

# THE VIRTUAL OBSERVATORIES: A MAJOR NEW FACILITY FOR X-RAY ASTRONOMY

N. A. Walton<sup>1</sup>, E. Gonzalez-Solares<sup>2</sup>, A. M. S. Richards<sup>3</sup>, and J. A. Tedds<sup>4</sup>

<sup>1</sup>Institute for Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

<sup>2</sup>Institute for Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

<sup>3</sup>The University of Manchester, Jodrell Bank Observatory, Macclesfield, Cheshire SK11 9DL, UK

<sup>4</sup>Department of Physics & Astronomy, University of Leicester, University Road, Leicester, LE1 7RH, UK

## ABSTRACT

We describe how the Virtual Observatory (VO) projects in Europe, the USA, Japan, and elsewhere are meeting the challenge of providing simple and efficient access to data, from the world's observational facilities, analysis tools and computational resources.

We note the pan European Euro-VO project, and its technological development arm - the VOTECH project (see <http://www.eurovotech.org>), with a focus on how it is designing the framework for mass scale access to not only major X-ray facilities now such as Chandra and XMM-Newton, but also to the large data from new facilities such as VISTA, ELTs, ALMA.

Science drivers from the X-ray research community are helping shape the development of the VO, with their requirements on access to both X-ray and supporting large multi-wavelength data sets, often coupled with comparison to theoretical simulations (e.g. in the study of galaxy clusters). We note relevant capabilities currently available, including science services from the UK AstroGrid project Future services, such as those tuned to generate 'observational' xray maps of gas in clusters, on demand from a range of input simulations will be outlined.

We give a specific usage demonstration of the AstroGrid system in mining X-ray and optical/IR data in the search for high redshift galaxies. We note how major missions and research consortia can be the first to benefit from the use of newly emerging VO systems.

Key words: Virtual Observatories, Distributed computing, X-ray astronomy.

## 1. INTRODUCTION

Astronomy is currently a golden age, with an explosion in the rate of new discoveries. These have been fueled by

the availability of high quality observations of the cosmos, from a range of major new facilities which together have opened up the sky at all wavelengths, from radio, through gamma ray. Key examples include the Hubble Space Telescope, the European Southern Observatory's (ESO) VLTs, the Keck Telescope, the VLA, WMAP, and recently Spitzer and Swift. The view of the X-ray Universe has been revolutionised with the advent of the Chandra and XMM-Newton telescopes.

Supporting these facilities, has been the systematic surveying of the whole sky at a number of wavelengths, e.g. 2MASS, the Sloan Digital Sky Survey (SDSS), The NRAO VLA Sky Survey. Additionally our understanding of the observed Universe is aided by comparison with increasingly sophisticated and physically accurate simulations (e.g. the Hubble Volume Simulations — <http://www.mpa-garching.mpg.de/Virgo/>).

Taken together, these data resources have enabled major new scientific break throughs across a wide range of areas from study of the Universe at the largest scales (e.g. analysis of the CMB, supernovae, and galaxy clustering to determine the cosmological parameters).

However, the observational facilities that we now possess, and plan to construct (e.g. ALMA, ELT's, XEUS) present astronomy with a range of challenges. There is more data (at the peta scale), more complex data (objects can routinely described by many hundreds of parameters), and data from more multiple sources. Further, there is increasing pressure to ensure maximum access to data from these high value facilities, and to allow maximum re-use of their data, to allow the extraction of the maximum amount of science.

This is the challenge which the Virtual Observatory (VO) initiatives are responding to. Thus, AstroGrid (UK), NVO (USA), Euro-VO (Europe) and others are aiming to provide simple and efficient access to reduced data products from the world's observational facilities (and indeed major theoretical simulations), through standardised interfaces, and with the provision of a wide range of pow-

erful tools and algorithms to enable the analysis of the data. This will facilitate and speed the creation of new science by allowing simpler and more powerful access to data and applications.

Here we briefly describe the specific example of AstroGrid and the Euro-VO, and how already they are providing services and capabilities of relevance to the X-ray astronomy community.

