A Framework for a Virtual Repository of Outreach Products

Lars Lindberg Christensen
AVO Outreach and Education Coordinator

Hubble European Space Agency Information Centre

v.1.1 Revised after Jan. 2005 AAS meeting

Abstract
The coordinated exploitation of astronomical science archive data through the Virtual Observatory will have a major effect on the way astronomers work. The exploding volume of incoming data and the emergence of technologies and tools to mine the archives will inevitably also have a knockdown effect and result in significant changes for outreach and education as well. There is undoubtedly a great potential for exploiting 'VO-data' (meaning data in the VO era) and facilities in the fields of education and outreach, but there is equally no doubt that this task is difficult and will need a coordinated worldwide effort.

Here we discuss some recent developments in this direction and outline a proposal for a framework for a virtual repository of outreach products (pretty pictures, videos and other digital material). This repository should allow the future exploitation of all kinds of outreach material in 'digital universes' by systematically linking resource archives worldwide.

Virtual Observatories
For scientists the Virtual Observatory (VO) will in the future link the archival data holdings of space and ground-based astronomical observatories, multi-wavelength catalogues and related computational resources. The VO will provide new opportunities for scientific discovery by enabling a new mode of research through the application of new database and data-handling tools that will become available during this decade.

Together with advanced instrumentation techniques, a huge new array of astronomical data sets will soon be available at all wavelengths from the radio to the X-ray and gamma-ray regions. These very large datasets will have to be archived and made accessible to scientists in a systematic and uniform manner to realise the full potential of the new observing facilities. It is difficult to illustrate the upcoming 'data explosion', but even now, if a scientist wants to download and analyse just one third of the Sloan Digital Sky Survey images it would take him about a year with a very good 200 KB/second network connection.

The Virtual Observatory initiative is currently building a global collaboration of the astronomical communities in Europe, North and South America, Asia, and Australia under the auspices of the International Virtual Observatory Alliance (IVOA).

Figure 1: The International Virtual Observatory Alliance partners.

Astrophysical Virtual Observatory

A major European component of the Virtual Observatory is the Astrophysical Virtual Observatory (http://www.euro-vo.org) that started in November 2001 as a three-year Phase A project, funded by the European Commission (FP5) and six organizations (ESO, ESA, AstroGrid, CNRS (CDS, TERAPIX), University Louis Pasteur and the Jodrell Bank Observatory) with a total of 5 M€. A Science Working Group was established in 2002 to provide scientific advice to the AVO Project and to promote the implementation of selected science cases through demonstrations. In 2004 the first science result – a systematic search of the archives, culminating in the finding of numerous obscured Type 2 Active Galactic Nuclei - was released as a consequence of these efforts.

There is no doubt that the Virtual Observatory will become a powerful tool, not only for scientific research, but potentially also as a means to promote a better understanding of the Universe in society and to provide for the educational needs of current and future generations. There is however a huge leap from the ‘dirty’ data in the archive to the ‘clean’ and refined data needed in an educational situation.

An educational VO

For a real “educational VO” the primary aim may be defined as: “Excite, inform and educate the public about space science and astronomy through access to real data, and serve as a catalyst for scientific and technological literacy.”

As we will see this is very difficult to achieve in practice. However, there are many reasons for trying to use real data, especially in education. Firstly the data are free. In addition real data and real science give students a sense of adventure and discovery. In some cases there is a real feeling of breaking new ground and the chance to make genuine discoveries. Finally astronomy projects that draw on real data can be a catalyst for learning about information technology.

Some small, but significant steps in the direction of opening the data archives to laymen, educators, and students have been achieved with our “ESA/ESO/NASA Photoshop FITS Liberator” (http://www.spacetelescope.org/projects/fits_liberator) and “FITS for Education” projects. For many years astronomical images from the world’s telescopes were only easily manipulated and displayed by astronomers and technical staff with access to collections of powerful, but often hard to use, software tools. With this free plug-in anyone can work with images and spectra from the NASA/ESA Hubble Space Telescope, the European Southern Observatory’s Very Large Telescope, the European Space Agency’s XMM-Newton X-ray observatory, NASA’s Spitzer Space Telescope and many other major facilities.

But for a real “educational VO” the goal might even be the creation of advanced “Digital Universes” that taps into science archives around the world and gives access to all data at the click of a mouse. A beguiling vision: but is it realistic?

The obstacles
In many ways this vision is not realistic, at least not within the foreseeable future. Education implies clarity and simplicity in presenting and using new ideas and principles, and so demands products at the highest level of ‘refinement’. Raw data are inherently ‘dirty’ and full of the complications that make the extraction of real science a challenge on all levels even to experienced scientists. These complications can often obscure the educational point. The VO interfaces that are currently under development are for experts only, requiring extensive training even for astronomers.

