guided by the advice of a dedicated Science Working Group (SWG). Here we briefly review the progress of the project.

Organisational Structure

The AVO project is led by the Executive Committee comprised of representatives of the six participating organisations. An advisory body, the AVO Science Working Group, is responsible for defining science goals and requirements and for reviewing progress.

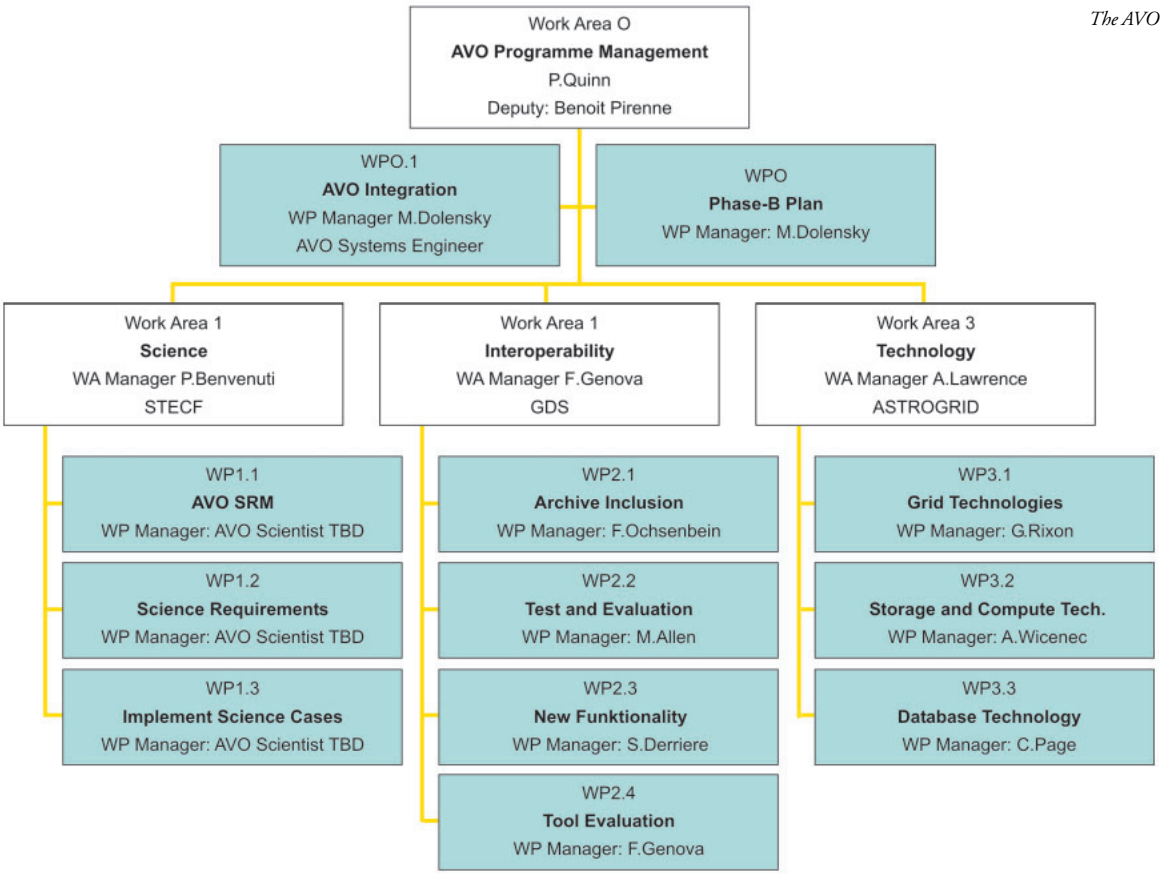
Three main working areas have been identified:
- Science
- Archive Interoperability
- GRID and Data Storage Technology

Work Area 1 – Science

The SWG nominated a subcommittee chaired by Nicholas Walton (IoA, Cambridge, UK) to select "use cases" for a first AVO software demonstrator planned for January 2003. It was decided to give priority to work on tools for finding targets for follow up spectroscopy based on the Great Observatories Origins Deep Survey (GOODS). GOODS (ESO Messenger 105, p. 40, www.eso.org/goods/) is a public, multi-wavelength survey covering two 150 arcmin^2 fields centred on the Hubble Deep Field North (HDF-N) and the Chandra Deep Field South (CDF-S).