## 2. EXEMPLAR SCIENCE FOR THE VO

As noted in the introduction, science drivers are shaping the development processes of the VO initiatives. Many of the science cases developed by science teams advising the VO projects include aspects requiring access to a variety of high-energy (X-ray, Gamma-ray) data resources. An example is an AstroGrid case, where the integration of data and tools to allow for the rapid follow up of SWIFT (Gehrels et al., 2004) based gamma ray alerts. Fig. 1 shows diagrammatically the data flows and processes involved, where data includes that captured rapidly, or that taken over longer time intervals monitoring fading sources.

## 3. BUILDING THE VIRTUAL OBSERVATORY

The creation of the worldwide global observatory framework involves the formation of an international partnership to agree vital interoperability standards. VO projects then build implementations adhering to these standards, taking advantage of newly emerging distributed computing technologies and fast networks.

### 3.1. The International Virtual Observatory Alliance

As described in subsequent sections, national and regional VO implementations, ensure that applications and services are developed which meet the prioritised science demands of their local communities.

The major VO projects recognised, however, that, at the low level, the various implementations would have to inter-operate. Thus a data resource published through the UK's AstroGrid should be accessible to the Euro-VO, the NVO and so forth. And likewise, resources and applications published in the the USA, Europe, should be discoverable and accessible to a user working through AstroGrid. This spurred the formation of the **International Virtual Observatory Alliance**<sup>1</sup> (IVOA).

Scientific and technical representatives from the global VO projects are thus agreeing these interoperability standards — covering areas such as 'registries' (the index

of resources), content descriptors (the astronomy dictionary to describe data), data models (providing described mechanisms to access astronomical data), data access, VO query language (one query language to enable querying of heterogeneous databases, registries, files, etc), and additional areas such as authorisation protocols. By adhering to the agreed IVOA interoperability standards in creating their VO systems, the user of one - such as AstroGrid - gains full access to the resources of the others.

An analysis of a standard VO architecture has been undertaken by the IVOA (Williams et al., 2004), giving an idea of the technical complexity underpinning any VO system. The various IVOA working groups maintain details of their activities on the IVOA wiki pages linked from <http://www.ivoa.net/twiki/bin/view/IVOA/WebHome>

### 3.2. The European Virtual Observatory

A number of partners (ESO, CDS-Strasbourg, AstroGrid, Jodrell Bank Observatory, and Terapix-Paris) in Europe formed the successful Astrophysical Virtual Observatory Project (AVO) see <http://www.euro-vo.org>. This three year (2001-2004) **EU Fifth Framework study programme**<sup>2</sup> produced the outline of the design for a Virtual Observatory in Europe. A key success of the AVO was the production of the AVO Science Reference Mission (SRM) document, which lays out a set of exemplar science cases which will benefit from, and in many cases require use of, VO capabilities, and thus impact on the design of the VO. The AVO SRM was produced by the AVO Science Working Group and is available online at [www.euro-vo.org/internal/Avo/AvoSRM/srm.pdf](http://www.euro-vo.org/internal/Avo/AvoSRM/srm.pdf)

With the completion of the AVO project, a more ambitious European Virtual Observatory (<http://www.euro-vo.org> Euro-VO) initiative is underway. This is composed of a range of European consortia representing major astronomical resource providers: ESO, ESA, AstroGrid (UK), NOVA (NL), INAF (Italy), CNRS (France), LAEFF (Spain) and GAVO (Germany). It is anticipated that it will be composed of three inter operating distributed structures:

1. **Euro-VO Facility Centre:** charged with providing user support and ensuring provision of registry services
2. **Euro-VO Data Centre Alliance:** providing a focus for the interface of major resource providers to the Euro-VO
3. **Euro-VO Technology Centre:** developing the infrastructure and technical components creating the Euro-VO. This final element is now funded.

### 3.3. The Euro-VO: Technology Centre

The Euro-VO Technology Centre (VOTC) is currently in progress through its VO-TECH project, a EUR 6.6M ini-

