INES: Astronomy Data Distribution for the Future

W. Wamsteker, I. Skillen, J. D. Ponz, A. de la Fuente, M. Barylak[‡], I. Yurrita[§]
ESA IUE Observatory, Villafranca, P.O. Box 50727, 28080 Madrid, Spain

July 1999

Abstract.
We describe here the structure under which the IUE Project will leave its archive at the completion of the distribution system and final data processing, within the INES system. The INES system is a total system, which comprises both the data in their final bulk processing mode for direct application to scientific analysis, as well as the software driving the distributed service for data retrieval directly by the end user. As a consequence of the expected long-term usage and support needs, it has been designed to require minimum maintenance costs and will not suffer single point failures because of the distributed nature. Integration of the INES system in a more general multi-wavelength archive for specialized analysis is anticipated to be relatively easy, so further evolution of the data availability will be driven fully by the users community and will evolve as the data needs of the community develop. The long term responsibility will be transferred from ESA to the astronomical community through the establishment of the INES Principal Centre at the LAEFF Institute of INTA in Spain.

Keywords: archive; IUE

1. Introduction

The INES (IUE Newly Extracted Spectra) System is a complete astronomical archive and data distribution system. Its production represents the final activity of ESA in the context of the IUE Project (International Ultraviolet Explorer) in its user services. INES contains, at the 98% level, the complete data set of 104,000 spectra obtained in 18.5 years of ultraviolet spectroscopy from space with the IUE satellite. As the usage of IUE data is expected to remain high for some time to come and the IUE Archive represents a unique historical record, as well as a very useful current research and educational tool, special effort was made to:

1. Assure ease of access and use of the data.
2. Bring the archive distribution system up in a currently up-to-date configuration, suitable to the tools available at most research institutes, universities and schools world-wide.
3. Limit the need for specialized Project knowledge or experience to use the data.
4. Structure the data such that further development can be easily implemented on top of the INES system without affecting its native functionality.
5. Prepare the system such, that the data can be used with ease by all astronomers and educators.
6. All data in the archive must be inter-comparable and can be used with the standard data analysis and statistical tools in use by the community.
7. Minimize maintenance costs associated with the archival support.

[‡] current address: ESOC, Robert-Bosch-Str. 5, D-64293 Darmstadt, Germany.
[§] current address: CDTI, P. de la Castellana 141, E-28046 Madrid, Spain.

As a result of these guidelines the INES system was developed under the WWW structure on a Linux platform, with essentially only public domain tools to support the basic archival functions. As a consequence of the extraordinary success of ULDA, the previous IUE Archival system (Wamsteker *et al.*, 1989), which was also based on a distributed archival structure (25 National hosts supported the retrieval of more than 235,000 IUE spectra between 1987 and 1995; see Table 1), the same philosophy underlies the INES system.

The INES system represents the final step in the astronomical data archival cycle, a Historical Reference Archive (Wamsteker and Griffin, 1995).

2. The Data

The INES system contains all the data in a reduced form, i.e. directly applicable to astrophysical analysis. The INES data set has been produced after science verification showed that the SWET extraction applied in NEWSIPS resulted in, at times, unreliable data, making a direct comparison of all data between each other not practical for users with no experience in the operational project. Also, the error structure of the data would not allow the use of the data with statistical tools. This is highly undesirable for long-term, limited support usage of the data. The details of the many improvements applied to the data contained in INES are described by Rodríguez *et al.* (1999), Cassatella *et al.* (1999), and González-Riestra *et al.* (1999). Direct examples of the application of INES data are Huélamo *et al.* (1999) and Tuerler *et al.* (1999).

3. The Distribution System

Apart from its improved data content, INES is a self-contained, internally consistent archival data distribution system. It is adjusted to adequately advanced communications technology, and is structured to be easily integrated in a logical way into more specialized archive systems, e.g. for multi-wavelength archival studies. The concept is based on a distributed archive system with National Hosts in different countries, and a Principal Centre from where the system is centrally maintained and organized. The larger data sets which can not be easily stored at the National Hosts are directly accessible to the user from the Principal Centre. The whole system is self-serving, so each end user can complete a total archival retrieval from his own institute, from data search to actual retrieval of the spectra (see Figures 1 and 2). The system is based on the Linux operating system, is PC driven, and does not require proprietary software. Data are delivered in FITS format to the user. The Principal Centre is located at the LAEFF (Laboratorio de Astrofísica Espacial y Física Fundamental) of INTA in Spain, which is intended to maintain and further develop the system for the world wide community of scientists. A mirror site for the Americas is maintained at the CADC (Canadian Astronomical Data Centre) in Victoria. At this time some 20-30 National Host installations are operational, one in each country. Information (such as the URL's of the National Hosts) can be obtained easily at the Principal Centre at the WWW Address:

