Astronomical Wide-field Imaging System for Europe

a partnership of

NOVA / OmegaCEN/Kapteyn Institute, Groningen - NL

INAF / Osservatorio Astronomico di Capodimonte, Napoli - I

CNRS / Terapix, IAP, Paris - F

ESO, Garching bei München - D

Max-Planck-Institute fur extraterrestrische Physik /Universitäts-Sternwarte München - D

co-ordinated by

OmegaCEN-NOVA/Kapteyn Institute- University of Groningen - NL

An on-going project which started from a FP5 RTD programme funded by the EC Action "Enhancing Access to Research Infrastructures".

www.astro-wise.org

What is Astro-WISE?

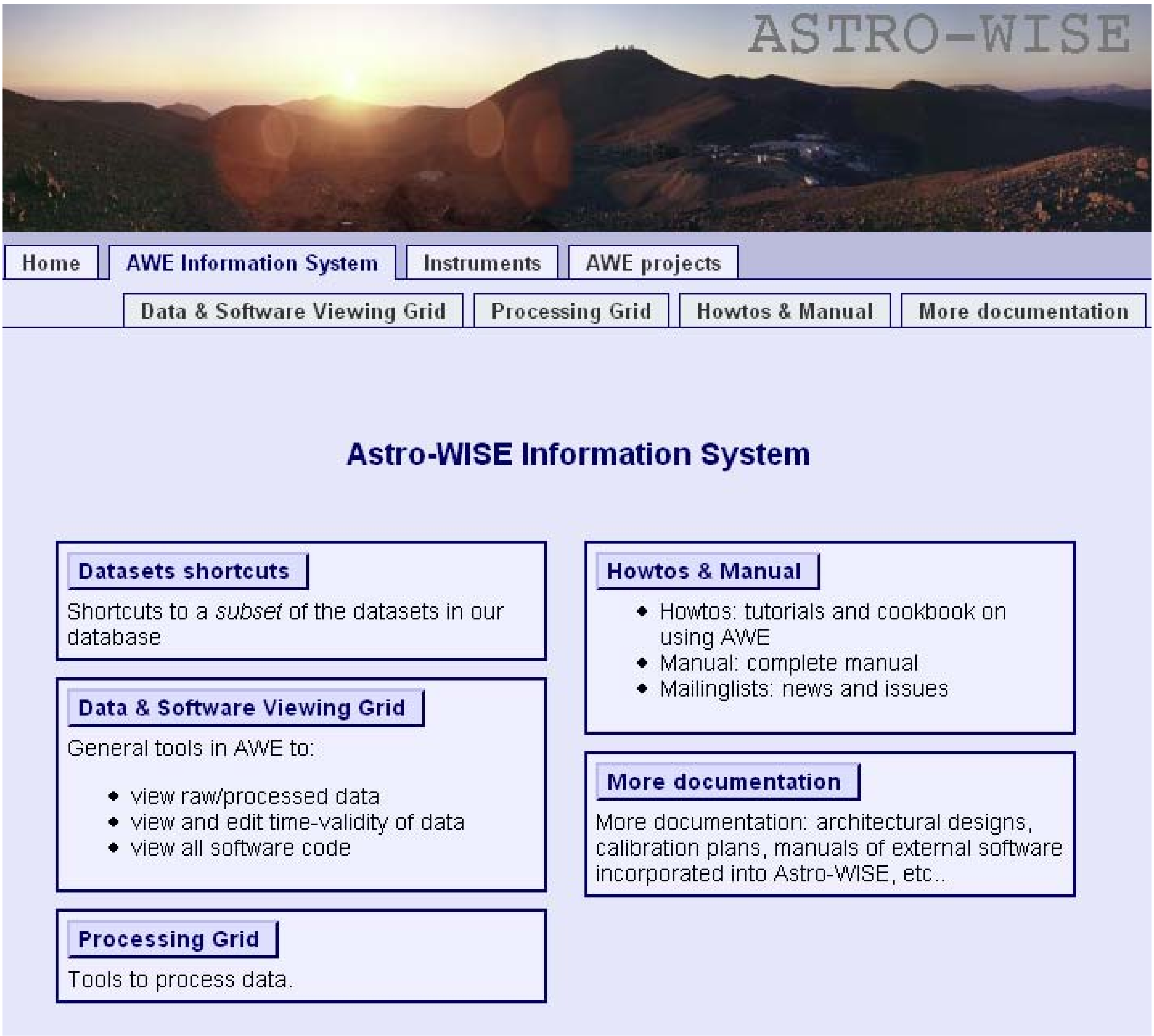
The Astronomical Wide-field Imaging System for Europe, Astro-WISE, is an environment consisting of hardware and software which has been developed to be able to scientifically exploit the ever increasing avalanche of observational data produced by science experiments. Astro-WISE started out as a system geared towards astronomy, but is now also being used for projects outside astronomy. Astro-WISE is an all-in-one information system: it allows scientists to archive raw data, to process and calibrate data, perform post-calibration scientific analysis and archive all results in a distributed environment. The system architecture links together all these steps of the data flow. The **complete** linking of **all** input and output in the data flow, including software code used, for arbitrary data volumes has only been feasible thanks to a novel paradigm devised by the creators of Astro-WISE, named “target processing”. The Astro-WISE information system started with one particular astronomical optical wide field imager, OmegaCAM. Subsequently, the system was expanded to host arbitrary optical wide field imagers, radio data and any other kind of digitized images.

Salient features of the Astro-WISE information system include:

- can be installed everywhere on local hardware, sharing remote hardware when needed
- fully scalable: database, parallelized code, compute and storage grid (Peta byte regime)
- no local installation required to use it
- can work with observational data of any digital imager
- share methods, raw and result data with collaborators globally, automatically, in real time
- connects directly to the e-infrastructures of the European Virtual Observatory (EURO-VO), AstroGrid and GRID (EGEE)
- tunable user privileges for non-public data
- provides a web-based archive (called DBviewer) to view an arbitrary subset of data of an arbitrary project at each stage of the processing (e.g., final catalogs, calibrated images of an imaging survey using multiple instruments, raw science data, raw calibration data)
- provides webtools to process arbitrary amounts of data
- provides a command-line prompt with the same capabilities
- based on the Python programming language
- easy to add own user provided analysis code in Python
- access to compute power of Astro-Wise clusters (grid infrastructure is part of the system)
- access to distributed data storage of Astro-Wise servers

Astro-WISE is an e-Science information system providing a distributed environment in which very large datasets from science experiments can be processed, scientifically exploited, published and archived.

Contact:
Prof Dr Edwin A. Valentijn
Coordinator of Astro-WISE Consortium
valentyn@astro.rug.nl
OmegaCEN – Kapteyn Institute – University of Groningen – The Netherlands



Astro-WISE Information System

Datasets shortcuts

Shortcuts to a *subset* of the datasets in our database

Data & Software Viewing Grid

General tools in AWE to:

- view raw/processed data
- view and edit time-validity of data
- view all software code

Processing Grid

Tools to process data.

Howtos & Manual

- Howtos: tutorials and cookbook on using AWE
- Manual: complete manual
- Mailinglists: news and issues

More documentation

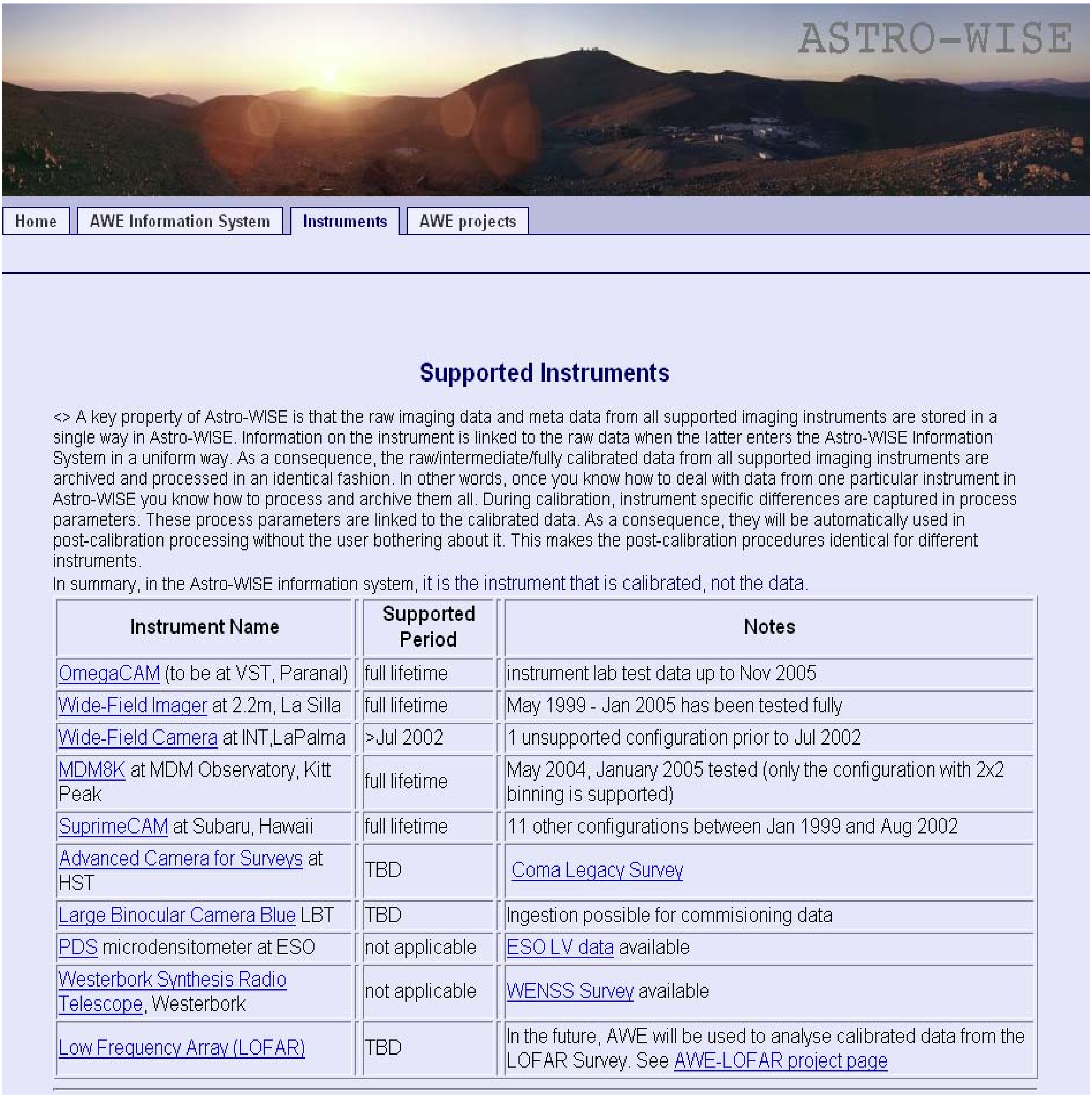
More documentation: architectural designs, calibration plans, manuals of external software incorporated into Astro-WISE, etc..

