PROGRESS OF THE VIRTUAL REPOSITORY

Lars Lindberg Christensen

ESA/Hubble, Munich, Germany

ABSTRACT

When talking to other 'communicating astronomy' colleagues one issue always comes up: how can we share our resources more effectively? And ultimately, how do we allow the public better access to images, videos, news and other material? Imagine a kind of "Google Universe". In 2004 a Programme Group called Virtual Repository under the IAU Working Group Communicating Astronomy was set up to stimulate work in this area.

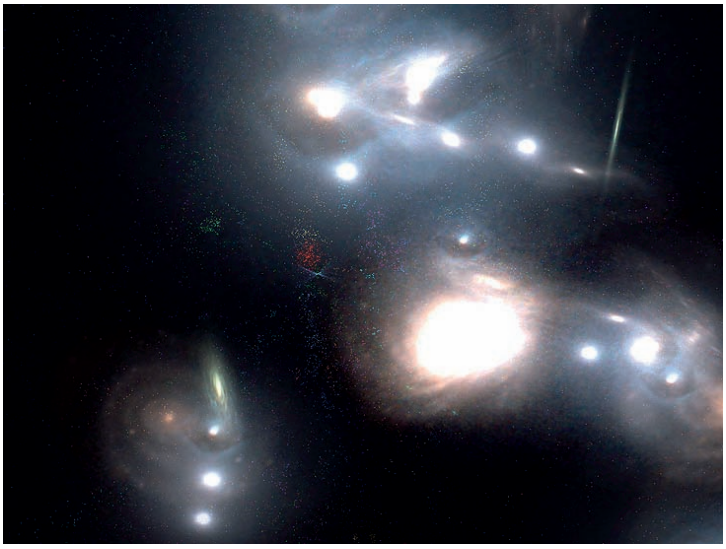

***Figure 1**. The dream of the Digital Universe is slowly becoming reality. Credit: Image from the Uniview application, property of Sciss AB Sweden, and the Digital Universe database, property of the American Museum of Natural History*

PROBLEM A – DATA IN EDUCATION

The coordinated exploitation of astronomical science archive data through the Virtual Observatory (VO) 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 knock-down 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 educational context, but there is equally no doubt that this task is difficult and will need a coordinated worldwide effort.

An ultimate goal for us, as science educators, is to give the students access to real data in a "Digital Universe". There are many reasons for trying to use real data in the education process. Firstly the data are free. Secondly contact with real data and real science gives students a sense of adventure and discovery. In some cases there is the exhilarating 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 educators, students and other interested parties have been achieved with our "ESA/ESO/NASA Photoshop FITS Liberator" (see http://www.spacetelescope.org/projects/fits_liberator and elsewhere in this volume) and "FITS for Education" projects. Using 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.

However for a real "educational VO" the goal might even be the creation of advanced "Digital Universes" that can tap into science archives around the world and give access to all data at the click of a mouse. A beguiling vision, but probably not realistic in the near future? Giving access to real data is a major undertaking and quite possibly not worth the enormous effort involved. Scientific data are complex and inherently "dirty", and teachers may lack the background and training to understand many of the issues.

PROBLEM B – IMPROVING ACCESS TO EPO IMAGES, VIDEOS AND NEWS

At the same time as the scientists are experiencing a "data flood", we in Education and Outreach (EPO) are also experiencing, and participating in, a similar parallel development. Larger and larger amounts of "polished" audiovisual multimedia materials are being made available to press, educators and lay people on the web. The volume of digital products—outreach images, videos and news—is increasing all the time and the trend seems to be accelerating. Vast quantities of 'clean' outreach material are available on the web today. The problem is that they are not linked systematically, and it is therefore next to impossible for the press and public to search these resources in a simple manner. Or, as I expressed myself bluntly during the conference: *"Heck, if **we** can't even find each other's stuff, how can we expect other's to?"*.

Ultimately, how do we allow the public better, and more coordinated, access to images, videos and other materials?

HOW DID PROBLEM B ARISE?

The Internet and the World Wide Web have been determining technologies allowing efficient access to information and provide a fruitful environment for the creation of new information. Although very innovative search tools, such as Google, have been developed to search the textual parts of the available information on the web, similarly efficient tools for searching audiovisual content such as images and videos do not exist.

Today's search engines work by searching and indexing the textual information in html text-pages on the web. Existing audiovisual search tools, such as Google Images, can only search textual information that is placed around embedded images on a web page. This information consists largely of random pieces of text that often have little to do with the actual images and furthermore only images embedded in html pages can be searched. All audiovisual content in image or video archives, or databases, cannot be searched in this way and thus excluding a large majority of existing audiovisual content by far. In addition, real archives are the preferred storage method for the highest quality content, i.e. the content closest to the scientific and cultural sources.

