

The ISO Data Archive

C. Arviset & T. Prusti

ISO Data Centre, ESA Directorate of Scientific Programmes,
Villafranca, Spain

Introduction

ISO was the world's first true orbiting astronomical infrared observatory. During its in-orbit lifetime from November 1995 to May 1998, it made about 30 000 individual observations of many types of astronomical objects, from within our own Solar System out to the most distant galaxies. These observations were made with a wide variety of spectral and spatial resolving powers and at wavelengths from 2 — 240 microns. The resulting database provides a treasure trove of information for further astronomical research.

The Infrared Space Observatory (ISO) Data Archive, developed by the ISO Data Centre in Villafranca, Spain, offers the astronomical community fast and easy access to all ISO products and related information through a pioneering World Wide Web (WWW) interface. The first release – issued on 9 December 1998 at <http://www.iso.vilspa.esa.es/> – has already been accessed by many astronomers and visitors wishing to query the ISO database and to retrieve important scientific data.

During the in-orbit operations, the data from the satellite were processed on a daily basis. In addition to being distributed to the astronomer who had requested it, each observation was loaded into an ISO product archive – essentially for operational use by experts located at the ISO Science Operations Centre in Villafranca, Spain.

ISO is now in a 3.5-year long 'post-operations' phase, designed to maximise its scientific exploitation. All data, along with the associated calibration information, had to be made readily accessible to the worldwide scientific community. This involved building a new and more modern archive structure and user interface, as well as reprocessing all the observations with the latest version of the 'pipeline' data processing software.

The resulting ISO Data Archive was opened to the public on 9 December 1998, only six months after the end of the operations. Through a pioneering user interface based on Java, complex queries can be issued to the

ISO database. Textual and visual presentations of the data aid in selecting observations for retrieval via FTP (File Transfer Protocol).

The ISO Data Archive was designed and developed at the ISO Data Centre in Villafranca. Due to continuous and fruitful cooperation between users and developers, it provides a unique, state-of-the art astronomical data archive.

Requirements for a new archive

Approximately one year before the end of ISO operations, it was decided to re-evaluate the existing ISO product archive concept. The new archive had to be more general-user oriented, open to the external world through modern and powerful technologies (e.g. WWW), and require less human intervention for its establishment and maintenance.

The approach adopted was to build something new out of the internal archive that would not only reuse, where possible, the existing technical and human expertise, but would also meet the following primary new requirements:

- open the archive to the external world via a fast connection to the Internet
- a WWW interface to the archive
- more powerful means for queries and retrieval requests
- proprietary data: access to each observing programme restricted to specified astronomer(s) for the first year after its completion
- provision of so-called 'browse products' to provide a fast and accurate impression of the data for each observation
- automatic processing of the ISO products from the telemetry
- uplink and downlink data held in a single database
- modular and flexible design that can evolve with additional user requirements.

A workshop was held in March 1997 involving ISO experts from the instrument institutes and the Science Operations Centre, as well as experts from other astronomical archives. This

led to a concept document, which defined the foundations of the new ISO Data Archive. Figure 1 gives a schematic view of the system as seen by the user.

A second workshop, involving astronomers with ISO experience ranging from novice to expert, defined the detailed user requirements and produced the first version of the ISO Data Archive User Requirements Document.

In parallel, an engineering team was established to follow-up the various software concept, design and development phases according to the ESA PSS-05 software engineering standards.

Design and development

The overall approach included regular interaction between the users and the development team in keeping with the ‘formality’ frame as defined in PSS-05 in order to avoid modifications without control. Therefore, many reviews were held during the software requirements and design phases. During the development phase, several beta releases of the ISO Data Archive were issued for testing. This interaction was an important factor for success as most of the look-and-feel issues and bugs were discovered and corrected before version 1.0 was released to the public. The overall project schedule from start to release of the ISO Data Archive version 1.0 is shown in Figure 2.

Database design

Early on, with the use of a CASE (Computer-Aided System Engineering) tool, a Logical Data Flow Model was set up to define the overall system followed by a Physical Data Flow Model for the architectural design. This was then developed into specific system components that could then be transformed into pieces of future software code.