`http://ines.vilspa.esa.es`

As the INES system is the logical modernization of the first self driven data retrieval system in Astronomy, the IUE ULDA, it is interesting to see how these data systems

Table I. National Host De-Archival from IUE/ULDA
Registered Users = 1,190 Scientists (18% IAU membership)
Total de-archiving supported (1988 until 1998): 235,956 (3.0 spectra/hr)

Country	IAU members	ULDA users	ULDA Use	
Austria	29	3	n.a.	
Belgium	81	21	2,140	
France	560	85	16,589	
Germany	421	116	20,317	
Italy	372	168	17,379	
Netherlands	159	57	5,269	ESA Member States
Nordic countries	181	63	5,077	(117,175 = 50%)
Spain	165	115	10,441	Total ESA Users: 850
Switzerland	56	34	14,349	
United Kingdom	477	181	25,492	
Portugal	15	7	122	
Percentage Archive users: 34%			Rate: 137 spectra/astronomer	
Brazil	87	39	18,827	
China	290	25	18,603	
Costa Rica	—-	4	174	
India	208	6	2,159	Developing Countries
Mexico	55	6	247	(47,404 = 20%)
Philippines	—	1	n.a.	Total D.C. Users: 131
Taiwan	19	15	1,795	
Turkey	47	2	786	
South Korea	21	2	113	
via VILSPA[1]		31	5,000	
Percentage Archive users: 13%			Rate: 360 spectra/astronomer	
Canada	207	42	5,929	
Czech	17	3	n.a.	
FSU	424	6	n.a.	
Israel	45	20	2,185	
Japan	349	6	219	
Poland	94	8	n.a	Other Countries
Slowak	18	3	n.a.	(71,077 = 30 %)
USA	2,059	121	2,694	Total other Users: 189
VILSPA[1]			52,006	
ST-ECF/ESO[2]			8,044	
Chile			43	

[1] Primary Host for IUE/ULDA Project. Considerable support for both countries without National Host (appr. 5000 spectra) and direct Project usage and support.
[2] Local usage only.

are contributing to the usage of archival data on a world wide basis. These results are shown in Table 1. This demonstrates that the IUE Project, through its observations and Archive, has not only had major impact in Astrophysical research, but also the archival use has contributed in a major way to the familiarization of scientists all over the world with space science, i.c. Astrophysics. The INES system fulfills the basic functions for fast convenient access, allowing the end user to be able to do archival astrophysics in an efficient way:

1. Finding the available data (search).
2. Previewing (evaluation of data quality).
3. Data selection (choosing).
4. Data storage by the end user (data retrieval).

All these functions are implemented in such a way that, even with no pre-knowledge of the IUE Project, data analysis can be efficiently done. The overall system of INES has been designed to be essentially platform independent for the end-user, requiring only standard Internet browsers to support data access. The data transfer functionality has been organized to minimize network traffic to assure that, also users with limited end-bandwidth, or heavy network traffic conditions, will not seriously hamper the access to the INES system and data. Already the availability of a prototype in a small number of national Hosts has shown that the INES usage is well received by the community. With the installation of the INES version 2.0 in all National Hosts the deployment of the overall operational INES system is foreseen for December 1999.

4. Summary

We have illustrated the capabilities and structure of the IUE INES system developed at the ESA IUE Observatory to support the long term availability of the 104,000 ultraviolet spectra in the IUE Archive. The long term responsibility of the distributed INES Archival system will reside in the LAEFF Laboratory at VILSPA. As has been shown, a good choice of Archive Architecture greatly enhances the usefulness of archival data in the wide community of astronomers. It can be expected therefore that the INES system will have an even stronger influence on the development of capabilities in the Basic Space Sciences for the developing countries than it predecessor, the IUE/ULDA, which has been an example of advanced astronomical data distribution and support during the last decade of this century.