<sup>1</sup><http://www.ivoa.net>

<sup>2</sup><http://www.cordis.lu/improving/infrastructure/home.htm>

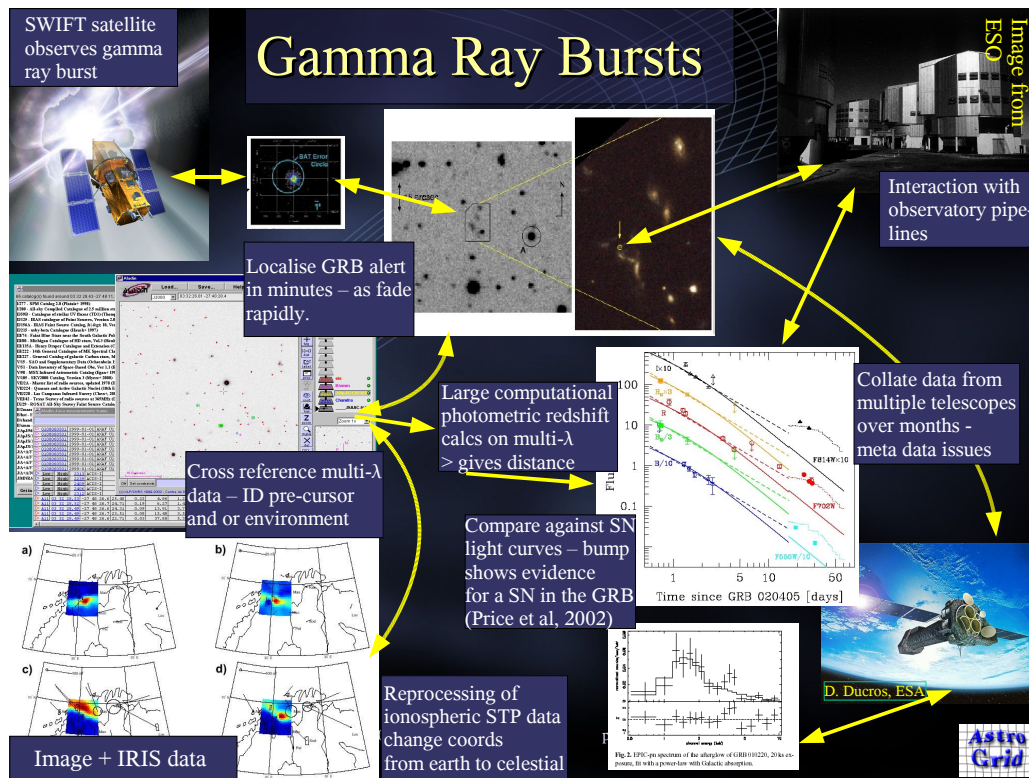


Figure 1. Complex work flows involved in following up Gamma Ray Burst alerts, ranging from the capture of rapid near real time earth based followup observations, through longer term monitoring of fading sources

tiative running from Jan 2005 to Dec 2007, which is carrying out feasibility studies and design work aimed at integrating new VO and distributed computing technologies into the Euro-VO. This work is closely aligned to ensure compatibility with, and input into, emerging global astronomical interoperability standards (e.g. from the International Virtual Observatory Alliance)

The VOTECH work is organised into three main strands:

1. Infrastructure: the VO middleware, e.g. Workflows, job execution, security, transport layer etc
2. New Tools: applications for the VO, e.g. Footprint, best fitting, SED builder, etc
3. Resource Discovery: finding the needle in the haystack, e.g. Building ontology's, dictionaries, resource browsers, etc
4. Data Mining and Visualisation: mass scale analysis, Large scale compute, multi dimensional visualisation, etc

The VOTECH project, is led by [AstroGrid](http://www.astrogrid.org)<sup>3</sup> in the UK, with partners: ESO, Centre National de la Recherche Scientifique (France) and the Istituto Nazionale di Astrofisica (Italy). In the wider VOTC context, ESA, through its [ESAVO](http://esavo.esa.int)<sup>4</sup> initiative is also working closely with the VOTECH project in developing new technologies for eventual deployment by the Euro-VO.

<sup>3</sup><http://www.astrogrid.org>

<sup>4</sup><http://esavo.esa.int>

### 3.4. Other VO Initiatives

In addition to the UK and European VO initiatives described above, a number of other major VO projects exist. These include the National Virtual Observatory (NVO) (see <http://www.us-vo.org> in the USA, and the Japanese Virtual Observatory (JVO) see <http://jvo.nao.ac.jp/>). All projects are committed to ensuring that the specific services that they provide are accessible to all other VO projects. Thus to an end user, use of any one VO project service, opens up access to data resources made available by a full range of global providers.