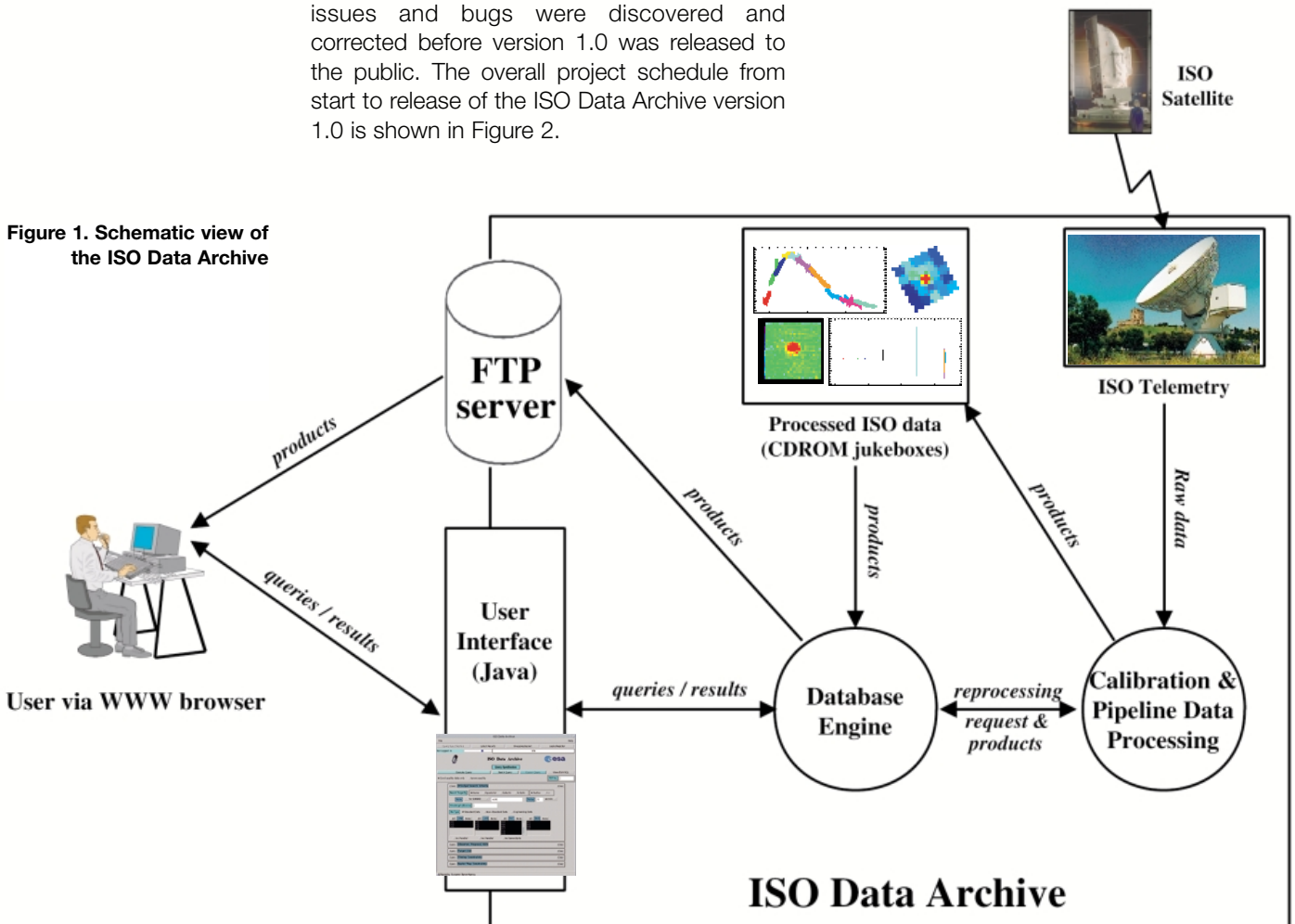
On the database side, a Conceptual Data Model was defined and extensively reviewed with the users. It was then transformed into a Physical Data Model to define a detailed database structure using a CASE tool that generated database tables and indexes from the model.

User interface

To design the user interface – once Java was adopted (see accompanying panel) – a ‘class tree’ was defined using the Rumbaugh Style Object Model.

One of the important new features of the ISO Data Archive is the provision of ‘browse products’ or ‘quick-look’ data associated with each observation. These products enable users to make informed decisions as to which