Contents

At the occasion of the full installation and the delivery to the EU of the Astro-WISE system we have compiled this booklet with some key descriptions of the system. Detailed information and “the system at work” can be viewed on the www.astro-wise.org web pages, from which some snapshots are reproduced on the cover and on these pages.

The paper printed in ERCIM news in 2006 gives a top-level description, while the ADASS paper printed in 2007 details some of the design features.

Astro-WISE plans to support several new Astronomical instruments which are expected to come on-line in 2008 and have operational lifetimes of at least 10 years. In the last section of this compilation we reproduce excerpts of the currently planned programme of Astro-WISE in 2008 and beyond.

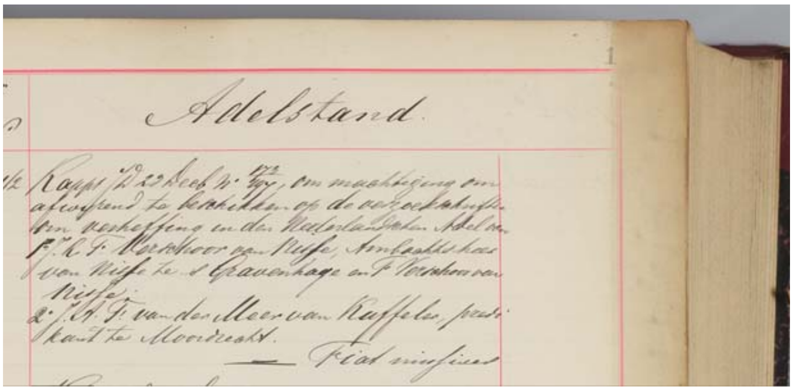


Supported Instruments

<> A key property of Astro-WISE is that the raw imaging data and meta data from all supported imaging instruments are stored in a single way in Astro-WISE. Information on the instrument is linked to the raw data when the latter enters the Astro-WISE Information System in a uniform way. As a consequence, the raw/intermediate/fully calibrated data from all supported imaging instruments are archived and processed in an identical fashion. In other words, once you know how to deal with data from one particular instrument in Astro-WISE you know how to process and archive them all. During calibration, instrument specific differences are captured in process parameters. These process parameters are linked to the calibrated data. As a consequence, they will be automatically used in post-calibration processing without the user bothering about it. This makes the post-calibration procedures identical for different instruments.

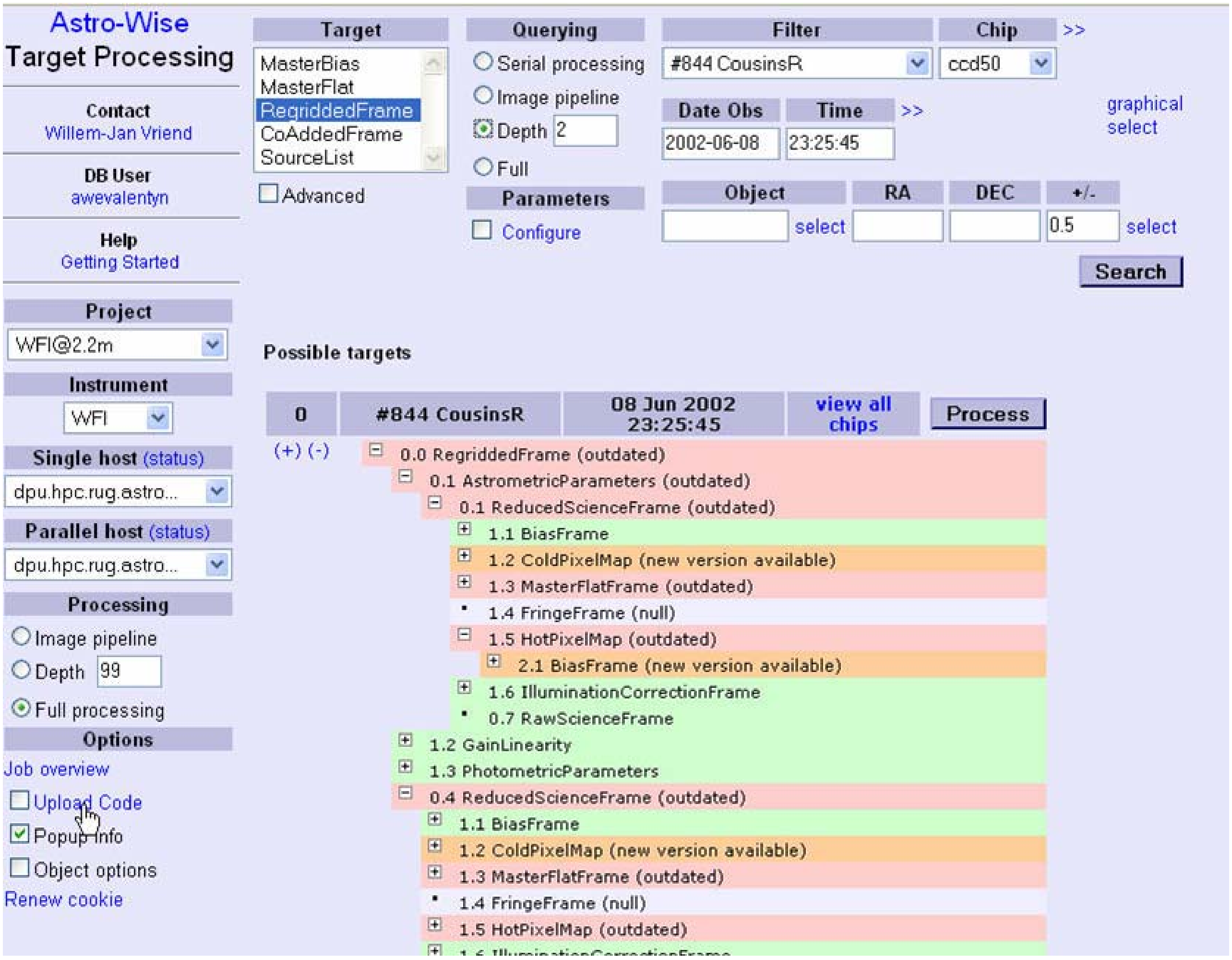
In summary, in the Astro-WISE information system, it is the instrument that is calibrated, not the data.

Instrument Name	Supported Period	Notes
OmegaCAM (to be at VST, Paranal)	full lifetime	instrument lab test data up to Nov 2005
Wide-Field Imager at 2.2m, La Silla	full lifetime	May 1999 - Jan 2005 has been tested fully
Wide-Field Camera at INT,LaPalma	>Jul 2002	1 unsupported configuration prior to Jul 2002
MDM8K at MDM Observatory, Kitt Peak	full lifetime	May 2004, January 2005 tested (only the configuration with 2x2 binning is supported)
SuprimeCAM at Subaru, Hawaii	full lifetime	11 other configurations between Jan 1999 and Aug 2002
Advanced Camera for Surveys at HST	TBD	Coma Legacy Survey
Large Binocular Camera Blue LBT	TBD	Ingestion possible for commisioning data
PDS microdensitometer at ESO	not applicable	ESO LV data available
Westerbork Synthesis Radio Telescope, Westerbork	not applicable	WENSS Survey available
Low Frequency Array (LOFAR)	TBD	In the future, AWE will be used to analyse calibrated data from the LOFAR Survey. See AWE-LOFAR project page



The first example of the Astro-WISE information system being used outside astronomy: images of handwritten texts from the Dutch government are currently being analyzed using Astro-WISE, aiming to provide easy access to 100’s meters of paper archives by means of intelligent computing. Handwriting recognition code is run on the IBM Blue Gene supercomputer to interpret the archives of the Kabinet van de[n] Koning[in]

Target Processing



Project: WFI@2.2m
Instrument: WFI
Single host (status): dpu.hpc.rug.astro...
Parallel host (status): dpu.hpc.rug.astro...
Processing: ○ Image pipeline ○ Depth 99 ● Full processing

Target: RegriddedFrame
Querying: ● Depth 2 ○ Full
Filter: #844 CousinsR
Chip: ccd50
Date Obs: 2002-06-08
Time: 23:25:45
+/-: 0.5

