THE ULTRAVIOLET DIGITAL SKY
(the GALEX Science Data Archive)

LUCIANA BIANCHI[1,2], AND THE GALEX TEAM [3]
[1] *JHU Center for Astrophysical Sciences, Baltimore, USA*
[2] *Astronomical Observatory of Torino, Italy*
[3] *CIT, JHU, LAS, CAL, UC Berkeley*

ABSTRACT.
GALEX (The Galaxy Evolution Explorer), approved as a NASA SMEX mission to fly in 2001, will perform imaging and spectroscopic surveys of the sky in the Ultraviolet (see Bianchi et al 1998, 1999a,c, Martin et al 1999). The GALEX mission will produce an unprecedented all-sky map in two UV bands (deeper than the POSS I and essentially without precursors in the UV), a nearly background-limited unbiased catalog of UV sources, UV images and spectra of nearby and distant galaxies, and a UV map of the background. The catalog of UV sources will contain an order of magnitude more objects than POSS, IRAS, EUVE, and Rosat provided for the infrared, extreme ultraviolet, and X-ray regimes. GALEX will also provide UV spectra of some 10^5 objects (to magnitude 23-25 in the deep fields). The GALEX archive will remain a long lasting powerful resource for a variety of research projects, and a road map for future space missions and ground based telescopes.

GALEX will produce a data rate several times higher than HST, and $\approx$ 3 TB of data in its 2.5 years planned lifetime. Other TB-size surveys at different wavelengths will be available after the GALEX mission, such as the SDSS (Sloan Digital Sky Survey) with imaging in 5 optical bands and optical spectra. Complex, multiparameter queries of terabyte-size archive are not possible with traditional indexing techniques. A new archive architecture (multi-dimensional tree-structure organization of the data that will allow a course-grain map of indexing parameters) is described that will allow complex queries to be performed on large databases, and to link databases at different wavelengths. Ultimately it will become possible to browse a multiwavelength digital sky.

1. The GALEX Science Data Products

Data are stored in the archive at three levels: 1) operational archive (raw data); 2) team only accessible data (for pipeline iterations, time analysis, recalibration); 3) science archive. We describe here the science archive. The GALEX science data products will be:

- The image of the whole sky in the two UV bands (SkyMap),
- the All-sky Imaging Catalog, with positions, UV magnitudes (two GALEX bands), cross correlation with known objects, source profile and postage stamp images. The catalog will contain: $\approx 10^6$ QSOs, $\approx 10^7$ spiral/irregular galaxies, $\approx 10^4$ elliptical, $\approx 10^3$ galaxy clusters, $\approx$ 3 x 10^5 hot white dwarfs, $\approx 10^5$ horizontal branch and sD stars, and $\approx 10^4$ cataclismic variable stars.

- UV Images of resolved nearby galaxies
- the Spectral Catalog, containing some 10^5 UV spectra, with reference data for the objects from the All-sky Imaging Catalog
- Subcatalogs of objects classified by astrophysical classes with cross-reference identifications to known objects
- a Variability Catalog obtained by re-examining all the photon-counting data for each point source catalogued after the reconstruction of the sky image. The *variability survey* in the Deep Fields will contain ≈10-300 photometric measurements of every object spaced over several months.

2. Powerful Browsing of TB-size Databases

GALEX will produce a data rate of 28 Gbit/day, higher than HST (currently 3-5 Gbit/day), and a final ≈ 2.5 TB size archive containing tens of millions of objects, and several hundred thousands UV spectra. For comparison, IUE produced 100,000 spectra of 10,000 objects in 18.5 yrs, HST has produced about 6 TB of data to date (of $\approx 10^4$ objects in $\approx 10^5$ datasets). The GALEX science archive will be larger than previous and existing surveys, however other projects planned to be operating at the same time as GALEX will have similar data rates and type. For example the Sloan Digital Sky Survey (SDSS) will have data rates even larger than GALEX.

Innovative approaches in the GALEX data storage architecture are incorporated in the design which 1) will make complex browsing of large data volumes quick and 2) will present the users more flexible on-line information than current table-format-only archives (e.g. IUE, HST), so that the search outputs can be easily visualized and preliminarily examined. In order to provide a UV sky survey as a "fundamental, long lasting" resource, which can support future investigations for decades, we plan an archive that can easily be transported to other platforms and expanded by linking and cross-correlating to similar databases at other wavelengths in a transparent way.

A hierarchical structure of the data will allow the users to browse quickly, and with a varied combination of query constraints, large data volumes. In some cases the query outputs will also be very large, and statistical in nature (rather than requests for individual datasets like e.g. HST) so the archive must support some visualization of the results for fine-tuning of the query parameters.

