The Role of Astronomy in Society and Culture
Proceedings IAU Symposium No. 260, 2009
D. Valls-Gabaud & A. Boksenberg, eds.

doi:10.1017/S1743921311002912

Virtual Observatories

Françoise Genova

Centre de Données astronomiques de Strasbourg, Observatoire Astronomique,
UMR 7550 (CNRS, Université Louis Pasteur),
11 rue de l'Université, 67000 Strasbourg, France
email: genova@astro.u-strasbg.fr

Abstract. Astronomy has been at the forefront among scientific disciplines for the sharing of data, and the advent of the World Wide Web has produced a revolution in the way astronomers do science. The recent development of the concept of Virtual Observatory builds on these foundations. This is one of the truly global endeavours of astronomy, aiming at providing astronomers with seamless access to data and tools, including theoretical data. Astronomy on-line resources provide a rare example of a world-wide, discipline-wide knowledge infrastructure, based on internationally agreed interoperability standards.

Keywords. Astronomical data, Data centres, Virtual observatory, Knowledge infrastructure

1. Introduction

On-line access to information has produced a revolution for society, including the scientific community, and astronomy has been a pioneering discipline in those changes. The context was favourable: astronomy is based on observations, and the importance of keeping and re-using data has long been recognized. There are major scientific objectives, since we perform long term observations of variable natural phenomena, and observe a large number of objects, with complex interactions and on many scales. We use many different techniques, very diverse instruments on ground- and space-based observatories, and large surveys. Multi-wavelength observations now make a significant and increasing fraction of publications, and are often at the core of our understanding of the physical processes at work in astronomical objects.

Another reason for the sharing of data is to optimize the science return of investments in large infrastructures, telescopes and satellites, and in the manpower- and time-demanding large surveys. A good illustration is provided by Wamsteker and Griffin (1995), who study the early example of the *International Ultraviolet Explorer*. Five times more publications were produced from data retrieved from the archives than from the originally selected observing teams. IUE was operational from 1978 to 1996, the data is still available on-line†, and it is likely that this proportion has still increased since the time of completion of the study.

"Data" in astronomy is now a huge amount of heterogeneous, distributed information, which includes observation archives, added-value databases such as SIMBAD (Genova, 2007) and NED (Mazzarella & the NED team 2007), bibliographic data from the academic journals and the ADS (Kurtz *et al.* 2000), and also more and more often theory data. A key element for the sharing of data is the existence of common standards. For example FITS (Wells, Greisen & Harten 1981), which has been established more than thirty

† ESA handed over the responsibility of the IUE archive, INES (IUE Newly Extracted Spectra), "to the world scientific community" in 2000, and the Laboratory for Space Astrophysics and Theoretical Physics (LAEFF), based at Villafranca del Castillo, Spain, acts as INES Principal Centre. Several copies of the database are made available by data centres around the world.

years ago by radioastronomers, is now a critical asset of astronomy, since it allows any astronomer to use data from any telescope.

In the following, the network of on-line resources which was developed soon after the advent of the World Wide Web will be described (Section 2). Then the astronomical Virtual Observatory (VO) will briefly be presented (Section 3), as well as the role of data centres in the VO context (Section 4). Finally the specificities and impact of the astronomical knowledge infrastructure will be described (Section 5 - Conclusion).

2. The network of astronomical on-line resources

Some astronomical data centres and services were established long before the advent of the World Wide Web. As explained above, IUE is an early example of an electronic, accessible data archive. Another early example of data centre is the *Centre de Données astronomiques de Strasbourg* (Genova *et al.* 2000), which was created in 1972. One of its major services, SIMBAD, was implemented in 1983, from the merging of two databases which had been produced at the beginning of the 70's. These services were already available remotely.

With the Web, the provision of remote access to resources and the usage of these resources became much easier, and astronomical data centres seized very quickly this opportunity to provide on-line services. At CDS for instance, the first service on the Internet was an innovative one set up in partnership with the European journal *Astronomy & Astrophysics* (Ochsenbein & Lequeux 1995): "large" tabular data published in the journal were not printed any more, but rather distributed in electronic form by the CDS. At the beginning of the service in 1993, the tables were made accessible by ftp (and they still are). They are now also included in the VizieR service, which allows users to query them by constraining values in any column of the table. A standard describing the physical and scientific content of the tables was defined (Ochsenbein, 1994) and discussed with other data centres and journals. It is now the international reference for the description of catalogues and published tables. Several other journals have adopted similar policies, and publish the tables in their electronic version. This has led to a real change in paradigm: numbers previously only available printed on paper became data which could be reused.

Other services have rapidly become key tools for astronomers, one of the most prominent being the Astrophysics Data System (ADS). ADS is the reference for astronomical bibliography, used daily by the community. Most observatory archives are also providing access to their data holdings.

Bibliography providers such as ADS, the CDS, the journals, NED, also used the capacity of building links between their Web pages: the on-line services were immediately networked with each other, taking advantage of an existing standard. At that time there was no generic standard to describe bibliographic resources, and it took years before publishers agreed on the Digital Object Identifiers (DOI). But interoperability between bibliograhic resources had been an issue for astronomy data centres long before the Internet, because they were exchanging data. NED and SIMBAD had defined in common the bibcode/refcode, a 19-character coding of published references (e.g., 2006A&A...447...89T), in 1989 (Schmitz *et al.* 1995). The *bibcode* was later heavily used and extended by the ADS and the electronic journals, and it has been the key of the rapid networking of astronomical resources. Observatory archives soon joined the network, by providing links between observations and the papers which use them.

