The Astronomical Information Infrastructure from the End-User Perspective

Sake J. Hogeveen

Astronomical Institute, Utrecht University, Princetonplein 5, P.O. Box 80 000, NL–3508 TA Utrecht, The Netherlands.
E-mail: *S.J.Hogeveen@fys.ruu.nl*

Information Technology (IT) today has found so many applications in astronomy, that we may speak of an electronic 'Astronomical Information Infrastructure' (AII). At this moment, the AII really is nothing but a collection of disparate services. Over the last few years the collection has grown so large, that it leaves the End-User in a rather desperate state. This paper provides an inventory of End-User problems, regarding network information systems, astronomical software packages, and electronic publishing. Solutions to the problems may be found through the coordination of efforts, and the definition of standards and principles. Examples of what may be achieved are the AstroWeb project, which provides a directory of astronomical IT services, the Starlink project in the United Kingdom, concerning astronomical software packages, and the initiatives of the American Astronomical Society with respect to electronic publishing. The paper also discusses the concepts of a modular client–server system, which may serve as the basic ingredient for a transparent and integrated Astronomical Information Infrastructure.

Introduction

The scientific practice of an astronomer involves literature and archive research, writing proposals, making observations, data reduction and analysis, building models, developing theories, writing papers and publishing them, collaborations and education. Due to revolutionary developments in Information Technology (hardware, software, *user interfaces*, and networking) during the last few years, computer applications are now available to assist the astronomer in each of these activities.

In fact, the applications are so many and so versatile, that we may speak of an electronic Astronomical Information Infrastructure, along side the traditional

information infrastructure of printed journals, letters, preprints, proceedings, transactions, (IAU) telegrams and circulars, archives of printed catalogues, glass plates, *etcetera*.

As a science software engineer at the Astronomical Institute of Utrecht University, I am, together with a colleague, supporting about 25 astronomers (permanent staff, post-docs, PhD students) in their use of computer facilities. At our institute, we are true End-Users [1] of astronomical computer and network facilities, as we only use them to the benefit of our astronomical research. We do not actively participate in the development of astronomical software, other than for local purposes.

We have made an inventory of the problems we encounter in the use of the facilities as they are provided by software developers (astronomical organizations, and commercial vendors), providers of network information systems, publishers, *etc.* We feel that our inventory touches on some of the more general aspects involved in the development of an *integrated* electronic Astronomical Information Infrastructure.

In this paper, we have grouped our findings into three sections, addressing the aspects of, respectively, network information systems, software packages, and electronic publishing. Section 4 discusses the concepts of a *modular client–server* system, that could help alleviate the problems reported in the earlier sections.

1 Network information systems

1.1 Locating facilities; integrated functionality

One of the first problems concerning network information systems is that of locating services. Today many organizations and institutes are making efforts to bring their own information on-line, but actually none of them has an obligation to provide an overview of all the astronomical on-line facilities that are currently available. The need for an overview has been recognized by Heinz Andernach, Robert J. Hanisch and Fionn Murtagh, and is addressed in their paper *Network Resources for Astronomers*[1]. Their paper is also available in an on-line *hypertext* version, [2] which allows easy access to the facilities and services mentioned in the paper. An even more ambitious programme to provide an overview, or directory service of on-line astronomical information

[1] Throughout this paper the 'End-User', the 'astronomer' *etc.* is referred to as 'he', to indicate the holder of the position: not his or her actual gender.

[2] http://www.hq.eso.org/online-resources-paper/rrn.html

and facilities, is the AstroWeb project[2]. This initiative is supported by a 'consortium' of seven institutions.[3] It seems the AstroWeb project is well established now, and that the members of the 'consortium' have *de facto* assumed the responsibility to provide this service for a longer period of time.