Some of the challenges are the cross-matching of images taken in various spectral bands, ranging from the radio to the X-ray regime, the cross-identification of objects in those bands, the integration of observation logs from several space and ground-based observatories into a common scheme and the homogenisation of heterogeneous catalogues in order to visualise the Spectral Energy Distribution (SED) as well as colour-colour plots. A further challenge is the re-extraction of sources from the original images, driven by user constraints. It is clear that not all of these goals can be met fully by January 2003, but work in all these areas has started. These activities are based on initiatives in the other work areas as described below.

The AVO Organigram



Meta-Data
Meta-data can be considered as triples consisting of a name/value pair as well as a subject. This third piece is crucial, since the subject puts meta-data into a context. The context can be expressed in several ways, but is necessary to ensure that the triple is unique. One possibility is to use the URL of the XML document to satisfy this constraint. The triple { Author, Max Planck, www.abc.org/apaper.xml } unambiguously declares Max Planck as the author of the given document.

DTD vs. Schema
Document Type Definitions (DTD) and XML Schema both describe the content and structure of a class of XML documents. DTDs came first and are expressed in Extended Backus-Naur Form whereas Schema are directly written in XML. In the opinion of the author DTDs are a subset of Schema in terms of functionality and that Schema will eventually replace DTDs, at least in the XML world. Some of the additional features XML Schema offer are: Namespaces (modularity), inheritance, numerous data types, so-called facets (regular expression patterns for validation, enumeration), default values, nil values, etc. Usage scenarios for Schema suggested by the W3C Consortium are: query formulation, open and uniform data transfer between applications and databases as well as meta-data exchange.

Web vs. Semantic Web
There is a fundamental difference between the Web and the Semantic Web: when we browse the Web we experience it as a system where the human reads Web pages. The Semantic Web, however, is a system where the machine can not only read, but more importantly understand the Web pages. In historical terms, the Semantic Web is currently moving from the Stone Age to the Bronze Age.

Work Area 2 – Interoperability

A need for common standards when exchanging data across archives to build higher level services was identified early on. When ESO hosted the very successful conference "Toward an International Virtual Observatory" in June 2002 a coordinating body for such a standardisation process was formed. It is called the International Virtual Observatory Alliance (IVOA). All VO initiatives known at the time in Australia, Canada, Europe, Germany, India, Japan, Russia, UK and USA joined this alliance. The members of these national and international initiatives agreed on a mission statement and plan until 2006 (www.ivoa.net).

Two important accomplishments are worth announcing in this context: VOTable and UCDs.

VOTable

The OPTICON working group released the VOTable V1.0 specification in April 2002 (cdsweb.u-strasbg.fr/doc/VOTable/). VOTable can represent astronomical tables using the eXtensible Markup Language (XML). VOTable is not meant to be a replacement for the FITS format. It is part of a transport layer to exchange data and meta-data (see box) in a hierarchical structure. There exists a DTD as well as a Schema definition (see box) of this format and prototypes of several software libraries supporting VOTable are available. To date there are implementations in Java, perl and C.

UCD Concept

A scheme of Unified Content Descriptors (UCD, vizier.u-strasbg.fr/doc/UCD.htx) has been developed as part of an ESO-CDS data mining project. The original goal was to map database columns with a distinct set of well described parameter identifiers. Such identifiers (UCDs) were grouped in a shallow hierarchy. 1,500 UCDs were sufficient to classify some 100,000 columns from 3,000 astronomical catalogues. These UCDs cover a lot of ground in terms of defining meta-data for resource registries and some common query specifications. UCDs are widely accepted in the IVOA community as fundamental building blocks for a vocabulary in astronomy suited to automated processing. The discussion now focuses on support for unit conversion and whether the UCD tree should be refactored to improve flexibility and simplify the handling of missing items.

Work Area 3 – GRID & Infrastructure

AVO formed a strategic alliance with the British Astrogrid project. The Astrogrid project started before AVO and so it was natural to share resources in this area in order to avoid parallelism resulting in incompatible approaches. Hence AVO and Astrogrid are collaborating in the fields of database, data mining, storage and Grid technologies.

The Near Future

The first science demonstrator will be released at a workshop hosted by the Jodrell Bank Observatory on January 20/21, 2003. It will focus on the GOODS case. Software components and data from GOODS will become publicly available at the same time. Details will be announced on the new AVO project page www.euro-vo.org. A second release is planned for the IAU general assembly in August 2003. This will be complemented by a detailed recommendation of IVOA due in January 2003 concerning the items that need standardisation. It is expected that the specification of an astronomical query language based on the evolving Semantic Web technology (see box) will rank high on this list.