Figure 1. Schematic view of the ISO Data Archive



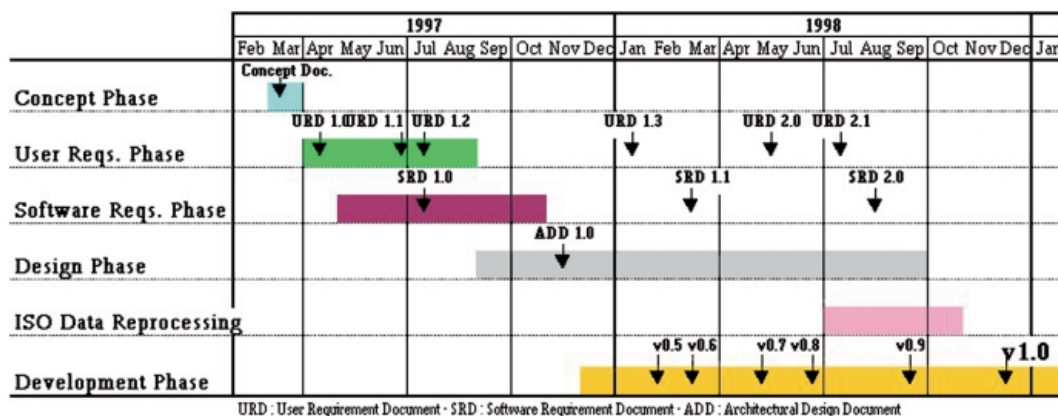


Figure 2. Overall ISO Data Archive project schedule

observations they want to download for detailed astronomical analysis. The quick-look products were developed as a joint effort between the ISO Data Centre, the Infrared Processing and Analysis Center (Pasadena, USA) and the French ISO Centre (Orsay, France). These products consist of:

- an icon (GIF file): a 56 x 56-112 pixel visual representation of the data content of the observation
- a postcard (GIF file): a screen-sized enlargement of the icon containing extra information about the observation (Figure 3 shows an example for each instrument)
- a survey product (FITS file): a fully processed product, which can be quickly downloaded by the user and manipulated using standard astronomical tools.

reprocessing of all data with the latest software version is foreseen towards the end of the archive phase, i.e. mid-late 2001. This reprocessing will create the final 'ISO legacy archive'. In the meantime, it is planned to have one major update per instrument per year of the calibration and processing software. To enable the astronomical community to take advantage of these improvements, an on-the-fly reprocessing facility for all data products will be provided through the interface to the ISO Archive.

Figure 3a. Postcard example of CAM observation

Data processing

An important design requirement was to automate data processing. During normal operations, an operator processed each day's data on two VAX/VMS machines, typically requiring about 24 hours. Up to three days' data could be processed simultaneously by three operators.

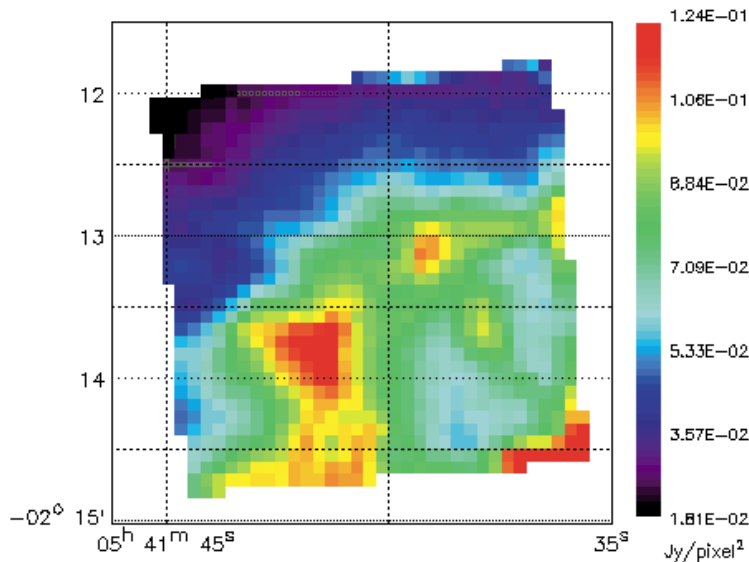
To solve this, all the telemetry (TDF) was put online on hard disk (about 350 GBytes) and parallel processing was performed over a cluster of six ALPHA/OpenVMS machines. These developments improved the processing rate to 20 days of data per day, more than six times the speed obtained during normal operations. Moreover, a single operator can now monitor it all.