Currently, the AstroWeb service may be regarded as a directory or 'yellow pages' service, and the various participating institutes are providing several, different ways to access and search the master database. The master database contains records in an agreed 'data interchange format', which accommodates key words, and short and long descriptions of an entry. This should allow for sophisticated indexing and, in combination with e.g. a WAIS[4] search engine, free search capabilities. We understand that work on such facilities is in progress. They are needed, to make the electronic AII as transparent and accessible for the End-User as the paper AII, with its long standing tradition of indexing and abstracting through libraries, library catalogues, and printed volumes of abstracts.

Quite another matter is that of the *integrated functionality* of services. The End-User is very interested in efficient ways of feeding the information, which he retrieved from one service, into another, so he may, in successive steps, obtain the information he is interested in. A researcher could, for example, start his investigations concerning a particular astronomical object with the SIMBAD[5] database. This provides him with references to literature and data sets. The researcher will then want to retrieve some of those data sets and the software to analyse them. Today he needs to consult various sources separately, to find out if, and where, these data and software are available at a site which may be accessed through the Internet. The subsequent retrieval and installation of data and software again require several actions from the researcher or his system manager. Effectively, the researcher needs to be familiar with things like 'ftp', 'archie', WWW, operating systems, *etc.* to be able to use the various on-line facilities to his benefit. Obviously, one would like the required actions to be simplified and automated, allowing the researcher to concentrate on the science, rather than on computer and network trivia.

In fact, there are two systems that provide precisely this kind of automated and integrated functionality: the Astrophysics Data System, ADS, operated by NASA, and the European Space Information System, ESIS, operated by ESA. However, at this moment in time the continuation of both systems is

[3] Involved in the AstroWeb project are André Heck and Daniel Egret (CDS), Anton Koekemoer (MSSSO), Don Wells (NRAO), Sergio Paoli (La Plata), Hans-Martin Adorf and Fionn Murtagh (ST-ECF), Bob Jackson (STScI), and Jose Daniel Ponz (VILSPA).

[4] Wide Area Information Systems.

[5] Set of Identifications, Measurements, and Bibliography for Astronomical Data, operated by the CDS, the Centre de Données astronomiques de Strasbourg.

rather uncertain (see section 1.2). The scope of the integrated functionality is, for both the ADS and ESIS, limited to the tasks that have been built into each system. Since no single effort will ever be able to cover all of astronomy, ways will have to be sought which allow integration of functionality across different information systems. This may be achieved by defining or adopting a data exchange and query format, similar to the FITS standard for filing astronomical data, and SQL, the Structured Query Language. Apart from this data exchange format, one would require a flexible or modular client–server system, with a client that is able to send data, that were retrieved from one server, to another, either directly or after some modification. The concepts of such a modular client–server system are discussed in section 4.

1.2 Continuity and stability of services

The ADS and ESIS both are the result of a major development effort. When they were made available to the astronomical community in 1993, they were state-of-the-art *distributed* systems for the search, retrieval, reduction and analysis of astronomical data. It seems that NASA and ESA considered these systems mainly as research and development projects, because after being operational for only two or three years, the parts of the ADS and ESIS systems, relying on *dedicated* client–server systems with their specific *graphical user interfaces*, will be closed down. Only the pace at which this will happen differs from within one year for ESIS, to several years (?) for the ADS.

The ADS and ESIS services will in the future be provided through the *general purpose* network information system World Wide Web, employing the *hypertext* client–server system on which the WWW is based. But particularly in the case of ESIS, where parts of the system will be transferred to other organizations and institutes, it remains to be seen whether the full functionality will, or even can, be provided in the same, or at least a very similar, way.

It is needless to say that these developments do not serve to increase End-User confidence in network information systems. End-Users are interested in *continuity* and *stability* of services, otherwise they will not invest time and effort in getting acquainted with a particular system and in learning how to use it.

