Baltic Astronomy, vol. 9, 672–675, 2000

ASTRONOMICAL DATA ACQUISITION, PROCESSING AND STORAGE: MODERN FACILITIES – CONCLUDING REMARKS

F. Pasian

Astronomical Observatory, Via Tiepolo 11, 34131 Trieste, Italy

Received September 1, 2000.

Abstract. A summary of the topics presented in Section 11 of the JENAM 2000 conference is attempted, and some remarks related to collaborative work and funding strategies are made.

Key words: astronomical databases

1. INTRODUCTION

While existing telescopes continue producing data and scientific results, and while historical data (e.g. photographic plates) need to be kept in proper archives and need to increase their usability, bigger telescopes and larger and more efficient detectors both from ground and from space are becoming available in this turn of century and millennium. We are getting into an era in which new-generation instruments will be producing terabytes of data a day. We are therefore confronted with non-trivial problems to solve in terms of handling these data, storing them so that they do not get lost, process them at the best of our competence to extract useful information which will help us to enrich our knowledge in astrophysics. Finally, we wish to share the results with the rest of the astronomical community, and thus provide the community with archives of calibrated and/or scientifically analyzed data.

We have tried to address such problems during these three very interesting days and we have had a good overview on what is going on in this field. In the following, some specific topics which are central

for research in the field of data management are extracted and briefly discussed.

2. MAIN TOPICS

Certainly the main topic discussed in Section 11 of JENAM 2000 is the one related to *archives and databases*. The contents of some of them were described, and the presentations themselves outlined the wide variety of interests and knowledge available in the community. It has been correctly stated that each archive knows best its data and can maintain them in the best way, because the expertise resides there: the key task of an archive is then to allow remote access to the data while sharing with the users the expertise required to "understand" the data. An extension of data archives can be therefore considered to be the distributed information systems, such as OASIS or IDIS, which aim at reaching the users community at their own sites, providing besides access to science data some value-added functionality as well.

The keyword for information systems to be really useful is *interoperability*, the way to see many archives and databases within a unifying view. For this to be technically feasible, and to happen properly from a scientific point of view, a crucial step is to solve the *cross-identification of sources* throughout various catalogues. The issue was discussed at length during the GCVS presentation and the MASCOT session. In these field there is a need to develop new collaborations and to improve the existing ones. A derived item, but of the greatest importance, is the need to allow *intelligent retrieval*, to gather the useful information from huge quantities of data. The "Kohonen map" search in bibliographical sources which has been implemented and is now active at the CDS is an excellent example, which needs to be followed by other data mining projects before the risk of drowning in the sea of information we produce ourselves becomes reality.

Another essential point for the data to be shared is the use of *standards*. The use of FITS is a must, and if the current standard does not meet the requirements set by the specific data we produce, an extension may be asked to a proper body (e.g. the European FITS Committee). From the point of view of designation of sources, the directions defined by the IAU Task Group on designations shall be followed.

The *technicalities* to build databases and archives were also discussed from the implementators' point of view (e.g. by the TNG and the BTA and RATAN 600 groups). In the discussion on tools to be used, a particular place was held by Database Management Systems. There is a wide variety of products being used: from the commercial ones (e.g. Oracle, Objectivity) having a high price, to lower price products, to public domain software, to code developed in-house. When cost becomes a major issue, then portability of software over different computer platforms, but also reusability of code by different projects and minimization of efforts would suggest to use software already available (e.g. public domain stuff accessible through Internet) rather than developing new systems.

A word of warning was spent on the problem of *degradation of electronic data.* While in the past photographic plates have guaranteed an efficient storage of data for many decades, the fast-cycle obsolescence of electronic media and devices call for technical maintenance of Data Centers. Magnetic media need to be regularly refreshed (i.e. implying dedicated manpower), while careful planning of media migration needs to be made to avoid the risk of having perfectly maintained media (magnetic and/or optical) with no device capable of actually reading them.

On a related topic, the *huge data flow* expected from new-generation instruments (examples are TERAPIX and the Wide-Field Imager at the LBT) is a problem in itself. Careful design of the tools for data handling and proper dimensioning of the hardware are necessary just to be able to acquire the data and to store them properly for further use.

Processing algorithms are another key point in the field of data management. It has often happened lately that the human skill in developing new ways of reducing/analyzing data does not go as fast as the development of computer power. During the session, there were presentations about

– increasing the efficiency in treating large quantities of data;

– processing unusual types of data (having unusual formats, or missing items);

– treating massive quantities of historical data (e.g. photographic plates) which are essential for long-scale variability studies.

Remote access to observing facilities was also briefly discussed in Section 11, both related to telescope control (e.g. in the Special AO), and to observation by astronomers. The issue of flexible scheduling

was mentioned, to allow the observing programs requiring special conditions to be carried out at the appropriate time.

From all discussions involving remote work, it has been noted that the *network* is still a problem sometimes and in some places. This is detrimental to our work, since it reflects in bad access to information.

3. CONCLUSIONS

The presentations and the discussions carried out during Section 11 of JENAM 2000 show that the community involved in data management (acquisition, processing and storage) is extremely lively and the results to be expected from the current activities are of great interest. This is particularly important, since archives are a democratic way of accessing top-level observational facilities: due to the over-subscription of all large observing facilities, archives can be an invaluable source of high-quality material for scientific and educational purposes. But to be really useful archives and databases of derived data need to be properly harmonized, to become a tool even more important for the community. The discussion on this harmonization must go beyond the boundaries of institutes or project teams, since funding constraints limit both the lifetime of projects and the operational capabilities of individual institutes.

And this leads to the problem of the procurement of funding to support of these scientific activities. The recent news on the withdrawal of IAU support to the GCVS group is a worrying sign. On the other hand, the IAU and EU support to the wide field plate archive project goes in the opposite direction. The limited funding the modern society is willing to offer to science implies that we need to survive in an environment which is getting far more competitive. But competitiveness shall not limit collaboration: on the contrary, exchange of scientific and technical information shall be fostered to create stronger and more competent groups across national boundaries. Scientific activity in data management must be politically visible at all the appropriate levels, to convince funding agencies that it is worth investing in science. Furthermore, ways should be pursued to define collaborative projects having high scientific and technical profiles to be submitted to proper super-national institutions (e.g. the EU) for funding.

Contents of Volume 9

Number 1

PROCEEDINGS OF THE FIFTH WET WORKSHOP, AUGUST 15–20, 1999

Number 2

PROCEEDINGS OF THE FIFTH WET WORKSHOP, AUGUST 15–20, 1999 (Continued)

Number 3

PROCEEDINGS OF THE FIFTH WET WORKSHOP, AUGUST 15–20, 1999 (Continued)

Number 4

PROCEEDINGS OF THE CONFERENCE "ASTRONOMICAL DATA ACQUISITION, PROCESSING AND STORAGE: MODERN FACILITIES", MAY 29 – JUNE 3, 2000