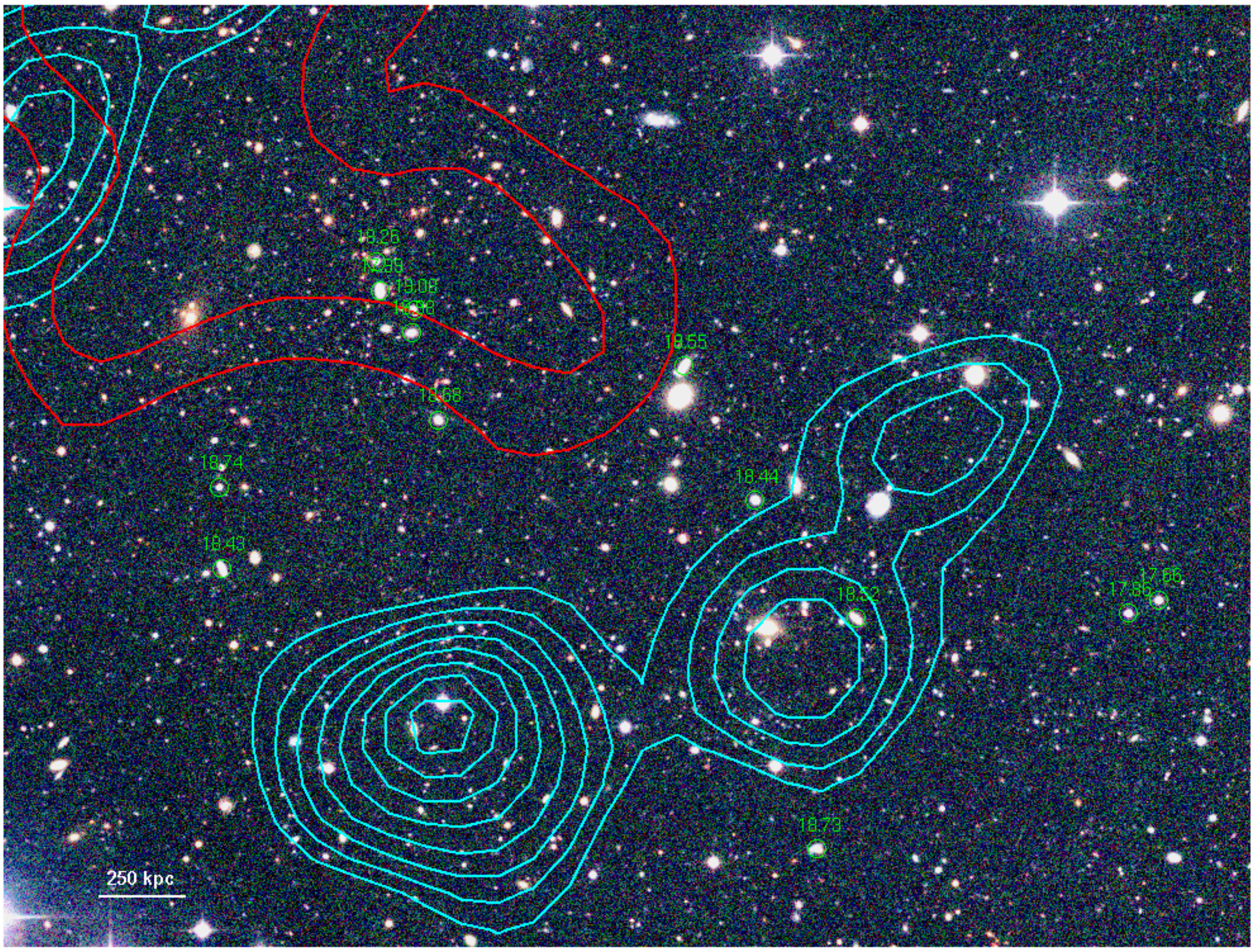
0	#844 CousinsR	08 Jun 2002 23:25:45

```
0.0 RegriddedFrame (outdated)
  0.1 AstrometricParameters (outdated)
    0.1 ReducedScienceFrame (outdated)
      1.1 BiasFrame
      1.2 ColdPixelMap (new version available)
      1.3 MasterFlatFrame (outdated)
      1.4 FringeFrame (null)
      1.5 HotPixelMap (outdated)
        2.1 BiasFrame (new version available)
      1.6 IlluminationCorrectionFrame
      0.7 RawScienceFrame
  1.2 GainLinearity
  1.3 PhotometricParameters
  0.4 ReducedScienceFrame (outdated)
    1.1 BiasFrame
    1.2 ColdPixelMap (new version available)
    1.3 MasterFlatFrame (outdated)
    1.4 FringeFrame (null)
    1.5 HotPixelMap (outdated)
```

Screenshot of the webservice of the Astro-WISE Target Processor. The users picks a target (desired result) and the system analyses the status of that target. If it has never been computed before the system initiates its derivation. If it has been made before, the system constructs a roadmap indicating which dependencies are outdated and require re-processing. The user decides to re-process dependencies and selects a cluster of nodes for the processing.

In the example, the user sets out for an astrometrically calibrated and undistorted image (“RegriddedFrame”) of an observation taken in the red band of the WFI wide field imager on 8 June 2002. The system finds a target conforming to the query, but concludes that better versions of the “ColdPixelMap” and the “BiasFrame” calibrations are available, and re-processing is required. Hitting the “Process” button initiates this. Dependencies can go over 20 levels deep, and users can tune the depth of the analysis and the re-processing. An extra graphical select service helps the user to define his target and even to select groups of targets (e.g., a night of astronomical observations).

Wide field astronomical imaging - finding rare objects



A typical Wide field imaging result from the WFI instrument at ESO.

The picture shows a small region of a much larger image taken of the merging galaxy cluster A2RXCJ0014.3-3022. Contours indicate surface densities of blue and red galaxies. The red contours form a part of the outer regions of the cluster. The blue contours represent an over density of blue galaxies in the inter cluster space. The labeled galaxies are very bright blue galaxies found in the outskirts of the cluster which appear to flame up when falling from outer inter cluster space into the cluster, suggesting we are witnessing the morphological transitions of galaxies when they fall into the cluster domain. The galaxy evolution part of the KIDS and VESUVIO programmes plans, among other things, to locate and study galaxies during their transitional phases.

Courtesy Braglia et al., Astronomy and Astrophysics 470, 425 (2007)

The Astro-Wise System: A Federated Information Accumulator for Astronomy

by Edwin A.Valentijn and Gijs Verdoes Kleijn

The progress of astronomy is about to hit a wall in terms of the processing, mining and interpretation of huge datasets. The Astro-Wise consortium has designed and implemented a fully scalable and distributed information system to overcome this problem for wide-field imaging. The same principles can be applied to other sciences.

Much of modern research involves the accumulation of huge amounts of digitized data. The analysis of this data by distributed communities represents a significant challenge to project management and ICT implementation, and is relevant to fields as diverse as biology, physics, astronomy, economics and cultural heritage projects. Furthermore, the projects are often global efforts requiring collaborators in many places to share, validate and combine processed data and derived results. It is therefore necessary to develop more efficient data lineage, mining and analysis systems to allow researchers to search intelligently through previously unmanageable volumes of data.

The Astro-Wise consortium has developed an information system to meet these challenges for wide-field imaging in astronomy. The Astro-Wise consortium is a partnership between OmegaCEN-NOVA/Kapteyn Institute (Groningen, The Netherlands; coordinator), Osservatorio Astronomico di Capodimonte (Naples, Italy), Terapix at IAP (Paris, France), ESO, Universitäts-Sternwarte & Max-Planck Institut für Extraterrestrische Physik (Munich, Germany).

Large data projects in high-energy physics, space missions and astronomy typically push data through various platforms in an irreversible way (eg a TIER node setting). In such a situation, the end user has little or no influence on what happens upstream. This 'classical' paradigm is characterized by fixed 'releases' of homogeneous, well-documented data products. In contrast, the Astro-Wise system allows the end user to trace the data product, following all its dependencies up to the raw observational data and, if necessary, to re-derive the result with better calibration data and/or improved methods.

This improvement is achieved by:

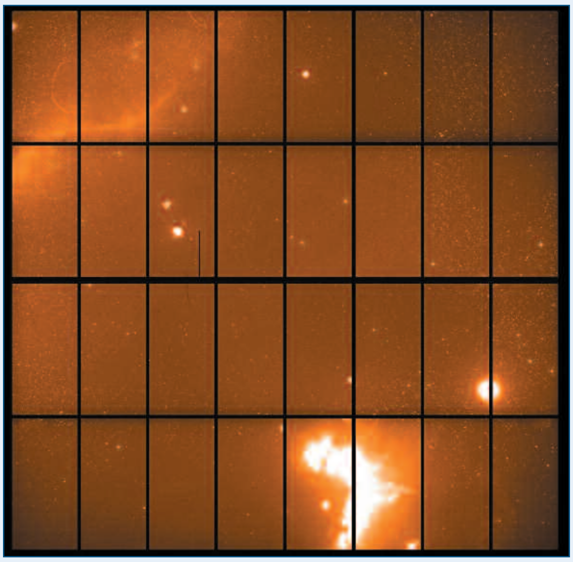
- emphasis on project management; enforcing a global data acquisition and processing model, while retaining flexibility
- translating the data model to an object model, with full registration of all dependencies
- storing all I/O of the project in a single, distributed database, containing all metadata describing the bulk data (eg images) and derived results in catalogue form (eg lists of celestial sources).
- connecting to the database a federated file server that stores hundreds of Terabytes of bulk data
- an own compute-GRID which sends jobs (including clients) to single nodes or parallel clusters, which then request data from the distributed database.