1.3 General purpose systems

Almost parallel to the *dedicated* client–server systems of the ADS and ESIS, several *general purpose* network information systems were developed, notably Gopher, WAIS, and the World Wide Web. We call them general purpose

systems, because their design allows them to be employed for the presentation of a wide variety of information, across scientific disciplines. The capabilities, of the *hypertext* based World Wide Web system in particular, are so versatile, that there is much less need for *dedicated* client–server systems.

The same remarks regarding continuity and stability apply with respect to astronomical information systems employing these general purpose client–server systems, as they were made concerning the dedicated systems of the ADS and ESIS in the previous section. The End-User should not have to worry about whether a particular system or service will still be here tomorrow, nor about what it will look like by then or how it should be operated. Of course, we are only just discovering the technology of network information systems, and both information providers and information seekers will have to explore the technology to find out about its capabilities. Changes are inevitable, as new information or new capabilities are added to the systems. However, when new systems are initiated, with the intent of providing a lasting service, changes to a system once in existence should be of an evolutionary nature rather than revolutionary, otherwise the End-User will loose track.

2 Software packages

2.1 Architecture, integrity, standardization, documentation

Software packages for the reduction and analysis of astronomical data are mostly developed at institutes related to space or ground based observational facilities. Most well known reduction packages, like Midas, Iraf, Xanadu, *etc.* were, at least initially, developed by scientists, from a scientific point of view, and not by software engineers. This is, until today, for most packages reflected in their architecture. Which means that they each have their own, very specific requirements about how and where they should be installed on a computer, or a cluster of computers. Often the reasons behind a particular set up are quite valid, and they can certainly be understood from a historical point of view, but they do not always coincide with what is practicable at a random institute, where these packages have to be installed. Crucial points in the development of many packages were the migration from VAX VMS to Unix based platforms, and from single station, stand-alone operation to operation in network environments with clusters of workstations.

Installation of these software packages at any given astronomical institute is often a laborious and pain-staking operation. Trial and error, and re-compilation of parts of the system, will eventually lead to a working local implementation of the package. But this practice is likely to compromise the integrity of the

package, jeopardizing the reliability of the scientific results.

Many of the installation problems might be overcome if the astronomical software packages were made to comply to the standards as set forth by the GNU project.[6] The GNU (Gnu's Not Unix) project was initiated by Richard Stallman, and it is aimed at developing and distributing free, sharable software. The GNU standards, or rather the principles underlying the standards, make sure that software packages complying to them are highly portable and easy to install. Where most astronomical software packages generally allow for only one, very rigid installation configuration, GNU software allows for almost every configuration one can think of: single user, multi user, special account, *etc.* Moreover, the installation of the software is automated. When the automatic installation procedure completed itself successfully, the chances of errors in the local implementation are minimal.

Many institutes are left to themselves when it comes to the installation and maintenance of astronomical software packages. In the United Kingdom a unique project was initiated in 1980. It is called STARLINK, and it looks after the installation and maintenance of many astronomical software packages for the astronomical community in the UK. Institutes that join STARLINK automatically receive updates of software overnight, without local system managers having to worry about each individual package or update. It hardly needs pointing out that this allows for a much more efficient allocation of human resources.

Another problem is the availability of software packages for different platforms. Generally, a software package is developed on a particular piece of hardware, and depending on the demand, ports to other platforms are made. It is not always possible to persuade a vendor to make a particular piece of hardware available (at little or no cost!) for porting a particular software package to that platform. However, the quality of todays networks allows for 'pooling' of hardware among astronomical institutes, thus allowing developers to remotely create a port on a hardware platform at a site which they may never visit in person.

The World Wide Web allows documentation of software packages to be made available on-line. At institutes where several scientists are using the same software package, the whereabouts of the paper manuals are always a problem. System managers at several institutes have made efforts to bring documentation on-line. Since these again are duplicate efforts, institutes and organizations that provide the astronomical software packages should be urged to also provide their documentation in an on-line version, when they are not already

[6] The GNU 'manifesto', and the text of the GNU standards, come with every distribution of GNU software. A very well known piece of GNU software is the Emacs editor.

doing so.