The data processing and calibration software has been, and will continue to be, constantly updated and improved, as the behaviour of instruments is better understood. Thus, at the end of operations, in May 1998, all ISO observations (around 900 days of data) were bulk-reprocessed (BKRP) with the latest version of the pipeline software to produce the first uniform 'ISO interim archive'. A second, full

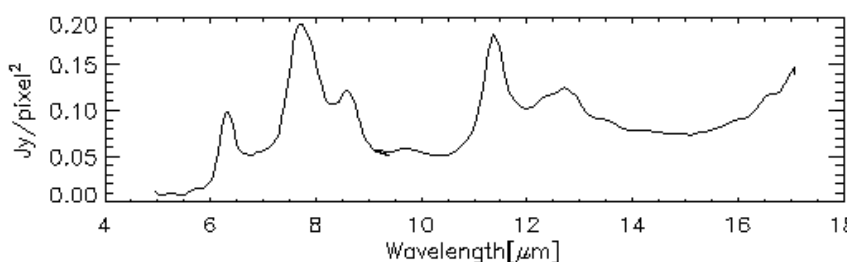
N2023-N 03-OCT-1997 CAM04

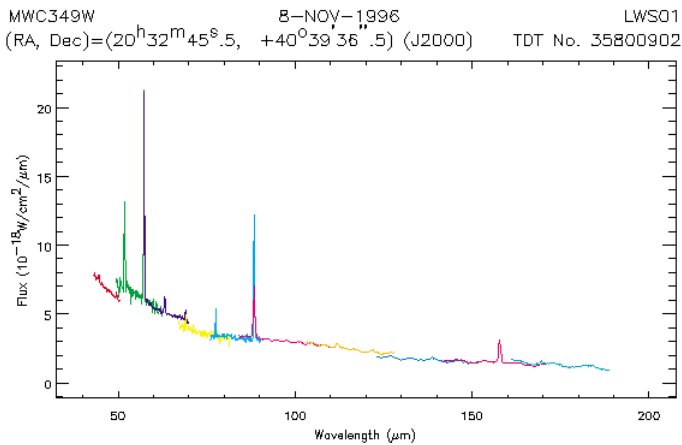
(RA,Dec)=(05h41m38.43s,-02°13' 4") (J2000) TDT No. 68701905

Mean CVF wavelength: 11.01 μm -- Pixel Field of View: 6.0 arcsec



OLP_701 - CALG_40 - CAM Browse V1.9





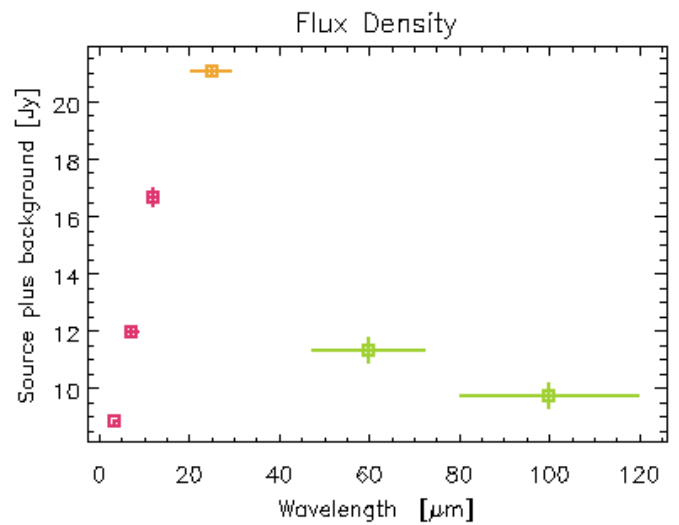
irsa Product v1.4 - OLP Version OLP_701 - CALG File Version CALG_40

Figure 3b. Postcard example of LWS observation

Figure 3c. Postcard example of PHT observation

Figure 3d. Postcard example of SWS observation

HD104237 2-MAY-1997 PHT03
 (RA,Dec)=(12h 0m 5.0s, -78°11'34.0") (J2000) TDT No. 53300120