## 4. ASTROGRID: THE UK'S VIRTUAL OBSERVATORY INITIATIVE

In the UK, the [AstroGrid](http://www.astrogrid.org)<sup>5</sup> consortium was formed and begun activities late 2001. AstroGrid is a substantial Virtual Observatory initiative funded through the UK's Particle Physics and Astronomy Research Council eScience programme. It is a consortium of eleven UK institutes and research establishments with expertise in data handling and computational science. After an initial definition and requirements analysis phase, it moved into a major development phase as described in [Walton et al. \(2004\)](#). The focus in 2005 and onwards is now on the

<sup>5</sup><http://www.astrogrid.org>

transition from a developmental system to a more operationally focused system, one robust and capable enough to support significant astronomer usage.

#### 4.1. AstroGrid Releases v1.0/ v1.1

With its version 1.0 release (May 2005), AstroGrid is providing a scientifically functional Virtual Observatory (VO) implementation, a system built conforming to emerging Interoperability and Grid standards. The v1.0 release (see [Walton \(2005\)](#) for a fuller description) provides a number of key capabilities:

1. **Data Set Access:** These access modules are run at each participating data provider - and are configured to allow visibility of the underlying data resources, through the AstroGrid framework. As an astronomer, one is able to perform database queries of remotely held data, utilising a standard query language (Astronomical Data Query Language, [ADQL](#)<sup>6</sup>, which is based on standard SQL). Alternatively they can access data held in flat files (e.g. FITS files), be they images (through a standard 'Simple Image Access Protocol' [SIAP](#)<sup>7</sup>, spectra, etc.
2. **Registry:** Each resource, be that data, application, hardware platform (e.g. CPU, disk, network), is registered to AstroGrid. This registry entry describes that resource, following the IVOA's interoperability standard. (Thus the AstroGrid registry 'harvests the registries of other VO projects, e.g. the NVO.) The registry entry is sufficiently detailed to give the system the level of knowledge about that resource, to make sensible assessments of whether it is relevant or not.
3. **Common Execution Architecture:** This provides a standardised framework in which any application can be run within the AstroGrid system. Each application available through the CEA is 'registered' in the **Registry**, thus is discoverable by the astronomer.
4. **Workflow:** this allows the ability to construct chains of processes, allowing user defined processing sequences to be carried out. This could be very simple, perhaps do a one step query of a database, or extremely complicated, with many steps. The workflow builder allows the creation of these work flows. They can be saved, and thus re-run at will.
5. **Job Execution Scheduler:** this underpins **workflow**. It takes the work flows submitted by the user, and sends the instructions contained therein to the various resources around the AstroGrid.
6. **MySpace:** This is the users own secure virtual storage made up of a number of large disk arrays hosted on the AstroGrid system. To the user, MySpace is storage space accessible by them, and available to store their work, thus data - especially intermediate data, work flows, results of catalogue queries and so forth. Because MySpace servers are located on fast networks, it is ideal for storing large data sets,

as these can be quickly uploaded from MySpace to applications that may be needed to work on them.

These services are available to the end user through either a web based portal or a client 'workbench application' (see Fig. 2). The system allows for the creation of communities (thus groups, access control and so forth), thus facilitating group working on shared data sets,

A range of client side applications configured to interface to the server services, such as the MySpace file storage system, are available to the astronomer to allow for visualisation and end stage analysis of results generated by for instance server side large scale processing jobs. These clients include: **Aladin** ([Boch et al., 2004](#)), a powerful image visualisation and catalogue access tool and **Topcat**<sup>8</sup> which enables visualisation and analysis of tabular datasets.

## 5. USING VO TOOLS

Today, and the coming years will see an explosion in the research possibilities opened up by use of new VO tools. Already there exists a large degree of useful capability available to the astronomer. In the next sections we describe the 'science services' available through use of the AstroGrid system, and note the use of Euro-VO developed systems in the discovery of black hole populations.

### 5.1. AstroGrid Science Services

The initial AstroGrid release represents a paradigm shift in the way in which astronomers approach and carry out on-line data analysis and interpretation. AstroGrid's workflow based system enables many complex tasks to be undertaken.