The database with all metadata and catalogues provides the infrastructure to develop tools for a variety of purposes. These include rapid trend analysis of data, complex queries and fast hunting for 'needles in the haystack' of Terabyte-sized catalogues. Thus, the system provides the user with fully integrated, transparent access to all stages of the data processing and thereby allows the data to be reprocessed and the system to be improved and expanded.

For a given project/instrument, the system initially starts in a naive, 'quick look' mode, which gradually improves as various researchers add refined information to the system under the supervision of project leaders. Approved calibration modifications automatically become public, beyond the project boundaries. A mechanism for quality control is implemented which allows for changes due to one of:

- true physical changes of parameter values
- improvements in encoded methods, or
- improved insight in either of these.

The core of the system exploits three properties in database environment. First, we apply the principle of inheritance using Object Oriented Programming (Python), where all Astro-Wise objects inherit key properties for database access, such as persistency of attributes. Second, the linking (associations or references) between instances of objects in the database is completely maintained, and for each



A 256 Mega pixel test image of the OmegaCAM instrument, which consists of 32 eight Megapixel CCDs.

bit of information, it is possible to trace those bits of information that were used to obtain it. Third, each step, and the inputs used for it, is kept within the system. The database grows constantly through the addition of new information or improvements made to existing information.

All system components are distributed over Europe, enabling research groups to collaborate on shared projects. Knowledge added by one group is immediately accessible by others via a Web portal, which includes data viewing, quality labelling and compute-services (see links). Currently, researchers use the Astro-Wise system with 10 Tbyte of astronomical images. Hundreds of Tbytes of data will start entering the system when the OmegaCAM panoramic camera starts operations in Chile. This camera is dedicated to various large surveys using the Astro-Wise system.

Astro-Wise coordinator OmegaCEN-NOVA is collaborating with the LOFAR consortium and CWI to explore usage of the Astro-Wise system for LOFAR, the next generation Low Frequency Array of radio telescopes, which is being built in the Netherlands and Germany. Astro-Wise can also be applied to other fields of science. The object-oriented use of the database allows for classes of objects dealing with arbitrary forms of digitized observational data. Scans of cultural heritage, DNA sequences, data from high-energy particle collisions or financial markets can be processed using similar principles to the images of the sky.

Links:
http://www.astro-wise.org
http://www.astro-wise.org/portal
http://www.astro.rug.nl/~omegacam
http://www.astro.rug.nl/~omegacen
http://www.lofar.nl

Please contact:
Edwin Valentijn, Astro-Wise Consortium
OmegaCEN-NOVA/Kapteyn Astronomical Institute, Groningen, The Netherlands
Tel: +31 50 3634011
E-mail: valentyn@astro.rug.nl

Astronomical Data Analysis Software and Systems XVI
ASP Conference Series, Vol. 376, 2007
R. Shaw, F. Hill and D. Bell, eds.

Astro-WISE: Chaining to the Universe

Edwin A. Valentijn, John P. McFarland, Jan Snigula†, Kor G. Begeman, Danny R. Boxhoorn, Roeland Rengelink, Ewout Helmich, Philippe Heraudeau, Gijs Verdoes Kleijn, Ronald Vermeij, Willem-Jan Vriend, Michiel J.Tempelaar

OmegaCEN, Kapteyn Astronomical Institute, Groningen University, PO Box 800, NL-9700 AV, Groningen, The Netherlands

Erik Deul, Konrad Kuijken

Sterrewacht Leiden, University Leiden, PO Box 9513, NL-2300 RA, Leiden, The Netherlands

Massimo Capaccioli, Roberto Silvotti

INAF, Osservatorio Astronomico di Capodimonte, Salita Moiariello 16, I-80131, Napoli, Italy

Ralf Bender, Mark Neeser, Roberto Saglia

†*Universitäts-Sternwarte München, Scheinerstr. 1, D-81679 München, Germany*

Emmanuel Bertin, Yannick Mellier

Terapix, Institut d'Astrophysique de Paris, 98bis bd Arago, F-75014, Paris, France

Abstract. The recent explosion of recorded digital data and its processed derivatives threatens to overwhelm researchers when analysing their experimental data or when looking up data items in archives and file systems. While current hardware developments allow to acquire, process and store 100s of terabytes of data at the cost of a modern sports car, the software systems to handle these data are lagging behind. This problem is very general and is well recognized by various scientific communities; several large projects have been initiated, e.g., DATAGRID/EGEE[1] federates compute and storage power over the high-energy physical community, while the international astronomical community is building an Internet geared Virtual Observatory[2] connecting archival data. These large projects either focus on a specific distribution aspect or aim to connect many sub-communities and have a relatively long trajectory for setting standards and a common layer. Here, we report "first light" of a very different solution (Valentijn & Kuijken, 2004) to the problem initiated by a smaller astronomical IT community. It provides the abstract "scientific information layer" which integrates distributed scientific analysis with distributed processing and federated archiving and publishing. By designing new abstractions and mixing in old ones, a Science Information System with fully scalable cornerstones has been achieved, transforming data systems into knowledge systems. This break-through is facilitated by the full end-to-end linking of all dependent data items, which allows full backward chaining from the observer/researcher to the experiment. Key is the notion that information is intrinsic in nature and thus is the data acquired by a scientific experiment. The new abstraction is that software systems guide the user to that intrinsic information by forcing full backward and forward chaining in the data modelling.

---

[1] http://www.eu-egee.org/

[2] http://www.euro-vo.org/pub/ and Padovani, 2006

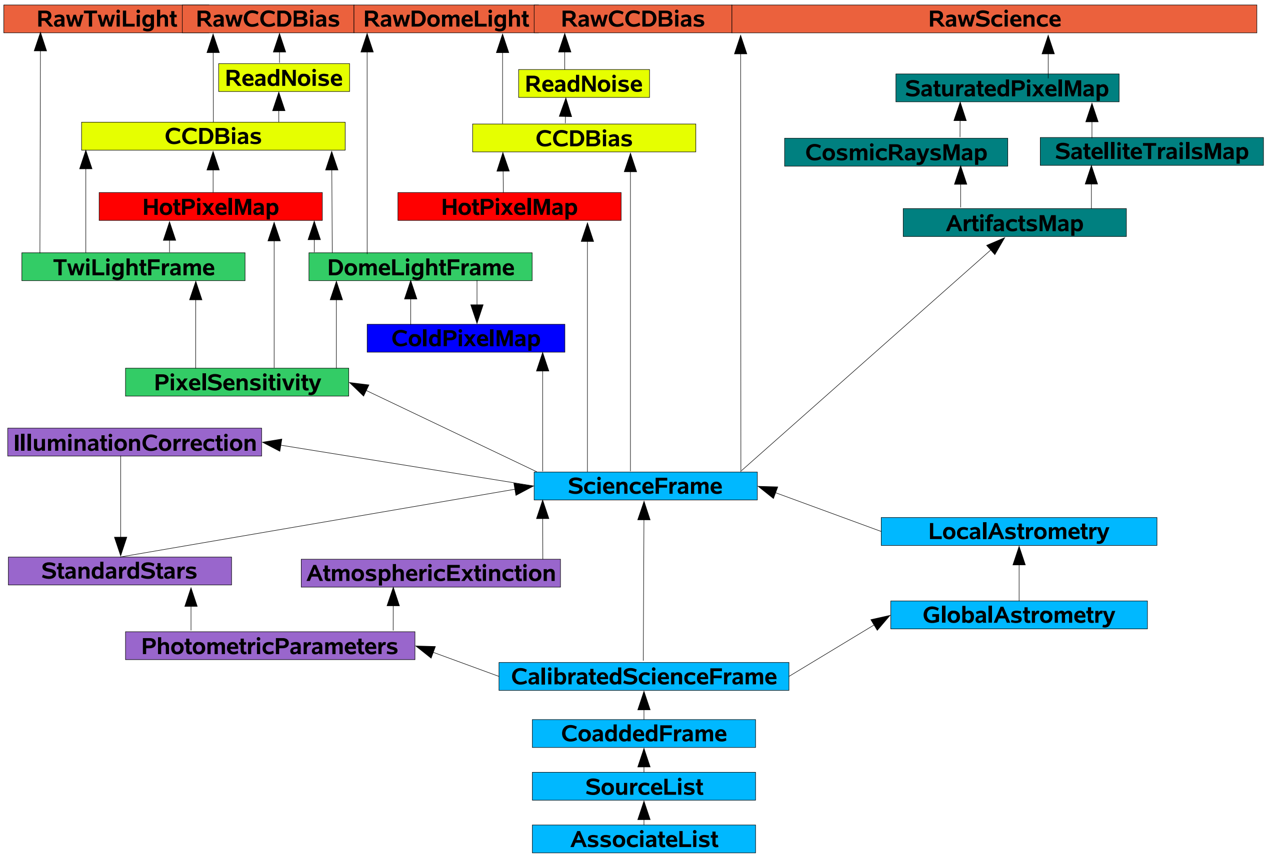


Figure 1. A *target diagram*: slightly simplified view of the dependencies of “targets” to the ocean of raw observational data of astronomical wide field imaging. Arrows indicate the backward chaining to the raw data.