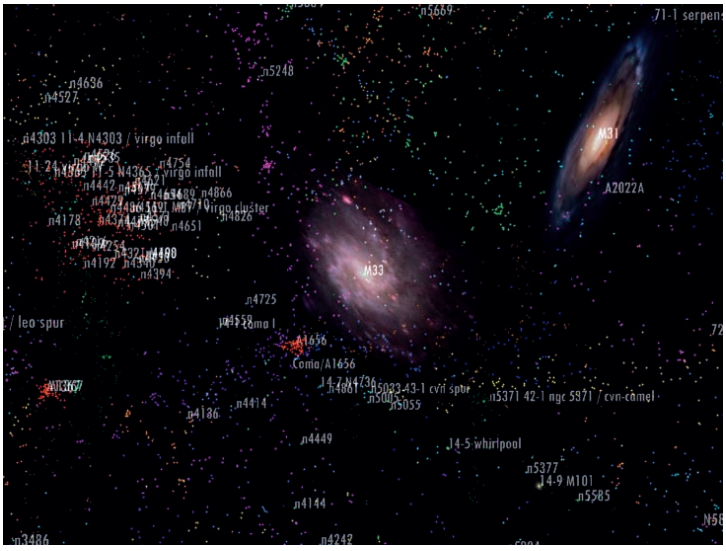

***Figure 2**. Placing EPO images in the right context within Digital Universe models in different environments such as the planetarium domes or on the desktop, may well be one the hottest issues in EPO over the next decade. Image from the Uniview application, property of Sciss AB Sweden, and the Digital Universe database, property of the American Museum of Natural History*

THE VIRTUAL REPOSITORY

What is needed is a framework that enables seamless searching in archival databases on the web—the Virtual Repository (VR). Here repository is used in the meaning of a 'place' where outreach and education resources are 'collected' and 'virtual' in the sense that no physical movement of data should take place—only a framework whereby the data can be accessed seamlessly in a sort of 'Virtual Observatory-style' is required.

The search should be advanced, and allow users to specify search criteria such as quality, size, popularity and more. The Virtual Repository project is dedicated to improve the accessibility and usability of digital astronomical material in a multilingual environment. The VR will coordinate collections in astronomical audiovisual archives worldwide and enhance the quality of the audiovisual material using well-defined metadata. The project will reinforce cooperation between digital content stakeholders such as the existing image archives at the large observatories. The aim is to give access to the unique resource that is the sky—a vast laboratory of science that is always in operation and accessible at all times to everybody.

A few possible applications of such a VR are:

1. Search engines (such as a "Google Universe")
2. Interactive click-and-point experiences in the planetarium dome ("let's look at the Orion Nebula in different wavelengths")
3. The sky on your home desktop: Links with existing commercial planetarium software (Redshift, Starry Night, The Sky etc.)
4. AstroKiosk (exhibition kiosks that automatically tap into, and exploit the daily stream of astro-news, provided they are coded with the right VR-metadata).
5. Educational material

There are no limits to the potential applications when a VR framework is in place, interlinking multi-wavelength images and videos and placing them in the right context.

On the technical, or implementation, side, the Virtual Repository is in essence a framework consisting of four components, namely:

1. Resource metadata tags attached to images, videos, news etc. (see Christensen, 2006, this volume)
2. A centralized organiser or controller (for instance the IAU Working Group Communicating Astronomy with the Public).
3. A list containing the data archives, i.e. a "telephone book" or a registry that contains metadata about data resources and information services
4. A definition of a protocol for communication between the physical repositories and the users. For instance through the Registry with the help of a VO-style Data Access Layer, such as the Simple Image Access Protocol (SIAP)

THE IAU VIRTUAL REPOSITORY PROGRAMME GROUP

The International Astronomical Union (IAU) Virtual Repository Programme Group was set up during 2004 and business meetings were held at the AAS meeting in January 2005 and at the CAP 2005 meeting in June 2005. The purpose of the group is: *"To construct the framework for a virtual repository to allow outreach resources across projects and country borders to be 'catalogued' in a virtual repository and accessed by educators, press, students and public through specialised visual tools combined with search engines."*

So far progress has been made in three different areas:

1. Concept: The concept of the Virtual Repository has been thought out, discussed and improved to such a degree that real implementation can start.
2. International collaboration: Contacts have been established with the group lead by Frank Summers at the Space Telescope Science Institute in the USA working on metadata tagging the outreach images from the Hubble Space Telescope.
3. Funding: A proposal is currently being worked on for the EC. The goal is to give *"European-wide access to the treasure trove of celestial images from anywhere, anytime; moving astronomical images into the 21st Century"*.
4. Implementation: A student from the University of Copenhagen, Kasper Nielsen, is currently studying the practical issues related to ways of realising the Virtual Repository.

The Programme Group's web page is:
http://www.communicatingastronomy.org/repository/virtual_repository.html
containing background material and a list of members.