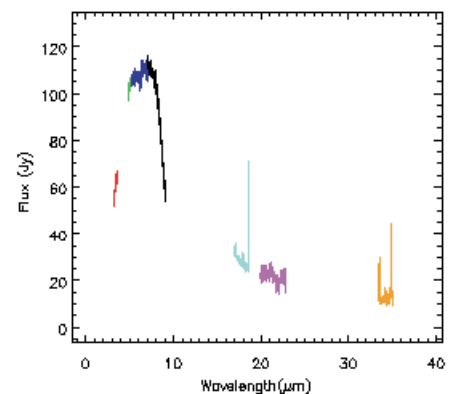
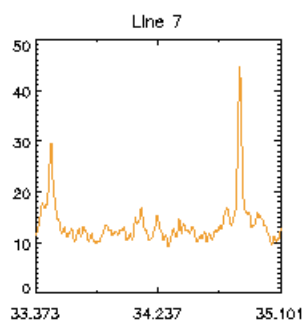
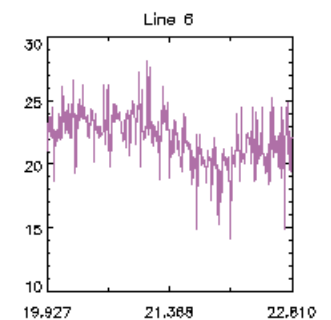
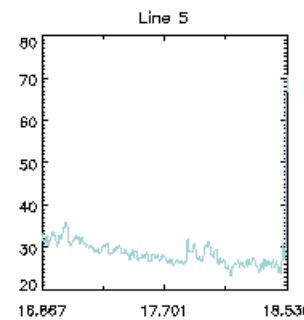
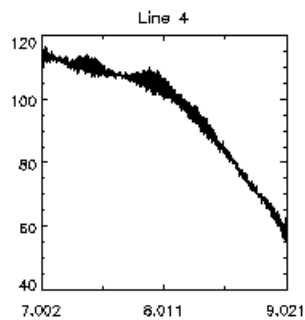
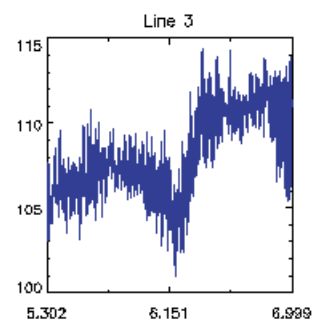
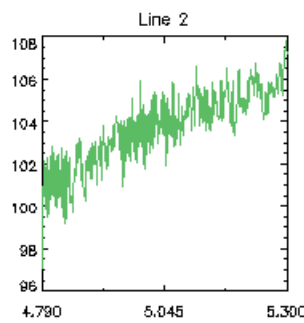
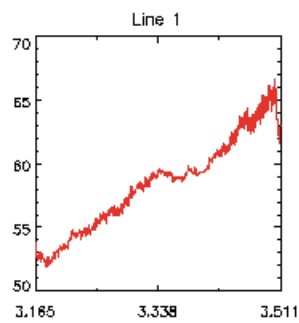


- CALG_40 - PHT Browse 1.4

IRAS 17080-3031
 (RA, DEC)=(17°41'12.8",

15-MAR-1997
 -30°32'29.0") (J2000)

SWS06
 TDT No. 48500603



Browse Product v1.3 - OLP Version OLP_701 - CALG Files version CALG_40

System design

The top-level system design is presented in a simplified form in Figure 4. This design integrated totally new parts (user interface, archive engine) with already existing systems (pipeline, update archive and CD creation) most of which needed some upgrade to fit into the new scheme.

The pipeline processing takes the telemetry from the TDF store and generates bulk-reprocessed products, which are ingested in the archive engine and copied onto CD-ROM jukeboxes. Via the WWW user interface, users can query the archive engine for products and can request those selected either from the prepared ISO data (stored product) or by triggering a processing request to obtain the latest version (processed product). In the case of the latter, the pipeline processing will generate the products on-the-fly and will return them to the archive engine. In both cases, the stored or processed products are put in the product distribution store and an e-mail is sent to inform the user that the requested products can be retrieved via FTP. If the products were requested on CD-ROM rather than via FTP, an operator will generate the CD-ROM, which is then mailed to the user.

This overall design has been very stable from the beginning and has managed to integrate all new user requirements to date.

The ISO Data Archive is connected to the Internet via a dedicated 2Mbits/second link to the European Academic Network called TEN-155, which ensures very high capacity connectivity between all European countries.

Currently, the ISO Data Archive consists of about 400 GBytes of data stored on around 600 CD-ROMs spread over two CD-ROM jukeboxes. The database (tables and indexes) volume is about 6 GBytes and resides on hard disk. The archive is connected to the Internet via a dedicated 2Mbits/second link.

ISO Data Archive usage

Prior to releasing the archive, all newly reprocessed data was distributed directly to its 'owners', i.e. the ISO observers who requested those observations. Additionally, all basic science products were exported to the instrument institutes (about 220 GBytes). These distributions were made so as to avoid an initial overload on the archive.