1. Introduction

The classical paradigm to handle data streams of large physical experiments, such as CERN-HEP, astronomical telescopes, or scientific satellites, is characterized by different layers in the project that apply certain algorithms to the stream of data and subsequently deliver the results to the next layer, following a so-called TIER architecture. This architecture can be characterized as data-driven and “feed-forward”. The construction of the different layers is often grown historically and relates to implicit or explicit project management decisions facing sociology, geography and interfacing of working groups with different expertise. Projects operated under this paradigm work with releases of datasets, which are obtained with a certain version of the code meeting certain quality control standards. Operators have a task to push the input data through the stream, often by means of a “pipeline” irrespective of whether the derived data items are actually used by the end users. Examples of successful forward workflow projects are numerous (e.g., Valentijn & Kuijken, 2004; Costa et al., 2003; Hiriart et al., 2003; Jung et al., 2003; Mehringer & Plante 2003; and Pierfederici et al., 2003). When funding for the operation centres dries up, a final release of processed data is made, and the project comes to an end.

2. The Problem

This “classical paradigm” has the advantage of publishing complete sets of data with well described methods, qualifications, and calibrations. However, it has bad scalability when exposed to modern huge data streams: re-processing all the raw data and storing the results for new releases when new computational methods, calibration strategies, insights or improved code becomes available is very difficult to impossible. And, moreover, why should operators re-process all data as long as it is even not known whether there are customers for the individual products? Though the classical approach has the obvious advantage of facilitating established releases, it cannot facilitate the demands of the end-user/researcher, who actually might want to evaluate specific questions, and specific results obtained with specific methods from (a subset) in the flood of physical observational data (including even holiday digital pictures or scanned documents of national libraries) with. Key to addressing the merits of a system for a researcher is to evaluate the question of what analysis on uncertainties the researcher facilitates in that system before (s)he submits his/her 4 sigma result to a scientific journal.

Such demands combined with the increased data rates from instruments and world-wide communities ask for a different approach. A first interesting step is set by the Space Telescope Science Institute, operating the Hubble Space Telescope, where users can ask for a certain data product which will then run the Calibration pipeline on-demand using the best available calibrations (Swade et al., 2001), a case of forward chaining. But what to do when a user discovers a very faint object and wants to inspect the robustness of the result by checking the dependencies on uncertainties of calibration parameters and applied methods? The user wants to interact with the data, derive his own result following his insights and methods, and preferably this knowledge is fed back into the system at the disposal of other researchers who optionally want to take advantage of the progress of insight. In the 1980s, solutions were explored in expert systems. Nowadays, with distributed communities working on enormous data floods, new artificial intelligence-like, distributed information systems are required to facilitate the scientific endeavour.

3. The Solution

A new generation of wide field astronomical imaging cameras includes the MEGACAM at the CFHT and OmegaCAM (Kuijken et al., 2004) at the VLT Survey Telescope (VST). OmegaCAM, with a one square degree field of view and a pixel size of 0.2 arc second, will deliver 256 mega-pixel images of the sky every few minutes, for 300 nights per year for many years to come. These experiments will produce 100's of terabyte per year of raw and processed data, e.g., to trace the effects of dark matter and dark energy. These requirements inspired an international consortium, Astro-WISE, to design and implement a completely different approach to connect the end-user to the experiment. Facing the problems sketched above we build an information system in which the role of the user (observer) is central. The system is built to handle queries by the user for his/her desired result, which we call a "target". A "target" could be an astrometrically and photometrically calibrated image, or the results for a set of calibration parameters, or a list of parameter values describing an astronomical object, or whatever "target" that is facilitated by the system. In the abstractions used for the software the "target" processing resembles that of the UNIX make metaphor, but in addition the dynamic aspects resembles to the "goal" in artificial intelligence systems.

4. How it Works

Following a query by the user, the system checks whether the "target" has been processed before,

- if not, it will be processed "on-the-fly", recursively following all its dependencies in an object model, similar to the Unix make metaphor. See the "target diagram" in Fig. 1.
- if already existent the system will check all its dependencies up to the raw data taken at the telescope (Fig. 1)
    - if all its dependencies are "up-to-date" the target object is returned.
    - if all or some dependencies are not "up-to-date" they will be re-computed on-the-fly, again according to the 'make' metaphor, but with optional, tunable "depth".

All data production in the system follows this schema, whether it concerns an individual user asking for a single special result (target), a calibration scientist determining the instrument behaviour over long periods or a production scientist deriving results for whole nights of observing. They all add knowledge to the system. Note that, in the end, it does not matter very much whether or not a target requested by the user has been processed before, it only affects the performance of the system. This way, the workflow in the ocean of data space is fully driven by the users quest. (see http://process.astro-wise.org/)

This behaviour is achieved by linking (referencing) all dependent data items in the information system. This is achieved by carefully specifying the observation procedure both for science and calibration observations and subsequently designing a data model, chaining the processing targets to the raw data. Next, the data model is converted into an object model, which in turn is ported in a commercial database (using object oriented "user-defined types" in a relational database). The database thus has full awareness of all dependencies. When compute scripts are run at a computer, all Class instantiations are automatically made persistent in the database forming a dynamic archive of all targets and added knowledge, while new and nearly "sacred" raw data is ingested as long as the acquisition continues. Table 1 summarizes some key differences between the Astro-WISE system and the classical approach.

Table 1. Characteristics of the Astro-WISE Information System

"Classical" Paradigm	Astro-WISE Paradigm
forward chaining	backward chaining
"tier" architecture	"target" architecture
driven by raw data	driven by user query
process in pipeline	process on-the-fly
operators push data	users pull data
results in releases	information system
static archive	dynamic archive
raw data is obsolete	raw data is sacred

Obviously, this can only be achieved by integrating all computing and storage hardware into a single information system. Figure 2 shows the peer-to-peer network employed for Astro-WISE. The network and its hardware can be viewed as an extension of the telescope and measurement apparatus hardware, together with the abstract software, forming a unity. The user, by requesting a target, triggers the whole chain from target to raw observational data of the measurement apparatus, a real virtual observatory.

All top level software in the system is written in the Python scripting language (See http://www.python.org/) which in turn calls C libraries. Following the LISP language, Python achieves a very high level of integration. It facilitates another crucial unification in the system: apart from the GUIs, all usage of the system is done from a Python prompt via a Python binding to the database. Novice users, advanced users, and scientific programmers/developers alike use the same environment, allowing a continuous transition between all levels. All Python open source facilities, e.g., visualisation libraries (MatPlotLib), numerical libraries (NumPy), etc., are at direct disposal of the user and are connected to the output of user queries to the information system. Users apply scripts provided by a federated CVS distribution of the standard Astro-WISE code base, but can also modify scripts, fine-tuning them or adding methods as long as they do not violate the object model. This way, both a simple standard version and user tuned versions of processing targets can be obtained simultaneously (different versions of the code are tagged and can be traced).

The database is partitioned in instruments and projects, facilitating individual read and write privileges for persons, groups and projects. Users set a context relating to their project and thus see only their partition in the ocean of data. The smallest project, the individual or anonymous Web based user doing his/her "afternoon" experiment is facilitated in a "MyDB" context, which can be upgraded to another, appropriate context/project at the end of the session. Project leaders have the responsibility for the project quality control. The code base supports this by built-in "verify", "compare" and "inspect" methods for each Class, where "verify" involves an internal code check, "compare" involves a database query to different instantiations of the same Class and "inspect" is a visual inspection by the user. Eventually, all quality control is maintained by a time-stamping mechanism on each Class instantiation, and a GUI is build for calibration scientists to supervise and alter the timestamps, the ultimate point where human insights add knowledge to the system.