3. The astronomical Virtual Observatory

The next step was to go beyond access to on-line resources through their own Web pages (which requires to learn the specifics of each service) and beyond networking of resources through http links between pages. The aim of the Virtual Observatory (VO) is to allow seamless and transparent access to data centres, and to provide new analysis and visualisation tools taking in particular advantage of interoperability. The VO infrastructure is a standard structure for data centres to publish their data and services.

The Virtual Observatory is one a the few truly global endeavours of astronomy: the idea is to implement seamless access to resources available **world-wide**. Creating this single VO requires to agree at the world-wide level on common interoperability standards for data description, access protocols, etc. The definition of this infrastructure layer is the role of the VO projects and of the International Virtual Observatory Alliance (IVOA) that they have created. IVOA coordinates the definition of interoperability standards.

The VO is fully science driven, and its goal is to help astronomers to do their science - to make all astronomy resources accessible as if they were on one's own desktop. Astronomers do not have to know about technicalities, but it is fundamental that they provide their requirements and feedback on the services and tools to the VO projects.

4. Data Centres in the Virtual Observatory

Data centres are key actors of the Virtual Observatory: they populate it with data and services. For the first time, the European project Euro-VO Data Centre Alliance (EuroVO-DCA) † has performed a census of European data centres to get a global picture of the on-line resources available or potentially available in Europe. The definition of a data centre was the following: provide a *service to the community*, with some *added value* built on *expertise*, some kind of *sustainability* and some concern for *quality*. We chose an inclusive definition of a data centre to get the whole picture, and the census has actually provided a lively picture of the very diverse landscape. Well known data archives and data centres have shown up, as expected. It also appears that the VO development, which has been well publicized and explained, especially in certain countries, and through dedicated actions of the European VO project Euro-VO, has strongly encouraged smaller teams to provide data and services to the community. It is also noticeable that a significant number of data centres providing or willing to provide theory data or services registered.

On the other hand, a census of large projects performed at the European level by the Astronet ERA-NET ‡ showed that most large projects are willing to provide science ready, VO-compliant data archives, which is also very encouraging.

5. Conclusion - The astronomical knowledge infrastructure

Through the partnership of many actors, astronomy has been building a discipline-wide, world-wide knowledge infrastructure, with a huge diversity of resources:

- large services maintained by international or national agencies, providing archives of large ground- or space-based instruments

† The EuroVO-DCA project (2006-2008) has been supported by the European Commission (Sixth Framework Programme, eInfrastructure *Communication Framework Development* initiative, Project RI031675).

‡ Astronet was created by a group of European agencies in order to establish a comprehensive long-term planning for the development of European astronomy, and funded by the European Commission in the Sixth Framework Programme as a European Research Area Network (ERA-NET).

- large systematic surveys of the sky, results of large simulations
- generalist value-added databases and services
- smaller, focussed contributions of scientific teams which share their expertise with the community.

The model of this knowledge infrastructure is based on open access to data and services, and is fully distributed with no central point. Common interoperability standards are discussed in the international alliance of national VO projects, which is a best effort endeavour but has defined rules and procedures.

This knowledge infrastructure has brought a revolution in the everyday work of scientists. The availability of on-line data and services has changed the way astronomers do research and has been critical for the development of multi-wavelength astronomy. It is also an extremely powerful tool to integrate the scientific community at the world-wide level: all astronomers, wherever they are working, and provided that they have an Internet access, can use the best data and tools. This is also true for education, and an excellent opportunity to raise the sometimes decreasing interest of students for science.

The sharing of knowledge is a global challenge for science and society. The way it can be organised is strongly dependent on the history and culture of each discipline. The astronomy knowledge infrastructure is a rare example of an operational, discipline-wide, world-wide knowledge infrastructure, widely used by the scientific community. VO-like projects are emerging in many disciplines, and we are willing to share lessons learnt.

References

Genova, F. 2007, in R.A. Shaw, F. Hill & D. J. Bell (eds.), *ADASS XVI, ASP Conf. Ser.*, 376, 145

Genova, F., Egret, D., Bienaymé, O., Bonnarel, F., Dubois, P., Fernique, P., Jasniewicz, G., Lesteven, S., Monier, R., Ochsenbein, F., & Wenger, M. 2000, *Astron. Astrophys. Suppl. Ser.*, 143, 1

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C. S., Murray, S. S., & Watson, J. M. 2000, *Astron. Astrophys. Suppl. Ser.*, 143, 41

Mazzarella, J. M. & The NED Team 2007, in R. A. Shaw, F. Hill & D. J. Bell (eds.), *ADASS XVI, ASP Conf. Ser.*, 376, 153

Ochsenbein, F. 1994, *BICDS*, 44, 19

Ochsenbein, F. & Lequeux, J. 1995, *Vistas in Astronomy*, 39, 227

Wamsteker, W. & Griffin, R. E. M. 1995, *Astrophys. Space Sci.*, 228, 385

Wells, D. C., Greise, E. W., & Harten, R. H 1981, *Astron. Astrophys. Suppl. Ser.*, 44, 363