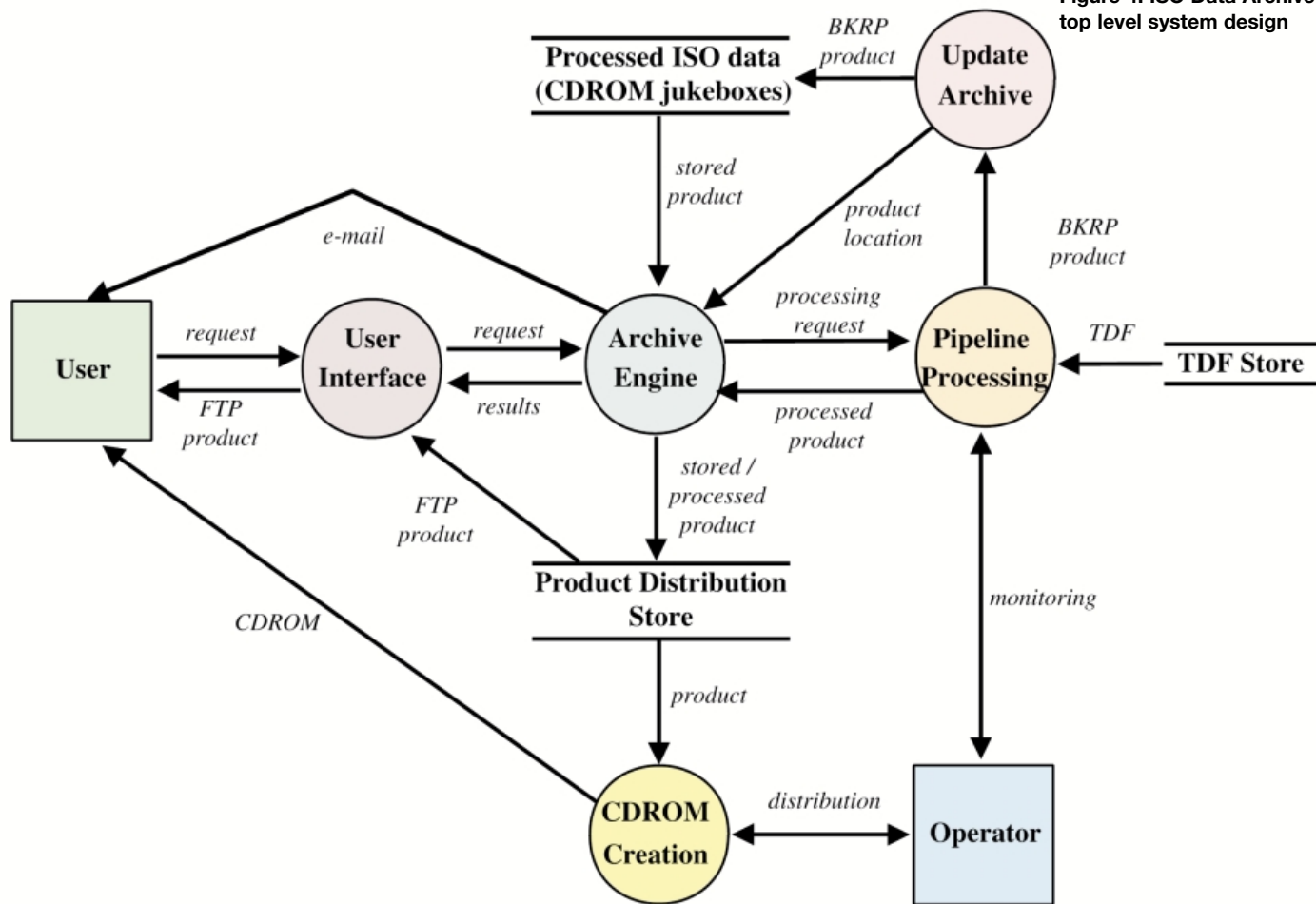


Figure 4. ISO Data Archive top level system design

The Choice of Java

At the start of the ISO Data Archive development phase, it was decided to adopt Java as the programming language for the user interface. At that time (1997), the choice was somewhat risky since Java was not always stable within the major WWW browsers. Nonetheless, its many other advantages made it the best choice.

Instead of using many standard WWW forms associated with cgi-bin scripts written in perl or C, Java allows integration of diverse facilities into a single application. Using the Java object-oriented approach, the interface is made of modular and 'living' components such as on-line help, subpanels and input fields according to other parameters. Control of displayed components on the screen is made easier and handling of quick-look data (e.g. icons, postcards) is more efficient. Moreover, this approach allows some level of user configurability.

Java also offers the possibility of doing various tasks in parallel, like consulting the help facility while querying the database.