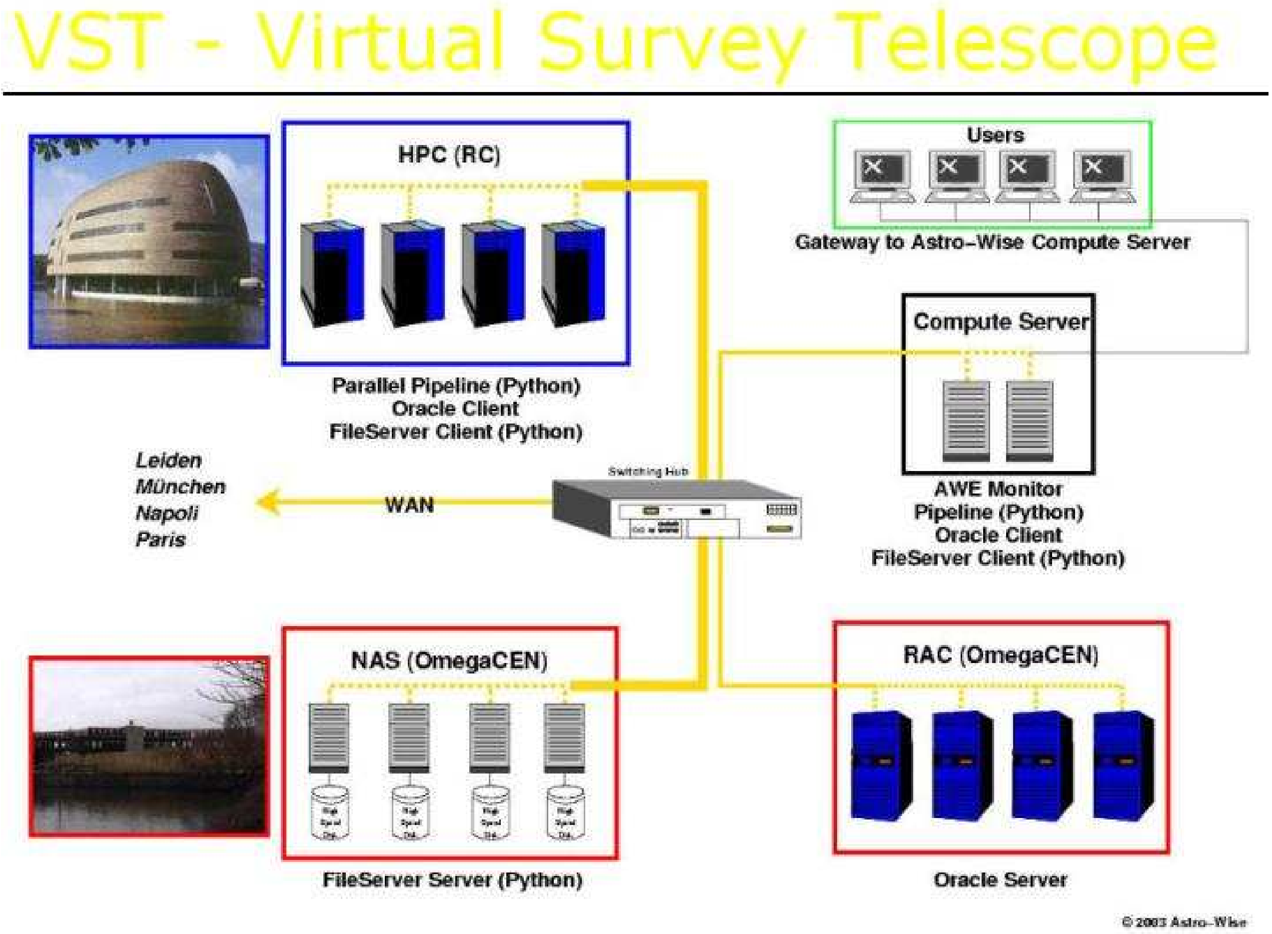


Figure 2. From virtual computer observing to real computer observing. The figure sketches the four fully scalable cornerstones for computer observing in a peer to peer network (with international federations).

In fact, an important design criterion was to allow a complete federation of the system, facilitating different research groups spread over Europe to share scientific projects. Currently, the system connects National data centres in the Netherlands (Groningen), Italy (Naples), France (Paris), and Germany (Munich), which in turn re-distribute the system to satellites (e.g., Bonn and Heidelberg). Similar hardware is installed at the National centres: the code-base is federated, the database is replicated, and the file-servers prevent duplication of massive storage of image data. Network traffic is minimised by allowing for terabyte-sized local cache for frequently used objects. All GUIs are web based. These include the database viewer (http://dbview.astro-wise.org/) and the GUI to evaluate and operate target processing at any selected Linux farm of the federation. User (client) altered Python scripts are sent to the processors in the federation as so-called Python "pickles" (persistent Python objects in a string representation).

The system has been fully implemented and currently operates the raw data of various astronomical wide field imaging cameras at ESO's 2.2m telescope in Chile (300.000 CCD read-outs), the ING telescope(Canary Islands), the Subaru telescope (Japan) and OmegaCAM for ESO's VLT survey telescope (currently test data only). During the implementation, while taking care of object model purity, we experienced an avalanche of benefits for which relatively little extra effort had to be made:

- The full linking of object dependencies allows full backwards chaining as employed by the artificial intelligence (Shachter & Heckerman, 1987 and Thompson & Thompson, 1985) and history tracking. Literally every bit of information that went into the target can be retrieved: the system provides a "tell me everything tool". In fact, this is an a priori ontological implementation, in contrast to the fashionable a posteriori, semantic web search engines.
- Publishing on Internet/EURO-VO (the European Virtual Observatory) is done by raising a flag. It is up to project managers to do this for classical paradigm static results or for Astro-WISE paradigm dynamic results.
- The control by the database over the parallel processing (e.g., SETI@home) on any number of nodes.The enabling of a international compute GRID.
- Enabling global astrometric and photometric solutions with increased accuracy, redirecting global database knowledge back into the system.
- The built-in workflow directly guides the user and no workflow systems are required.

Next to facilitating statistical studies, the system is optimized for "needle-in-the-haystack" kind of searches, finding extremely rare astronomical objects, such as moving, solar system objects, or variable objects like ultra-compact binary systems of white dwarf stars or distant supernovae. Next to all image Meta-data, the astronomical source parameters are stored in an Oracle database and Oracle partitioning is used to quickly address up to 100 Tbyte data volumes. Fast astronomical object associations (a posteriori associations) are made in "many-to-many" mode using both native database indexing and positional HTM[3] indexing. The linking (joins and references) are maintained at an extreme level. For example, in the KIDs 1500-5000 degree survey nearly 1 Tbyte of linked data items is anticipated.

The avalanche of side products and its rapid implementation are thanks to several unifications which are achieved by merging various pointer mechanisms, such as those provided by object oriented programming (inheritance), class/attribute persistency, database internal links (joins, references), and namespace handling by the Python scripting language. In the context of physical information theory, all these links can be viewed as kinds of memory addresses and facilitate forms of information transfer, gearing the user to intrinsic information, and facilitating the handling of the dispersion caused by the measurement apparatus, camera, and the off-line computing hardware. Particularly, the option to inspect partial derivatives of dependent parameter, by re-deriving results, allows the user to inspect the information content in the data (Frieden, 1998).

5. Conclusion & Acknowledgements

Our key-concept, involving novel approaches to maintain data associations in federations of integrated pipelines and archives, can be applied to arbitrary forms of digitized observational data, ranging from

---

[3] http://www.sdss.jhu.edu/htm/

DNA sequences to numerical simulations and national libraries (e.g., centuries of Dutch governmental handwritten records being scanned and entered into the system will be processed with pattern recognition techniques). After all, this is all part of our Universe that we observe, and with the coming decade, will be copied into our hardware at multi-petabyte scales to be interpreted by a next generation of scientists.

The project received funding from the European Union through the EC Action *Enhancing Access to Research Infrastructures*, a FP5 RTD programme, the Netherlands Research School for Astronomy NOVA and NWO. Correspondence and requests for materials should be addressed to E.A. Valentijn, Coordinator, Astro-WISE *valentyn@astro.rug.nl*.
Supplementary Information can be found on www.astro-wise.org

References

Padovani, P., et al. 2006, in ASP Conf. Ser., Vol. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique (San Francisco: ASP), 771

Valentijn , E.A., & Kuijken, K. 2004, Astro-WISE: Toward an International Virtual Observatory, Proceedings, ed. P.J. Quinn, K.M. Gorski, & Springer, 19

Costa, L.d., et al. 2003, in ASP Conf. Ser., Vol. 314, ADASS XIII, ed. F. Ochsenbein, M. Allen, & D. Egret (San Francisco: ASP), 3

Hiriart, R., et al. 2003, in ASP Conf. Ser., Vol. 314, ADASS XIII, ed. F. Ochsenbein, M. Allen, & D. Egret (San Francisco: ASP), 74

Jung, Y., et al. 2003, in ASP Conf. Ser., Vol. 314, ADASS XIII, ed. F. Ochsenbein, M. Allen, & D. Egret (San Francisco: ASP), 764

Mehringer, D.M. & Plante, R.L. 2003, in ASP Conf. Ser., Vol. 314, ADASS XIII, ed. F. Ochsenbein, M. Allen, & D. Egret (San Francisco: ASP), 42

Pierfederici, F., et al. 2003, in ASP Conf. Ser., Vol. 314, ADASS XIII, ed. F. Ochsenbein, M. Allen, & D. Egret (San Francisco: ASP), 476

Swade, D.A., Hopkins, E., & Swam, M.S. 2001, in ASP Conf. Ser., Vol. 238, ADASS X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne (San Francisco: ASP), 295

Kuijken, K., et al. 2002, ESO Messenger, 110, 15

Kuijken, K., et al. 2004, SPIE, 5492, 484

Shachter, R.D. & Heckerman 1987, *Thinking Backward for Knowledge Acquisition*, AI Magazine, (8,3), 55

Thompson, B.A. & Thompson, W.A. 1985, *Inside an Expert System*, Byte, 10, 315

Frieden, B.R. 1998, *Physics from Fisher Information*, Cambridge University Press

Astro-WISE European programme 2008 and beyond

The EU **FP5**-RTD project Astro-WISE delivered an EU wide federated information system, facilitating distributed production and research analysis of astronomical optical wide field imaging surveys. The system connects, *in real time*, distributed databases and processing grids at national and satellite datacenters, where operators and researchers *produce and analyze* petabyte of observational data for a variety of astronomical optical imaging projects. The entire information system has been built and is distributed over the involved National nodes and satellites.

The objectives for the years to come are i) to further deploy the infrastructure for new very large astronomical and multidisciplinary surveys involving at least 3 petabyte of raw observational data acquired during the equivalent of about 3000 nights of observing, ii) to involve and coordinate new communities, iii) to achieve the massive ingestion and warehousing of data in the system, iv) to facilitate top scientific research and v) to further develop and disseminate the system for multi-disciplinary use.

Astro-WISE supports a set of top scientific objectives for which national funding is available.
The science goals reflect the most compelling issues in their fields: dark matter, dark energy, the origin of galaxies and their mysterious morphological transformations, the uniqueness of our solar planetary system, and the usage of artificial intelligence to decipher unprecedented volumes of historical documents. All the information and knowledge created by the programme is open for public research, and numerous smaller research projects plan to use the system. In fact, 19 consortium members are already committed to use the system, in turn servicing users at 44 additional institutes.

The geographical distribution of the information is provided by the system itself, which allows Europe-wide distributed research groups to work cooperatively while optimally profiting from the work done elsewhere. So, in addition to very innovative information technology, the programme facilitates collaborative astronomical research.