However, it is recognised that for the new user, the basic system may require a significant learning curve before the user is able to fully exploit the system. Therefore, AstroGrid provides a number of exemplar 'science services' which solve specific astronomical problems. These science services, provide a simple way in to the system for new users of AstroGrid to understand the operation of the system, are ready to use in science analysis, and can be adapted by the astronomer if required. The services include:

1. **Redshift Maker:** Photometric red-shifts use broad band photometry to measure the red-shifts of galaxies rather than spectroscopy. With large imaging surveys like the Sloan Digital Sky Survey ([Abazajian et al., 2003](#)), the Isaac Newton Telescope Wide Field Survey (WFS) ([McMahon et al., 2001](#)), 2MASS (see <http://www.ipac.caltech.edu/2mass>), WFCAM on the United Kingdom's Infrared telescope (UKIRT),

<sup>6</sup><http://www.ivoa.net/twiki/bin/view/IVOA/IvoaVOQL>

<sup>7</sup><http://www.ivoa.net/Documents/latest/SIA.html>

<sup>8</sup><http://www.star.bristol.ac.uk/mbt/topcat/>



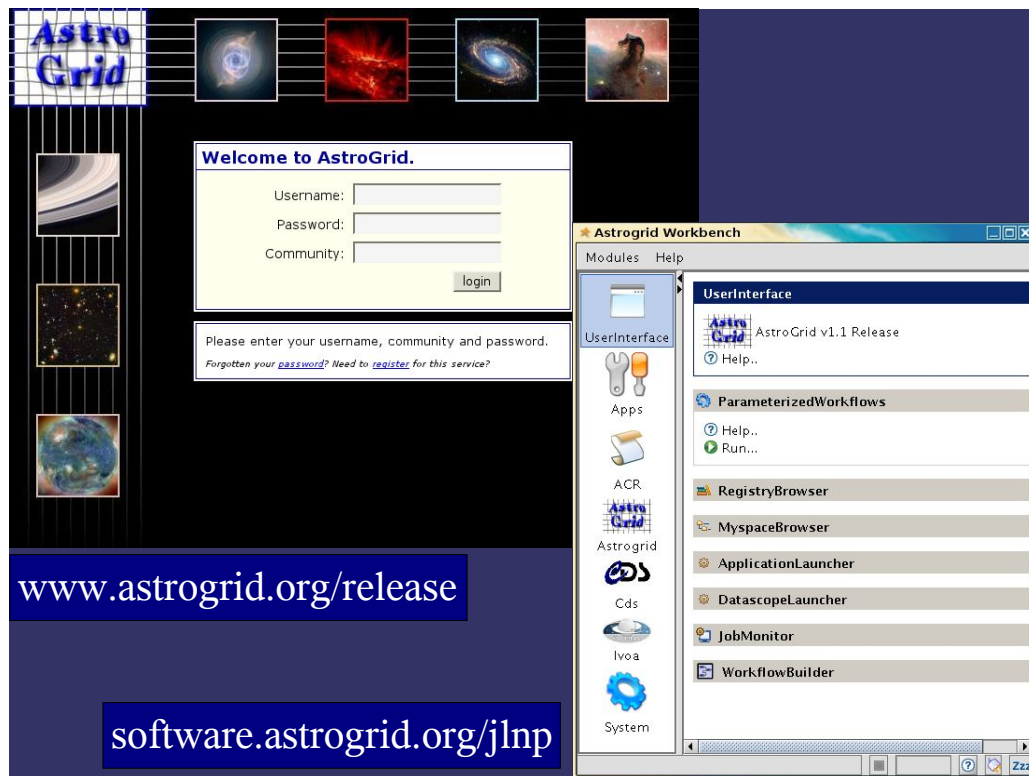


Figure 2. Entrance into the UK AstroGrid system, via either the portal or the desktop workbench

it is practically impossible to carry out a spectroscopic campaign to determine the red-shifts of all the sources detected. Photometric red-shifts have larger uncertainties than spectroscopic red-shifts, but they are a way of determining the properties of large samples of galaxies.

The redshift maker allows the user to enter a positional search. The workflow then handles a number of tasks. The INT's WFS reduced image data files are searched for multi-colour broadband image data. If available, these images are automatically transferred from the archive and sent to an application which extracts object catalogues from each of the images. These individual catalogues are then associated to produce a master catalogue of objects with their multi-colour broadband fluxes. This catalogue is then fed into a further application which produces photometric red-shifts to each of the objects based on their colours.