2.2 General purpose software packages

There are several 'general purpose' software packages in use in astronomy, like IDL (the Interactive Data Language) from CreaSo, and Mathematica, the system for 'doing mathematics' from Wolfram Research Inc. Part of the strength of these packages lies in their 'user libraries': collections of routines written by users, containing algorithms for frequently occurring astronomical problems. For IDL there is an astronomical user's library; for Mathematica there currently is not. But for Mathematica there are libraries from mathematicians, engineers and physicists, and it might be worthwhile to investigate the usefulness of these libraries for astronomers. End-Users would benefit from a repository where reviews of libraries, or of specific routines, are being deposited.

3 Electronic publishing

3.1 Submission: author mark-up

Since a few years ago, all the major astronomical journals accept *compuscripts*: manuscripts of papers, prepared by authors by means of the 'computer typesetting' and 'document processing' systems TEX and LATEX, which are submitted on computer media or through the networks.

Currently, every major journal provides its own (La)TeX macro package, by means of which authors are supposed to prepare their compuscript for a particular journal. The macro packages serve two purposes: one is to obtain the specific *layout* for a particular journal, and the second is to provide mark-up tags for astronomy specific constructs (notably symbols), which are not available in standard TEX and LATEX. The extra tagging facilities differ from one macro package to an other, so the author needs to be aware, from the outset, to which journal he will submit his paper.

Only the American Astronomical Society provides a general macro package covering several journals, with additional macro packages taking care of the (mainly typographical) fine details of a particular journal. The AAS claims[3] that their LATEX macro package allows an author to concentrate on the contents of his paper, rather than on its layout, by the concept of *generalized* mark-up. However, every macro package based on LATEX, including the one provided

by *Astronomy and Astrophysics*, largely adheres to this concept, because it is inherent to LaTeX.

End-Users would benefit considerably from the development of an AstroTeX package, which combines the astronomical tagging constructs required by each individual journal, and allows the typographical requirements of individual journals to be met by additional or optional macro packages. They would then only have to concern themselves with the tagging requirements of the common AstroTeX package, while the additional packages would allow them to see what their paper would actually look like (layout) in a particular journal.

The development of AstroTeX would also allow for a considerable improvement of user support. Currently, the support of astronomical (La)TeX macro packages is left to local system managers and TeX 'Gurus', who have to familiarize themselves with the different packages, and cope with their anomalies and other shortcomings. When AstroTeX becomes a project, supported by all astronomical publishers and organizations, on-line 'help desks' could be set up to which individual astronomers could then refer for support.

In 1994 a new LaTeX version, LaTeX 2_ε[4,5], was released. Maybe this release is a good moment to start working on AstroTeX, and combine in it the efforts, that would be required from the various publishers and organizations anyway, to upgrade their individual macro packages. At the same time, proper BibTeX macros could be developed for the efficient compilation of lists of references. The WWW, combined with systems like WAIS, allow for the setup of a *distributed* database of astronomical references in BibTeX format, from which every astronomer could benefit.

3.2 Distribution: the Electronic Journal

The next step, after electronic submission of manuscripts, is the electronic distribution of astronomical journals. In fact, the AAS has already started publishing volumes of its journals on CD-ROM, and the *Astrophysical Journal Letters* are currently available on-line. The ADS provides on-line scanned images of ApJ (Letters) from 1975. In Europe, the CDS provides on-line abstracts of volumes of *Astronomy and Astrophysics* and the P.A.S.P.

These initiatives indicate what the first electronic journals will look like: very much a linear continuation of the printed journal, but distributed with electronic means. The electronic journal allows the astronomer to browse *hypertext* versions of abstracts, and even full papers, on-line. When the astronomer has found the information he is looking for, (portions of) the paper may be retrieved, to produce a paper copy on a local printing device.