Deploying Astro-WISE for Europe: collaborations, nodes, surveys and science

Currently, national datacenters in the Netherlands (Groningen), Germany (Garching), France (Paris) and Italy (Napoli) and satellites at Leiden (NL), Nijmegen (NL) and Bonn (D) support the Astro-WISE e-Infrastructure. In the near future Astro-WISE plans to deploy the system at new satellite nodes.

Astro-WISE's infrastructure is unique in connecting databases that contain images, meta data, 'event data' and links, to software pipelines. Astro-WISE developed a novel extreme data lineage paradigm: 'full backward chaining', facilitating 'target processing'. Thanks to this achievement, all data items from raw to results are connected, facilitating quality control during production as well as critical research and posteriori inspections at any desired level. Astro-WISE's e-science infrastructure is fully scalable by design, and hence it can handle very large (petabyte) data sets and compute grids in a distributed community. The huge data volumes require an innovative dynamical approach, in which results can be re-derived, customized to each user's specific needs. To this end, the various calibration data and other input files are distributed over a network, which connects the data centres. The system itself provides the geographical distribution of the key information (both methods and data). This is particularly useful for the production of **large surveys**, because it allows scientists at a number of data centres to work cooperatively in an efficient manner and to optimally benefit from the work done at each of the centres. Smaller projects can profit as well, for example using the calibration parameters derived by the large surveys and applying them to their own data, irrespective of where they were derived.

The following tables list the planned satellites and new communities which plan to connect to the system.

Satellite nodes, via National nodes
Consortium Members

Institute	Location
Astronomy Dept, Radboud University	Nijmegen, NL
Leiden Observatory, Leiden University	Leiden, NL
Argelander-Institut für Astronomie, Bonn University	Bonn, D
Astronomical Institute, Ruhr-Universität Bochum	Bochum, D
Max Planck Institute for Astronomy	Heidelberg, D
INAF-Astronomical Observatory of Padova	Padova, I

Deployment of new communities - Consortium members

Institute	Location
Pontificia Universidad Catolica de Chile	Santiago, Chile
Laboratoire d'Astrophysique	Marseille, France
Niels Bohr Institute, University of Copenhagen	Copenhagen, Denmark
Institute for Computational Cosmology, Durham University	Durham, UK
Instituto de Astrofísica de Canarias	Canary Islands, Spain
Artificial Intelligence Institute at the University Groningen	Groningen, The Netherlands
Radio Observatory- ASTRON	Dwingeloo, NL

The system is ready for deriving, hosting, and warehousing the data of new generations of massive data accumulators and is scalable up to the petabyte data regime. Extensive web services have been put into operations to facilitate the users of the system.
While Astro-WISE was originally conceived to support modern optical wide field imaging experiments such as the VLT Survey Telescope (VST) of the European Organization for Astronomical Research (ESO), the system has attracted additional communities and experiments that wish to propel their research using this infrastructure:

- ESO decided to initiate large public imaging surveys for its two new telescopes at Paranal (Chile), the 2.6 meter VST and the 4 meter Visual Infrared Telescope (VISTA), by launching a peer reviewed announcement of opportunity for the majority of the available time at the telescopes. The production, processing and analyses of these large surveys will be handled by selected international consortia, which in turn deliver the results to ESO. Most of the big VST public surveys and some of the VISTA public surveys intend to use the Astro-WISE infrastructure for their work.
- The radio astronomical community is building a new Low Frequency Radio telescope (LOFAR) scheduled to be fully operational in 2009. LOFAR as an instrument will have a data rate of 10 terabyte/day which will be handled at the RUG-CIT by a dedicated central processing system, consisting of a 12,288-processor IBM BlueGene/L system together with a large Linux cluster. The plan is to ingest the outcome of the initial processing for two projects, the LOFAR Surveys and the Epoch of Reionisation, into the Astro-WISE system, thus connecting the LOFAR machinery to the end researcher.
- The Coma Cluster Legacy Survey, a large imaging project with the Hubble Space Telescope using the Advanced Camera for Surveys (ACS), is using Astro-WISE for data dissemination and detailed galaxy structure parameter derivation and analysis.
- The European partner (Max-Planck-Institut fuer extraterrestrische Physik - MPG) for the Panoramic Survey Telescope & Rapid Response System (Pan-STARRS) plans to use Astro-WISE for the production and eventual dissemination of the huge wide field imaging survey carried out with the Pan-STARRS telescope 1 (PS1) at the University of Hawaii's Institute for Astronomy. The data when delivered through Astro-WISE will serve a wide range of science goals.
- Several other wide field imaging facilities, such as ESO's Wide Field Imager at the 2.2m telescope at La Silla, Chile (WFI@2.2m), the wide field imager (WFC) at the UK-NL-Spain Observatorio del Roque de los Muchachos (Canarias, La Palma), the Suprime-Cam at the Japanese 8 meter Subaru telescope are supported by the system and attract new users who wish to reduce or analyse the enormous data archives.
- The artificial intelligence research group at the University of Groningen, together with the Dutch National Archive (Cultural Heritage Collections group) plans to use Astro-WISE to access kilometers of paper archives (30 terabyte of scanned pages) by means of intelligent computing. Handwriting recognition code is run on the IBM Blue Gene supercomputer to interpret the archives of the Kabinet van de[n] Koning[in] (Cabinet of the Queen), next to intense usage of the LINUX cluster. Similarities with astronomical survey IT issues are remarkable.

For all these projects, pioneer studies have taken place during the last year ensuring the technical feasibility of operating these projects in the Astro-WISE information system.
Astro-WISE's way of addressing the data reduction is to calibrate instruments over given periods of time, rather than calibrating individual data items. The following table lists the planned support of observational infrastructures, while in the next section we summarize the key science cases and individual projects and surveys which plan to use the Astro-WISE system for the production, calibration and science analysis.

Observational science infrastructures supported by Astro-WISE

Discipline	**Supported infrastructures in Astro-WISE**		
	Acronym	**Description**	**Level**
Ground-based optical astronomy	OmegaCAM@VST	OmegaCAM imager at the ESO VLT Survey Telescope, Chile	raw
	WFI@2.2m	ESO Wide Field Imager at 2.2m telescope, Chile	raw
	SuprimeCam@Subaru	SuprimeCam Camera at Japanese Subaru- 8 m telescope	raw
	Pan-STARRS	Panoramic Survey Telescope and Rapid Response System, USA	image
	WFC@INT	Wide Field Camera, Isaac Newton Telescope, Spain	raw
	SDSS	Sloan Digital Sky Survey catalogs, USA	image
	ESO-LV	ESO Lauberts & Valentijn Catalogue, ESO	image
	LBT	Large Binocular Telescope, USA	raw
Ground-based infrared astronomy	VISTA	ESO Visible & Infrared Survey Telescope for Astronomy ,Chile	raw + image
	UKIDSS	UK Infrared Digital Sky Survey, UK	image
Ground-based radio astronomy	LOFAR	Low Frequency Array, The Netherlands	image
	WENSS	WSRT Northern Sky Survey, The Netherlands	image
Space-based optical astronomy	ACS@HST	Advanced Camera for Surveys on board the Hubble Space Telescope, USA	image
	GAIA	Candidate for further deployment, ESA	-
Cultural Heritage	SCROOGLE@ASTROWISE	‘Script Google’ for handwritten archives	raw

Science cases for Astro-WISE

Tracing dark matter and dark energy

(KIDS, VIKING, Pan-STARRS-3π)

Since the cosmological supernova survey results in 1998 it has become clear that about 95% of the energy in the current day Universe resides in an invisible form. About 70% is thought to be dark energy and 25% dark matter. The nature of dark matter remains a mystery. Its distribution can be simulated theoretically by means of large computer calculations, but the results can only be tested indirectly, by comparing to the distribution of galaxies. The nature of dark energy and how its density evolves as a function of cosmological time is also unknown. Together they form major open issues in astrophysics. Dark matter and energy can be traced via multi-band imaging of large parts of the sky. The dark matter distribution is traced via its gravitational lensing of light: the bending of light through gravity. Large area optical surveys can map via lensing in exquisite detail the distribution of dark matter around galaxies and galaxy clusters, at radii beyond those that can be probed with the traditional dynamical measurements where the model predictions are clearest. Together with the redshifts of the galaxies (determined from multi-band photometry) the lensing measurements trace the growth of structure with redshift (i.e., time), and provide a powerful probe of the expansion history of the universe and hence of the dark energy as a function of cosmological time.

The deep Kilo Degree Survey (KIDS) + VISTA Kilo-degree INfrared Galaxy survey (VIKING) combination is specifically designed to trace both dark matter and dark energy, while the Pan-STARRS 3π steradian Survey (Pan-STARRS-3π) will address mainly dark energy. The surveys will result in the detection of hundreds of millions of galaxies out to redshift ~1 and hundreds of terabytes of science and calibration and data.

Galaxy evolution - finding and tracing morphological transitions

(KIDS, VIKING, VESUVIO, VST16, Pan-STARRS-MDS, Coma-LS and SUDARE)

The galaxies of the Universe reside in clusters of galaxies connected by filaments --made up of loose groups of galaxies-- together forming a ‘cosmic web’ around large voids which are practically devoid of galaxies. Galaxies flow from the low density regions, the ‘field’, to the high density regions of massive rich clusters of galaxies attracted by the gravitational pull.

As they flow, an extremely complex interplay occurs between galaxies themselves (i.e., galaxy mergers) and galaxies and their gaseous intra-galactic medium. These processes transform the galaxies in terms of morphology and stellar populations, due to the gravitational / hydrodynamical interactions and the bursts of star formation that they trigger.