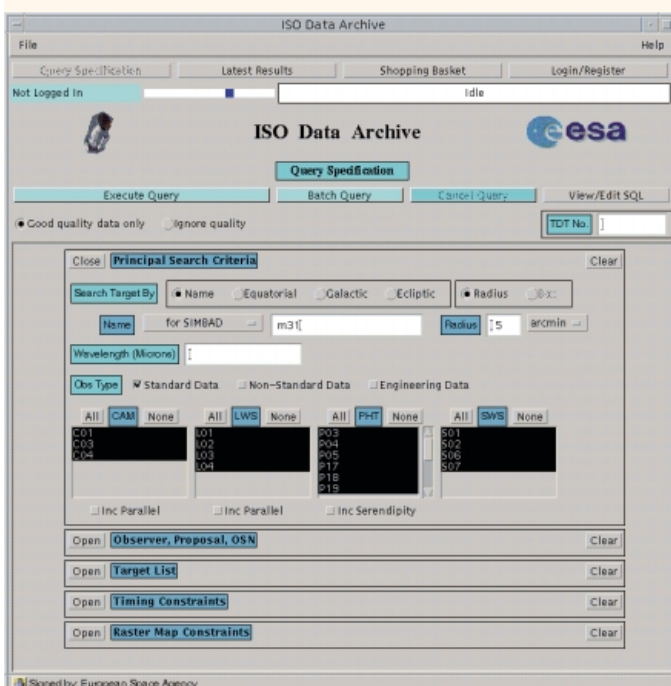
Once the ISO Data Archive Java applet (current size is around 500 kbytes) is downloaded to a user's computer, many operations are performed using the client computer's own resources (e.g. CPU, memory). This avoids unnecessary interactions with the ISO Data Archive server for faster access. Additionally, using a specific Java driver for the SYBASE database enhances access to the user interface.

Using the Java object-oriented approach and the Model View Controller (MVC) paradigm, the ISO Data Archive is modular and flexible, easier to maintain, and source code documentation is automatically generated. This approach is particularly appropriate for developing complex and portable graphical user interfaces viewable through WWW browsers. The effort required for developing and maintaining such an archive using WWW forms and cgi-bin scripts would probably have been much higher.

However, using a new technology was not problem-free. The first beta releases suffered from the then-poor implementation of Java in the most popular WWW browsers. To circumvent this, without extra effort, a standalone Java application of the ISO Data Archive was made available. This could be installed by the user and ran without a browser, thereby avoiding Java instability.

But with the ensuing beta releases of the ISO Data Archive and the improvement of the WWW browsers, the ISO Data Archive v1.0 issued in December 1998 was very stable and reliable, in both stand-alone and browser environments, and the choice of Java definitely paid off.

How to Use the ISO Data Archive



User registration and login

Though anyone can query the archive, a user name and password must first be requested and issued. The login mechanism is necessary to protect proprietary data, helps in monitoring archive usage and avoids possible disturbances from hackers. Additionally, an e-mail address is required for notification from the archive.

Making a query

Having reached the ISO Data Archive on the WWW <<http://www.iso.vilspa.esa.es/>> (Data Archive section, General user), and having started the Java applet, users can define a query as shown in Figure 5.

The *Query Panel* consists of various sub-panels, which can be opened or closed to specify query parameters.

The *Principal Search Criteria* panel, opened by default, allows the astronomer to query against general parameters such as:

- target name as given by the proposer or as known in the various astronomical archives (e.g. NED and SIMBAD)
- coordinates
- wavelength ranges
- type of observing mode (AOT) used by the satellite
- instrument

Figure 5. Making a query within the ISO Data Archive

Aimed more at expert users, the interface supports querying of non-standard modes, including engineering data and additional observations obtained in parallel with other instruments, or while slewing between targets. Queries can also be made against the 'quality' flag associated with each observation.

Other sub-panels are more specific and support searches against parameters like:

- observer or proposal name, proposal text, observation number
- date, time, uplink software version, revolution or phase in orbit
- target name list
- observation list
- raster map parameters.

For each of the parameters, help can be obtained by clicking the mouse button on the field.

Once the user has completed the search criteria, the query can be executed. If the query is too complex to complete quickly, the system will trap it, stop it, suggest that the user runs it in batch mode, and send an e-mail when the results are available.

Getting the results

Once a query has been executed, the ISO Data Archive returns the list of ISO observations matching the constraints in the *Latest Result* panel. A subset – sized to fit the screen – of the returned list of observations will be displayed. The user can scroll through this list using navigational buttons.

Figure 6 shows an example of the information displayed for each observation. An icon (one of the browse products) on the right-hand side of the screen gives an immediate visual impression of the data content of the observation. The postcard can be viewed by clicking on the icon. The amount of information and observations displayed can be customised; the main parameters are:

- target name
- observing mode
- coordinates
- observation type and number
- field-of-view
- wavelength range
- observer and proposal identification
- date, time and observation lengths
- quality of the data.

More detailed information, linked and associated files are offered for each observation.