With time, data issues will become more and more complicated. Today there are ever more data, more wavelength regions, more telescopes, more detectors and more calibrations to follow. On top of this, VO concepts are very abstract and difficult to explain in an educational situation. Teachers themselves have often not had the right training in dealing with data or using advanced astronomical image processing tools such as IRAF and MIDAS, and may not have the right connections with scientists who could help.

A framework for a virtual outreach and education repository
If a real educational VO is not a realistic proposition for the next five or ten years, then we must turn instead to some smaller and more realistic goals. It is in a way paradoxical to discuss access to real data when there already today are vast quantities of ‘clean’ outreach and educational material available on the web. The problem is that they are not linked systematically, and it is therefore next to impossible for educators, laymen and the like to search these resources in a simple manner.

Here we describe a realistic first step towards an “educational VO”. The framework outlined here will link outreach products and resources worldwide, ultimately allowing the creation of various ‘digital universes’ by professional companies or other organisations.

Today, most public outreach resources, most notably images, do not adhere to any specific standards for archiving. Imagine if the wonderful collections of press release materials from ground and space-based missions could include common information (known as metadata), such as their positions on the sky, object names etc. An elaborate and standardised system could be envisaged whereby the world’s archives of more refined outreach and education products such as ‘pretty pictures’ and videos could be tied together and made accessible.

This would make it possible for outreach offices as well as third-party companies to build automated tools that could interface with image databases on the Internet and allow exploration of this treasure. Anything from simple searches using existing, and very powerful, internet search engines up to interfaces to fully three-dimensional 'digital universe' settings is conceivable as an outcome of such a framework. One could imagine using outreach images in live planetarium shows, in comparative multi-wavelength views, as a teaching aid and in many other places. Finally, and most important of all - so long as future PR images are compatible with some yet to be agreed upon standard, the treasury of 'mouse-click accessible' images will grow from day to day.

This is a challenging, but manageable, task. It demands consensus and collaboration among the entire outreach and education community - from the people creating the 'pretty pictures' (image processing specialists), via the data Holdings (the outreach archives) to the different end-users such as educators and 'visualisers', who use the resources to visualise 'digital universes'.

The aim of such a system could be stated: "To allow outreach resources to be 'catalogued' in a virtual repository and accessed by educators, press, students and public through specialized visual tools combined with search engines."

An imaginary example of the outcome of such a system is shown in Figure 2.

Figure 2: The vision of a Digital Universe. Wearing Virtual Reality helmet and gloves we fly through a database of star positions and images of objects placed in the correct positions. Occasionally we stop to query for information about interesting objects we find along the journey. Credit: NOVA/NCSA.

All parties will gain by such a system. By working together and defining a common set of systems and formats more people will see and use the resources from each group. This should be enough in itself to get the ball rolling.

For the sake of simplicity such a system might be termed a: "Virtual Repository". Here repository is used in the meaning of a 'place' where the outreach and education resources are 'collected', and 'virtual' in the sense that no physical movement of data should take place. Instead a framework is put in place whereby the data can be accessed seamlessly in a similar style to that envisaged for the Virtual Observatory.

The components of the "Virtual Repository"

The first component of the "Virtual Repository" is a *central coordinating organisation* to endorse the necessary formats for metadata and protocol. This rôle would sit most effectively with the "Virtual Repository Programme Group" within the International Astronomical Union Division XII Working Group "Communicating Astronomy". The programme group was created following the Global Hands-On Universe meeting in St. Petersburg in 2004: http://www.communicatingastronomy.org/repository/.

The second component should be a well-defined list of *metadata* descriptors that would always accompany products such as images and videos. A draft for such a list is given below. It is currently under discussion and, once agreed, will be submitted for final endorsement by the IAU.

Metadata keywords:

- **FILE:**
  1. Product type ("image") [image/video/text]
  2. File format ("tiff") [tiff/jpeg/avi/mpeg-2/]
  3. Original dimensions (=NAXIS1/2) ("2100 x 2304 pixels")
- **ID:**
  4. Identifiers (e.g. "heic0412a, opo0420b")
  5. Observatory ("1: Hubble Space Telescope")
  6. Instrument ("WFPC2")
  7. Dataset names (VO compliant if possible): ("ivo://ESO.HST/U2JZ0607B, ivo://ESO.HST/U2JZ0603B, ivo://ESO.HST/ U2JZ0605B, ivo://ESO.HST/U2JZ0601B") [ivo://AuthorityID/ResourceKey]
  8. Data holding: ESO.HST
  9. Creator ("41: Hubble European Space Agency Information Centre")
  10. Publisher ("Lars Lindberg Christensen")
- **PROCESSING:**
  11. White level (z1)
  12. Black level (z2)
  13. Stretch function
  14. Scale factor
  15. Offset
- **INFO:**
  16. Quality ("2")
  17. Further information link ("http://hubblesite.org/newscenter/newsdesk/archive/releases/1995/45/image/a")
  18. Comment ("This spectacular color panorama of the center the Orion nebula is one of the largest pictures ever assembled from individual images taken with the Hubble Space Telescope. The picture, seamlessly composited from a mosaic of 15 separate fields, covers an area of sky about five percent the area covered by the full Moon.")
  19. Credit ("ESA & NASA")
  20. Image release date ("02.01.1995")
- **ASTRO:**
  21. Wavelength range ("502-658 nm")