References

Cassatella, A., *et al.*, 1999, Astron. Astrophys. Suppl. Ser., submitted.
González-Riestra, R., *et al.*, 1999, Astron. Astrophys. Suppl. Ser., submitted.
Huélamo, N. , *et al.*, 1999, Mon. Not. R. A. S., submitted.
Rodríguez-Pascual, P.M. *et al.*, 1999, Astron. Astrophys. Suppl. Ser., in press.
Tuerler, M., et al., 1999, Astron. Astrophys. Suppl. Ser., 134, 89.
Wamsteker, W., et al., 1989, Astron. Astrophys. Suppl. Ser., 79, 1.
Wamsteker, W., Griffin, R.E.M., 1995, Astrophys. Sp. Sci., 228, 385.

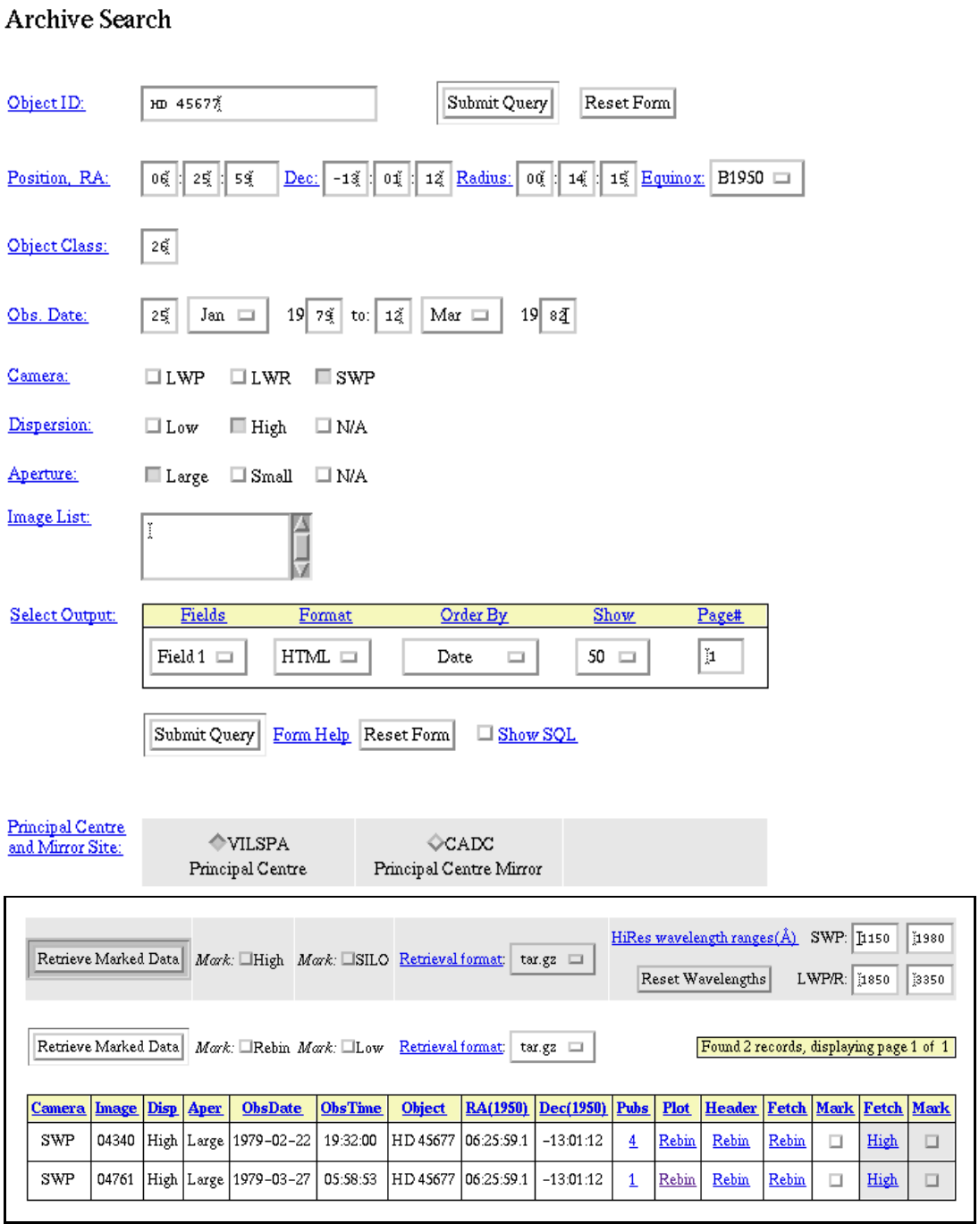
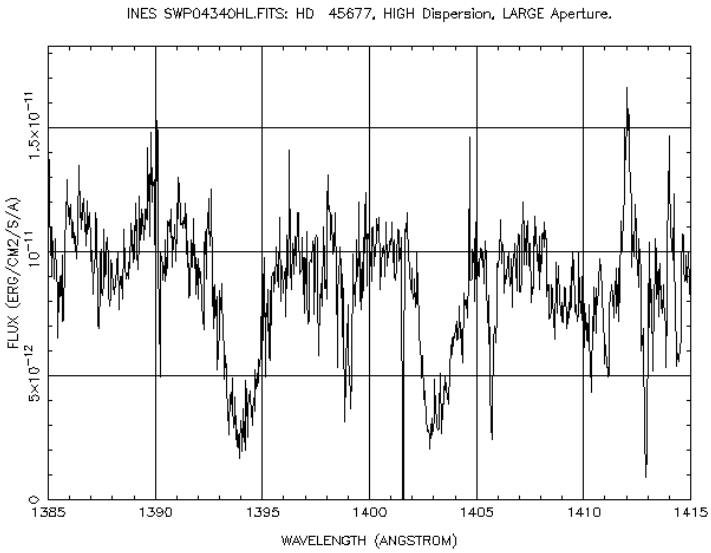