Requesting the data

Another innovative and friendly feature of the ISO Data Archive is the wide selection of possibilities offered for the retrieval of data. Observations can be selected for retrieval by moving them to the 'shopping basket'. From this subpanel, the user can select the level of products to be retrieved for each observation individually, collectively, or a combination thereof, i.e.:

- Raw Data
- Basic Science (default)
- Fully Processed
- Quick-Look
- Custom
- All.

Once completed, the user presses the 'Submit Request' button. The user interface then determines the volume of data, checks whether the daily quota for the user has been exceeded and asks the user for a compression format. The download time by FTP (default means of data retrieval) can be estimated if required. The archive processes the data retrieval task and copies the data into an FTP (public or secure) area and stored for a period of seven days. The user is informed via e-mail.

Alternatively, the user can request the data on CD-ROM. In this case, an operator at the ISO Data Centre creates the CD-ROM and sends it by normal mail to the user. However, due to the excellent Internet connectivity of the ISO Data Archive, all data requests to date have been handled by FTP rather than the more expensive and slower CD-ROM route.

Observation ID	Instrument	Proposal	Date	Time	Quality
M31_B01LGE	SWS01	00h42m46.01s +41d16'11.9"	20-Dec-1996	21:09:09	1140 sec Good
M31_Center	PHT22	00h42m44.40s +41d16'08.5"	26-Jan-1997	15:14:54	2416 sec Good
M31	LMS01	00h42m44.40s +41d16'08.5"	18-Jun-1997	13:23:49	3160 sec Good
M31_PAN	PHT40	00h42m44.40s +41d16'08.5"	15-Dec-1997	22:30:57	6066 sec Good
M31	CAM04	00h42m44.30s +41d16'09.0"	20-Dec-1996	22:22:45	700 sec Good

Figure 6. Viewing the results from a query

When the archive was released, only some 6% of the standard observations were available. Nevertheless, the archive has been extensively and smoothly used by a significant number of users (see Table 1). Access will continue to increase in the future with the availability of all the data by August 1999, and as new versions of calibration and processing software with further improvements are released.

Table 1. Archive usage

Accesses	around 8000
Hosts	around 1000
Users	156
Data requests	around 2000
Data Volume	33 GB

Future developments

A phased approach is being taken for the overall development. Version 1 of the ISO Data Archive implements most of the original user requirements. Version 2 (mid-1999) is aimed at incorporating the specific needs of expert users. Version 3 (end-1999) will fully implement all user requirements. Among other features, it is planned to:

- add more data to serve the needs of instrument experts
- add extra subpanels to the user interface to access these new data
- implement the ISO products on-the-fly reprocessing
- integrate links to and from other astronomical archives
- provide a direct interface to the ISO Data Archive for other astronomical archives and programs
- add interactive browse product display tools
- implement a 'user-defined function': essentially scripts that can be run on the data products themselves; these are being developed for expert users by the ISO Spectrometer Data Centre (Garching, Germany)
- consider putting all ISO products on hard disk rather than on CD-ROM jukeboxes.

Conclusions

The ISO Data Archive, designed and developed by the ESA ISO Data Centre in Villafranca Spain, has been available to the external world since 9 December 1998. It provides fast and easy access to ISO observations data via a modern, innovative, efficient and effective user interface based on Java technology.

After a smooth design and development phase with constant collaboration with science users, several beta versions were tested by end-users

in order to release a stable and high-quality archive.

By the end of 2001, all ISO products will have been reprocessed with the latest calibration software and final improvements will have been implemented to leave the legacy ISO Data Archive to the world astronomical community.

Acknowledgements

We would like to thank the software team who designed and developed the ISO Data Archive: Richard Carr, John Dowson, Antonio de la Fuente, José Hernández, Neil Jenkins, Arno Plug, Andy Pollock, Pedro Osuna, Richard Saxton, Alex Scohier and Colin Todd. We owe special thanks to Deborah Levine for all her work on the user requirements. Kieron Leech is thanked for the user test team.

We also thank all others, too numerous to mention by name here, who participated in the ISO Data Archive project and contributed to its successful first release.

The ISO Data Archive is one element of the ISO Post Operations phase, which is a collaborative effort coordinated by the ISO Data Centre in ESA, Villafranca, Spain and includes six national data centres: the French ISO Centres (Orsay-Saclay, France), the UK ISO Data Centre (Rutherford, UK), the ISO Spectrometer Data Centre (Garching, Germany), the ISOPHOT Data Centre (Heidelberg, Germany), the Dutch ISO Data Analysis Centre (Groningen, the Netherlands) and the Infrared Processing and Analysis Center (Pasadena, USA). 