22. Corner coordinates (ra, dec, Epoch 2000) “(04 12 12, -05 04 32) (04 12 04, -05 04 32) (04 12 10, -05 07 32) (04 12 04, -05 07 32)”
23. Creation type (“real”) [real/simulated/artwork]
24. Target name (“M 42”)
25. Other Number of exposures (“4”)
26. Exposure times in seconds (“320, 300, 700, 900”)
27. Object class/subclass (“nebula”)(“emission”):
    - Solar System
        - Venus
        - Mars
        - Jupiter
        - Saturn
        - Uranus
        - Neptune
        - Pluto
        - Planetary Moon
        - Planetary Ring
        - Weather / Atmosphere
        - Minor Body
        - Asteroid
        - Comet
        - Kuiper Belt Object
    - Star
        - Binary Star
        - Brown Dwarf
        - Constellation
        - Massive Star
        - Neutron Star
        - Nova
        - Protostellar Jet
        - Protoplanetary Disk
        - Pulsar
        - Star Field
        - Star with Planet
        - Supernova
        - Variable Star
        - White Dwarf
    - Star Cluster
        - Globular
        - Open
    - Nebula
        - Dark
        - Emission
        - Planetary
        - Reflection
        - Supernova Remnant
    - Galaxy
        - Cluster
        - Dwarf
        - Elliptical
        - Interacting
        - Irregular

    - Magellanic Cloud
    - Quasar / Active Nucleus
    - Spiral
- Exotic
    - Black Hole
    - Dark Matter
    - Gamma Ray Burst
    - Gravitational Lens
- Cosmology
    - Distant Galaxies
    - Intergalactic Gas
    - Universe: Age / Size
- Facilities
- Miscellaneous

The third component some sort of *dynamic data Holdings list*. This list contains the 'addresses' of data Holdings ('pretty picture' archives etc.). Some natural things to have in this list would be the data Holding query format and metadata conversion rules as well as the Holding type (EPO group, school, robotic telescope). Any interested group with data to share, should in principle be allowed to sign up to the registry so that resources can be disseminated as widely as possible.

Perhaps this could be done as Registries as is done by the VOs: "Registries function as the "yellow pages" of the Virtual Observatory, collecting metadata about data resources and information services into a queryable database."

The fourth component would be the *communication* between the user and the data Holdings. This could perhaps be done through the Registry with the help of a VO-style Data Access Layer, such as the Simple Image Access Protocol (SIAP).

With the development of the upcoming FITS Liberator version 2.0 a step in the direction of the second item on this list by adding agreed-upon metadata (presumably in XML-format) to the jpeg and tiff 'pretty pictures' that are stored in outreach archives around the globe.

Summary

The basic framework outlined above needs a collaborative endorsement of the fundamental elements at each level in the list. For the first item this is a *central coordinating organisation*, for the second a fixed list of *metadata* descriptors for outreach and education resources such as images and videos. For item three a *dynamic data Holdings list* containing 'addresses' of the archives of images and videos is required and finally a protocol for *communication* between user and data Holdings is needed.

If this basic framework were successfully put in place it would open the door to the 'Digital Universe'. Search engines could be built that would enable laymen and educators to search globally for pictures of individual galaxies and stars. Visualisers such as the Redshift and Starry Night planetarium software or full-immersion 'real' planetarium dome systems such as Evans & Sutherland's Digistar 3 could place the outreach images in the right context on the sky and link to textual information. And this would just be the beginning...

Literature

- IVOA documents, general: http://www.ivoa.net/Documents/latest/
- VO Overview: http://www.ivoa.net/Documents/latest/IVOArch.html
- VO overview 2: http://www.ivoa.net/Documents/Notes/CODATA/IAUCodataReport2004-20040910.html
- Resource Metadata for the Virtual Observatory: http://www.ivoa.net/Documents/latest/RM.html
- Uniform Content Descriptors valid words (metadata tag values): http://cdsweb.u-strasbg.fr/UCD/ucd1p-words.txt
- Metadata list (towards the bottom) for NASA's Space Science Education Resource Registry: http://ossdev.stsci.edu/Registry/registry.cgi?stage=faq
- Gateway to Education Materials schema: http://thegateway.org/about/documentation/ApplicationProfiles-9-8-04
- Dublin cores overview: http://dublincore.org/documents/dcmes-xml/index.shtml