The redshift maker, when run by the user through the AstroGrid Workbench, asks for a basic set of parameters, in this case the position on the sky. The workflow is then automatically run. However, the workflow is also saved to the user's MySpace area. The user can then adapt the workflow, by altering the default configuration parameters for each of the steps and applications in the workflow, and run their personalised workflow. This might allow for instance the user to adjust the detection level at which objects are extracted from the image files.

Over the coming year, this service will be ex-

panded to allow for the selection of image data from a wider collection of image survey's such as UKIDSS<sup>9</sup> (Warren, 2002) - a public access infrared imaging survey). Importantly this service is ideal for use in analysis of deep X-ray data observation of galaxy clusters, when comparing with for instance deep optical and IR data to determine the distances of objects contained within those fields (e.g. Kodama et al. (2004)).

2. **Colour Cutter:** allowing object retrieval from a number of large databases based on user selected colour cuts. The user can configure this service to access X-ray catalogues such as 1XMM/ 2XMM (XMM SSC Consortium, 2003)
3. **Astroscope:** returning multiple resources concerning an area on the sky, adding value and building on the NVO developed 'Datascope'<sup>10</sup> product.
4. **Solar Movie Maker:**

The AstroGrid system is also of relevance to the solar physics community, through the access to solar data sets and tools. With this science service, the solar physicist can enter a time range, the service then will locate, retrieve, process and create movies from SOHO Extreme ultraviolet Imaging Telescope (EIT), SOHO Coronal Diagnostic Spectrometer (CDS) or TRACE image data over the desired time range.

<sup>9</sup><http://www.ukidss.org>

<sup>10</sup><http://heasarc.gsfc.nasa.gov/cgi-bin/vo/datascope/init.pl>

## 5.2. The AVO Demonstrator Products

Through the AVO programme, science cases were used to define certain VO experiments bringing together specifically developed tools for these cases. One of these was demonstrated in 2004, where various deep survey data from the GOODS (Giavalisco et al., 2004) were mined using VO tools to discover a significantly enlarged population of obscured, high luminosity Type 2 QSO's. These indicate the presence of a larger population of massive black holes and their host active galaxies than was previously predicted. This work was enabled via the access to key X-ray data including that from Chandra, with objects from that been cross matched with deep optical survey data. The work is fully described in Padovani et al. (2004).

## 6. CONCLUDING REMARKS

AstroGrid and other VO's are providing systems, with sufficient flexibility, to enable astronomers to carry out a wide range of astrophysical data analysis tasks in support of their research. For the X-ray community, the Virtual Observatories provide access to many major relevant X-ray and related data and application resources.

We note that the initial deployment of AstroGrid release v1.0 provides the world's first fully functional integrated virtual observatory system, an emerging scientific capability with the potential equivalence of a major new observational facility. Users can find more details of how to use the AstroGrid system at <http://www.astrogrid.org/release>

## ACKNOWLEDGMENTS

AstroGrid is funded by the UK's PPARC through its eScience programme. The Euro-VO VOTECH project is supported by the EU under the FP6 Research Infrastructures programme, contract no. 011892-DS. The AVO project was funded by through the EU Fifth Framework programme under contract HPRI-CT-2001-50030.

## REFERENCES

- Abazajian K., The Sloan Consortium, 2003, ApJ 126, 2081
- Boch T., Fernique P., Bonnarel F. 2004, in Proc ADASSXIII, eds F. Ochsenbein, M. G. Allen, D. Egret, ASP Conference Proceedings, 314, 221
- Gehrels N., et al. 2004, ApJ, 611, 1005
- Giavalisco M., et al. 2004, ApJ 600, L93
- Kodama T., et al. 2004, MNRAS 350, 1005
- McMahon R. G., et al. 2001, NewAR, 45, 97
- Padovani P, Allen M. G., Rosati P., Walton N. A. 2004, A&A 424, 545
- Williams R, et al., 2004, IVOA Note 2004-06-14: <http://www.ivoa.net/Documents/Notes/IVOAArch/IVOAArch-20040615.html>
- Walton N. A., Lawrence A., Linde A. E., 2004, Proc SPIE, 5493, 146
- Walton N. A. 2005, A&G 46, 23
- Warren S. 2002, Proc SPIE, 4836, 313
- XMM-Newton Survey Science Centre, Consortium, 2003, VizieR On-line Data Catalog: IX/37