The catalog data are inherently multidimensional (each object has position, fluxes, size, etc.), and searching for specific classes of objects requires the possibility to define complex domains in the N-dimensional space, including spatial relations. A fast, indexed, but complex search of Terabytes in N-dimensional space is not possible with traditional indexing techniques. Thus we will create a structure where data are quantized into "containers" (each container including objects from a same spatial region, and with similar parameters). The containers represent a coarse grained map of the data, enabling us to cut down the response time to a complex query by a very large factor. The storage hierarchy adopts a sky tassellation (AstroMap: Szalay and Brunner 1998; Szalay 1998) which allows spatial indexing of the objects. AstroMap is also adopted by other projects,

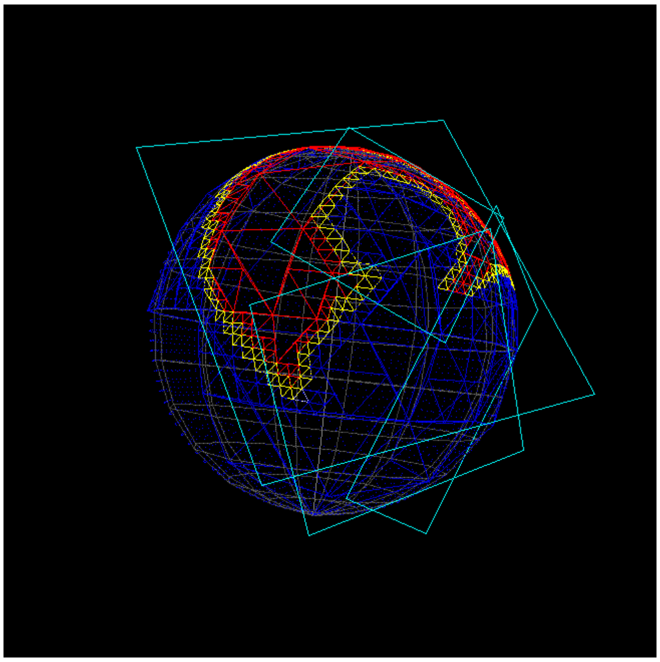

Fig. 1. Representation of a tridimensional query performed over the hierarchical map of data. The sky (spatial coordinates, but also any other combination of parameters) is subdivided into triangular grids from coarser to finer. Grey triangles represent sets of data whose parameters satisfy the query definition, white triangles contain sets of data that might or might not satisfy the query: only these subsets need to be scanned. For more details, see Bianchi et al 1999b.

as a proposed community effort to use a common partitioning of the celestial sphere (e.g. SDSS, Szalay 1998; GSC-2, Greene et al. 1998).

3. The Multiwavelength Digital Sky

In a few years, it will be easy to handle and browse large volumes of data. Powerful statistical searches for objects with given combinations of [multiwavelength] parameters over large areas of the sky will be possible, allowing to exploit effectively all the potential for discoveries and unprecedented information contained in the UV sky. The availability of multiwavelength, terabyte size, archives from GALEX and other survey projects will open a new way of doing science, with entirely archive-centered research projects.

The GALEX archive structure, compatible with the SDSS (Szalay 1998) archive, that should be completed approximately at the same time, is being adopted by other projects as well (FIRST, 2MASS, GSC-2, POSS2). This will ultimately allow to link archives from different projects. The multiwavelength archive, which will evolve in the conceptual "Digital Sky", will link GALEX to other archives, supported separately but that will share a common structure and seamless cross-referencing scheme (see also Szalay 1998, Bianchi et al. 1999b).

References

Bianchi, L., Martin, C., et al.: 1998, in *UV Astrophysics beyond the IUE Final Archive*, R. Gonzalez-Riestra, W.Wamsteker and R.A. Harris eds., ESA SP-413, p. 797.

Bianchi, L., et al.: 1999a, *Mem. Soc. Astron. It.*, Vol. 70, N.2, in press

Bianchi, L., Szalay, A., et al.: 1999b, in *Ultraviolet-Optical Space Astronomy beyond HST*, J.M. Shull, A.L. Kinney and J.A. Morse eds., ASP Conference Series, Vol. 164, p. 194.

Bianchi, L., et al.: 1999c, *Mem. Soc. Astron. It.*, this book

Greene,G., McLean,B., Lasker, B., Wolfe, D., Morbidelli, R., Volpicelli, A.: 1998, IAU Symp. 179, 474.

Martin, C., Bianchi, L., et al. 1999 in *Ultraviolet-Optical Space Astronomy beyond HST*, J.M. Shull, A.L. Kinney and J.A. Morse eds., ASP Conference Series, Vol. 164, p.182.

Szalay, A.: 1998, AAS 192, 640.

Szalay, A., Brunner, R.: 1998, IAU Symp 179, 455.