Traditionally, publishers are responsible for the production and distribution of printed journals, and libraries are responsible for their archiving, to thus ensure prompt and indefinite access. Initially these responsibilities will not change much, when publishers are to provide the on-line version of the current (annual) volume of a journal, and libraries to collect the CD-ROMs of completed annual volumes.

However, it is quite likely that todays electronic journals will evolve into electronic 'information systems', taking much more advantage of the capabilities of computerized systems. These information systems will not only allow for the electronic publication of scientific papers, but also for the inclusion of fragments of computer code, visualizations, animations and the like. Contributions to information systems may not even be in the form of papers anymore, but of pieces of computer code, which add functionality to the system, for example for the symbolic computation of magneto-hydrodynamic processes under astrophysical conditions.[7] Also, sophisticated search and retrieval capabilities will be developed. Some of these services may be offered by publishers. Part of the archiving is then no longer in the hands of university libraries or astronomical organizations, but in those of commercial enterprises. This requires safeguards against the loss of scientific information, for those occasions where a commercial information provider is unable to continue his services.

We must be cautious for high expectations regarding the development of these astronomical information systems. In the end, it is the scientists who have to contribute to them. To be able to do so, they will have to adapt to the concept, and master the tools that allow them to produce their contributions. Such developments call for the assistance of IT specialists, with a background in astronomy.

4 The modular client–server system

To accommodate information exchange through Internet, several protocols have been developed: telnet, ftp, archie, gopher, veronica, WWW, wais, *etc.* Modern *clients* of network information systems, like Mosaic and Netscape, 'speak' several of these protocols, allowing a user to connect to information servers based on a variety of protocols. Expectations are that future *clients* will be of a modular nature, each module being tuned to a specific protocol. A *client* may be updated to accomodate a new protocol or new features, simply by adding or replacing a module.

[7] Authors of such contributions may receive full scientific credits, because a contribution will be evaluated by referees before it is added to the system.

The concept could be used in astronomy, to add functionality to an astronomical network information system, but also for the installation of software packages. The first step in software installation would then be to download a special module, related to the specific package, which completely automatically takes care of the subsequent retrieval, installation, and *verification* of the software, much along the lines discussed in section 2.

Concluding remarks

Turning the currently available collection of astronomical IT services into a transparent and integrated electronic Information Infrastructure, requires a coordinated effort. Initiatives, like the AstroWeb project concerning network information systems, the Starlink project concerning astronomical software in the UK, and those of the AAS concerning electronic publishing, are indicative of what may be achieved, and deserve to be extended to the entire astronomical community. Also, the AII can be made 'intrinsically coherent', by adopting or establishing certain standards and principles. Efforts in this direction will increase the ease and efficiency by which the End-User, the astronomer, can do his scientific work. It will also allow for a much more efficient allocation of human resources (system managers, software engineers), at local, regional, and global levels, towards the development, maintenance, and support of the AII.

References

[1] H. Andernach, R.J. Hanisch, and F. Murtagh, *Network Resources for Astronomers*, Proc. Astron. Soc. Pac. 106, p. 1190, 1994.

[2] R. Jackson, D. Wells, H.M. Adorf, D. Egret, A. Heck, A. Koekemoer, and F. Murtagh, *AstroWeb. A database of links to astronomy resources*, Astron. Astrophys. Suppl. 108, 235-236, 1994.

[3] American Astronomical Society, *Special Insert on Electronic Publishing in Astronomy*, AAS Newsletter 62, 1992.

[4] L. Lamport, *LaTeX, A Document Preparation System*, Addison-Wesley, Reading, Massachusetts, 2nd ed., 1994.

[5] M. Goosens, F. Mittelbach, and A. Samarin, *The LaTeX Companion*, Addison-Wesley, Reading, Massachusetts, 1994.

The hypertext version of this paper also contains references to on-line sources. They are in the form of *hyperlinks*, which are invisible in the paper edition.
URL: `http://cdsweb.u-strasbg.fr/waw/hogeveen/hogeveen.html`.