Figure 1. This figure shows the flow of the search, selection and data transfer capabilities of the INES system. They are driven from the box in the bottom of the figure which shows the results of the search shown in the upper part. The different source of the data supplied is indicated by the grey areas (Principal Centre) and the white fields (National Host). These differences are completely transparent for the end user. To avoid the transfer of large amounts of unnecessary data for the high dispersion spectra, a smaller wavelength range, containing only information of interest, can be selected.



SWP04340RL.FITS Browse Plot

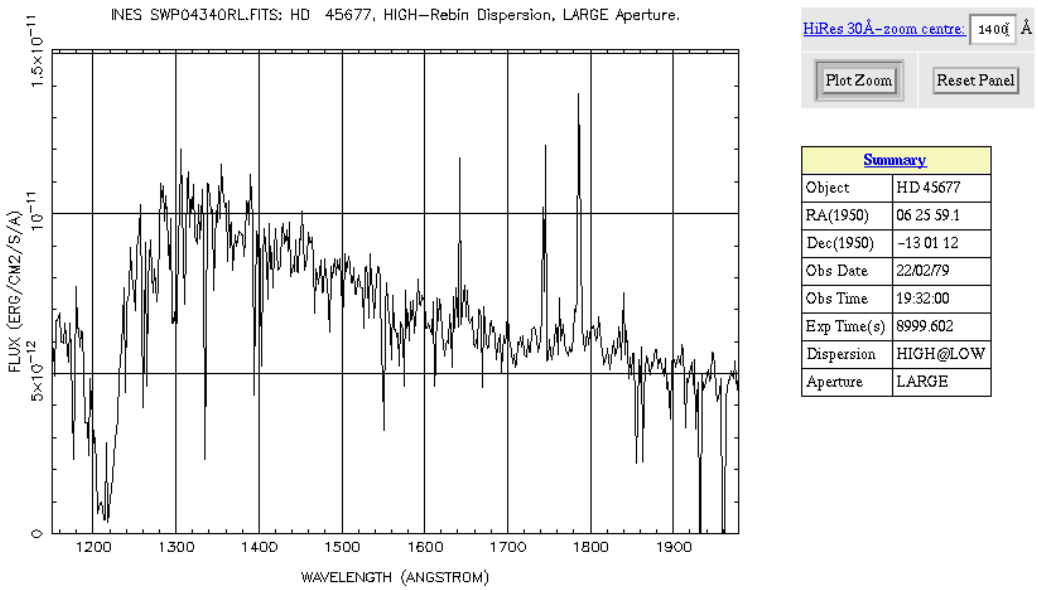


Figure 2. This figure shows the data previewing capabilities supplied within the INES system. These are filled respectively, from the National Host (*Browse plot* in the lower part), and the Principal Center (high dispersion shown in the upper part). The *Browse plot* shows the rebinned high resolution concatenated spectrum at low resolution binning (not convolved to low resolution). The rebinned spectrum can also be directly compared with real low resolution data. The active box *HiRes 30 Å zoom center* allows the choice of the part of the high resolution data which will be shown. The concatenated high resolution data reside in the Principal Centre. The distributed nature of the archive assures that the system response will remain good for the end user.