Astronomical Data Analysis Software and Systems IV
ASP Conference Series, Vol. 77, 1995
R. A. Shaw, H. E. Payne, and J. J. E. Hayes, eds.

WWW as a Support for the Long Term LBT Archive

L. Fini

Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, I-50125 Firenze, Italy

Abstract. Any large and complex project has the need to build and maintain an archive of all of its related information throughout its life. The archive content is typically multimedia in nature, consisting of drawings, pictures, papers and documents with different formats, spreadsheet data files, plots, and so on.

When a project is the common effort of a number of institutions, as in the case of the Large Binocular Telescope Project, where several institutions in different countries are cooperating, the most natural structure of the archive is a distributed one, where each party has responsibility over a part of the database, but shares the data with the others. A completely distributed system, i.e., a system where a given data item is stored in a single site, would be impractical for both security and access efficiency reasons. A mixed approach, where stable data items are duplicated at all the participating sites, while the work-in-progress ones are stored only where they are generated, seems to be a safer one. WWW services can then provide a consistent access to the whole archive for all the involved parties.

1. The LBT Project

The Large Binocular Telescope (LBT) is a telescope for infrared and optical wavelengths, equipped with two 8.4 m mirrors on a single mounting, to be located in the northern hemisphere at Mt. Graham (Arizona, USA). The technical details of the project are outside the scope of this presentation and can be found elsewhere (Salinari & Hill 1994; Hill & Salinari 1994).

The project is jointly developed by a number of institutions in different places: the University of Arizona, the Arcetri Astrophysical Observatory in Florence, and the Research Corporation in Tucson, together with some contractors both in USA and in Italy. All the involved parties will need to interact, cooperate, and exchange information throughout the project life. The LBT project is thus "distributed" in nature; pieces of information such as papers, reports, drawings, etc., are produced at far apart places and must be efficiently shared among all the involved parties.

2. The LBT Archive

The LBT archive is an information system which stores a huge number of items (or data files) from various sources and of different nature, gathered during the development phase and the operating life of the telescope. These include: (1) scientific papers, such as scientific justifications of the project, results of the preliminary studies, and so on, (2) technical reports, (3) drawings: mechanical, optical, electrical, etc, (4) datasheets of the commercially available parts and subsystems, (5) software documentation, including the full source code, (6) manuals, with different purposes: operating, troubleshooting, maintenance, observing, etc., probably in hypertext format, (7) statistical data, to maintain an historical log of weather conditions, seeing, instrument related data, etc, (8) observing Programs, to maintain an historical log of telescope usage, and (9) operation log: complete data on telescope operation to support troubleshooting, to improve usage efficiency, etc. All the above data items must be accessible "on-line" throughout the entire telescope life. A considerable fraction of them will continuously grow in size.

The LBT archive will be of central importance for a number of activities: coordinating telescope design and development, supporting instrumentation design and construction, properly operating the telescope, troubleshooting and maintenance, supporting astronomers who want to submit observation programs and the board which must review submitted programs, and refurbishing and modifying the telescope during its operating life.

3. Data Types

Due to the number of different sources of the data files to be stored in the archive, it will not be possible, or even desirable, to define a fixed list of supported formats. even though a set of standards can be agreed upon in order to avoid an uncontrollable explosion of formats. In the future, new formats may be adopted, perhaps requiring the conversion of existing data files; new tools could be developed; sounder standards could emerge; and so on. Although a part of the archive could be stored in the form of "finalized" documents (e.g., using PostScript, GIF, etc.) many data files will need to be stored in some "source" form (e.g., TEX/LATEX, HTML, and various native formats for word processors, spreadsheets, CAD systems, databases, etc.) because they will be subject to modifications and updates.

All the supported formats will require the availability of adequate "viewers" or browsers to be easily used; this will likely set the ultimate constraint to the number of supported formats, especially when the long life span of the archive is considered.

4. Archive Requirements

The LBT archive system must fulfill a number of conflicting requirements:

- **Generality**. It must support and manage a number of different standards and file formats.

- **Durability**. It must follow the entire project life, which is a pretty long time, especially when the development speed of the supporting technologies is considered. It must, therefore, survive a number of technological updates and changes.
- **Security**. The archive must survive various kind of failures, both those of the supporting hardware and those due to human errors.
- **Flexibility**. It must allow complete reconfigurations of the supporting hardware/software structures. It must allow easy introduction of new formats.
- **Expandability**. It must allow the growth in size of its content, possibly by adopting new storage technologies as soon as they are needed and available.

5. A Distributed Mirrored Archive

The LBT archive will be hosted on many nodes of a LAN. Some are the nodes used for the development of the data items themselves, and so store the dynamical part of the archive. One or more other nodes may have only archiving functions, storing the static data files.

The "active" workstations (i.e., those where the development is actually performed) are integral parts of the archive. In this way a single, unified approach can be used to access both the "dynamic" and the "static" pieces of information.

6. Data Duplication

Most of the archive contents will be duplicated at the participating sites in order to obtain faster access to data, to lower the demand for network bandwidth and, at the same time, to increase the global system security; storing many copies of sensible data in far apart places is traditionally the best way to guarantee data integrity when faced with hardware and human failures and environmental threats.

7. Data Dissemination

The implementation and maintenance of the LBT archive will require the development of procedures to ensure the proper synchronization of the copies of the archive, i.e., to provide for the distribution across the network of new data items from the sites where they are generated to the other sites.

A number of measures can be adopted in order to increase the efficiency of the process and/or minimize the usage of the network bandwidth. Files to be downloaded could be compressed before transmission and decompressed before storing (although in many cases the compressed form could be stored instead). Sites could agree upon best time for downloading based on locally defined tables of "transmission costs." Huge downloads could be split across several days or, perhaps, delayed to next holiday.

Proper crosschecks must be performed after each downloading session to ensure that each updated file is received and that its content is not corrupted. For this purpose, the update lists will contain checksum information about the files so that file integrity can be verified.

8. Data Access and Browsing

The most important function of the archive, at least from the point of view of operations, is the ability for users to navigate through lists of data items, perform searches, access and view single data items, etc. Hypertext documents will impose a logical structure on the actual physical layout of the archive. This scheme allows many independent logical structures to be defined, each suited to the particular needs of a different kind of user. Other functions are needed in order to allow efficient retrieval of data items, such as searching through lists with various selection functions, textual searches in documents, and so on.

9. Tools

The main procedure for updating the archive is structured as a client-server system; each node will periodically run the client procedure, requesting update lists from the other nodes and downloading the required files. Each node, when requested, will start the server procedure to fulfill the request. The client-server mechanism is supported by the FORM capability of the *http* protocol, and is implemented with *Perl* scripts. Further work will include development of a user interface based on the *Mosaic* browser, and the implementation of searching functions, for which the *GlIMPSe* system is being explored.

References

Salinari, P., & Hill, J. M. 1995, Proceedings of the SPIE Conference on Large telescopes (Hawaii, SPIE), in press

Hill, J. M., & Salinari, P. 1995, Proceedings of the SPIE Conference on Large telescopes (Hawaii, SPIE), in press