Hence, the galaxy properties and demography of the filaments, field and cluster differ. For example, low density regions are dominated by galaxies with large disks, whereas the high density regions of clusters are dominated by spheroidal galaxies. Somehow the disks disappear or are destroyed in high density regions and result in a build-up of spheroidal galaxies up to stellar masses much larger than those observed in the low-density field. Thanks to the development of the new generation of wide-field imagers it is for the first time possible to follow and thus understand this process of transformation by mapping at high spatial resolution the large areas spanned by the voids and outskirts of clusters all the way to central cluster cores. KIDS, VIKING, the VST/OmegaCAM Exploration of SUpercluster Voids and Intermediate Objects (VESUVIO), VST16 and the Pan-STARRS Medium Deep Survey (Pan-STARRS-MDS) are such large area surveys. There are two complementary approaches. First, the optical imaging of dense cluster regions at uniquely extremely high spatial resolution such as the Hubble Space Telescope Coma Cluster Legacy Survey (Coma-LS). Second, to measure the occurrence rate of supernovae which are the end products of stellar evolution and hence this rate traces the rate at which galaxies formed stars. The program “Supernova Diversity and Rate Evolution” (SUDARE) will map the supernovae rate at a redshift ~0.5.

Science Targets supported by Astro-WISE

Science Targets	**Science Target projects propelled by Astro-WISE**			
	Acronym	**Description**	**Allocated Observing time (nights)**	**Total data volume input/processed (Tb)**
1.Tracing dark matter and dark energy	KIDS	KIlo Degree Survey	500	30/45
	VIKING	VISTA Kilo-degree INfrared Galaxy survey	250	60/45
	Pan-STARRS-3π	Pan-STARRS 3π steradian Survey	690	4200/4300
2.Galaxy evolution-finding and tracing morphological transitions	VESUVIO	VST/OmegaCAM Exploration of SUpercluster Voids and Intermediate Objects	150	10/15
	KIDS	KIlo Degree Survey	500	30/45
	VIKING	VISTA Kilo-degree INfrared Galaxy survey	250	60/45
	Pan-STARRS-MDS	Pan-STARRS Medium Deep Survey	345	2100/2150
	COMA-LS	Coma Cluster Legacy Survey	150[a]	0.1/0.2
	VST16	16 band wide-field survey	65	3.5/5.5
	SUDARE	Supernova Diversity and Rate Evolution	24	2/3
3.Very Early Universe: finding the first galaxies and the first light	ULTRAVISTA	Ultra-deep survey with VISTA	200	3/4
	LOFAR-EoR	Epoch of Reionisation with LOFAR	120	1000/1100
	LOFAR-Survey	LOFAR Surveys of high-redshift galaxies	300	224/230
4: The history of our Milky Way	STREGA	Structure and Evolution of the Galaxy	45	3.5/5.5
	OmegaWhite	Evolution of very short period white dwarf binaries	35	3.0/5.0
5.Detecting extrasolar planets	OmegaTranS	OmegaCAM Transit Survey	35	15/30
	PanPlanet	Pan-STARRS Transit Survey	50	20/40
6. Datamining handwritten cultural heritage	PUBLIC SCROOGLE	Public access to pattern recognition results of hand-written Dutch National archives	-	30/10

[a]) in units of number of orbits

Very Early Universe: finding the first galaxies and the first light

(ULTRAVISTA, LOFAR Survey and EoR)

It has become clear in recent years that the very early Universe at a redshift z>10 did not yet contain stars and galaxies but consisted of mostly neutral gas. We now stand at the final frontier in our quest to understand the end of these ‘Dark Ages’ and the formation of the first galaxies. The current situation is tantalizing; we know of several hundred galaxies up to redshift z=6 but there currently exist only three credible galaxy candidates at z > 6.5. We know so little about the crucial epoch 6 < z < 10 because selection via optical imaging can only be achieved out to z=6. To probe to higher redshifts requires near-infrared imaging of sufficient area and depth. The large field of view in combination with the sensitivity of the near-infrared VISTA telescope is unique in the world. Science operations for this telescope will start early 2008 and the facility will stay unique until ~2015. The ULTRAVISTA project will exploit this capability with an ultra-deep imaging survey of the ‘COSMOS field’.

Thanks to its unprecedented sensitivity, two projects with the radio telescopes of the Low Frequency Array (LOFAR) offer alternative radio approaches to explore the end of the Dark Ages. *LOFAR surveys* will detect extremely distant galaxies with strong radio emission, the radio galaxies, using an empirical correlation between radio spectral steepness and distance. Only LOFAR has the sensitivity and field of view to efficiently pick out radio galaxies at redshifts z> 6, which is not possible with present radio telescopes. This first time inventory of extremely high redshift radio galaxies will (i) constrain the formation of supermassive black holes (radio emission is powered by those black holes), (ii) obtain detailed view of the interstellar medium, the fuel of star formation, which shows up as absorption of the radio emission and (iii) identify proto-clusters, as distant radio galaxies have been shown to pinpoint those.

The *LOFAR Epoch of Reionisation (EoR) project* will detect directly the extremely subtle radio signal (the redshifted 21cm line emission) from gas around the EoR which ended the Dark Ages. Research in the last few years suggests that there may have been extended, or even multiple phases of reionisation, the start possibly being around z~10-20 and ending at z~6. Using LOFAR, the redshift range from z=11.4-6 can be probed for the 21cm line emission. It is with LOFAR that one can for the first time determine the redshift range in which the bulk of the neutral hydrogen became ionized and if it happened in a single transition phase or through multiple phases of reionisation. Comparing simulations from the BlueGene supercomputer to the exquisite details of the LOFAR data, the project can constrain even which objects (very early ‘Population III’ stars, galaxies, quasars) or processes were responsible for reionizing the Universe.

The history of our Milky Way

(STREGA, OmegaWHITE)

Evidence has emerged in recent years that the Milky Way and its surrounding galaxies have experienced a complex history of interactions and mergers. The next step in this field is to map large areas of the Milky Way and its surroundings. Such observations yield also a uniquely detailed view on millions of stars yielding to detect statistically meaningful samples of poorly understood rare stellar populations and very short-lived evolutionary stages which are crucial in our understanding of stellar evolution.

At OmegaCAM@VST, the program “Structure and Evolution of the Galaxy” (STREGA) will image 150 square degrees chosen to survey the suspected interactions between the Milky Way and the Fornax and Sculptor dwarf galaxies and the Pal3 and Pal12 globular clusters.
The imaging survey OmegaWHITE will monitor at least 100 square degrees to find periodic photometric variations that are the fingerprint of ultracompact binaries. The binary stars are so close to each other that gas transfer has a strong impact on the evolution of the stars in a very complex way. Thus, these systems are unique laboratories for the evolution of all binaries which make up ~50% of all Milky Way stars. Furthermore, understanding the life of ultracompact binary stars is crucial for understanding gravitational waves (ultracompact binaries are the strongest

sources of gravitational waves) and the progenitors of Supernova Type Ia used to trace the acceleration of the Universe.
OmegaWHITE expects to return a factor ~7 more ultracompact binaries than currently known.

Detecting extrasolar planets

(OmegaTranS, PanPlanet)

The search for extra-solar planets is one of the newest, most exciting and fastest developing fields of astronomy. The ultimate goal is the detection of habitable planets, to assess how common life is in the Universe. When a planet moves in front of its host star it covers a small part of its surface. Monitoring the star's brightness one can see the resulting periodical drop of 1-2% in the light-curve for several hours. To find planets using this ‘transit’ method is a huge observational challenge. One limitation is the small chance to find a planet's orbit edge-on. In addition one needs to monitor the target stars for a reasonable amount of time in order to observe multiple transits and to be able to determine accurate periods. For a transit survey to be successful ten thousands of stars have to be observed more than a thousand times in order to detect a few planets. The new OmegaCAM@VST and Pan-STARRS telescopes are ideal to survey vast numbers of stars for transiting planets: per square degree, several 100.000s of stars can be monitored simultaneously. OmegaTranS - the OmegaCam Transit Survey- and PanPlanet will observe several fields close to the Galactic plane hundreds of times per year to catch transits. The number of new transiting planets detected per year for OmegaTranS (~10) and PanPlanet (~30) will increase within a few years many times the number of currently known transiting planets.

Datamining handwritten cultural heritage

(PUBLICSCROOGLE)

During the last two years, fast progress has been made in the field of data mining handwritten collections via projects running at the Artificial Intelligence (AI) department of Groningen University. The availability of high-performance computing facilities has made it possible to explore new concepts in machine learning and pattern recognition. Combining brute force (correlator) matching with traditional machine learning has proven to be capable to open up the access to large collections of scans of handwritten text via continuous web-based transcription of text, concurrent with nightly retraining computation sessions. This requires (1) a continuous availability of high-performance computing; (2) persistent storage of valuable cultural-heritage images for several decades; and (3) a fusion of modern web-based data access for both cultural-heritage researchers and computer scientists working on pattern recognition and machine-learning problems. A pilot experiment by the AI group and OmegaCEN using the Astro-WISE system has demonstrated that much is to be gained from joint access to a central data-operations and science-operations center.
It is essential that the underlying architecture allows general-public browser access to portions of the data on the one hand, and access by supercomputing facilities of the same data on the other hand. Having scanned one book, i.e., the year 1903 of the 'Cabinet of the Queen' at the Dutch National Archives, it now becomes a scientific and technical challenge to scale up towards providing access to the full range of years from 1901 to 1915. Continuous, persistent and open access to huge collections of handwritten manuscripts can be achieved using the Astro-WISE system. The software tools needed for this non-astronomical application will be developed by the Artificial Intelligence group at Groningen University in collaboration with the Astro-WISE consortium and the Groningen University Computing Centre.

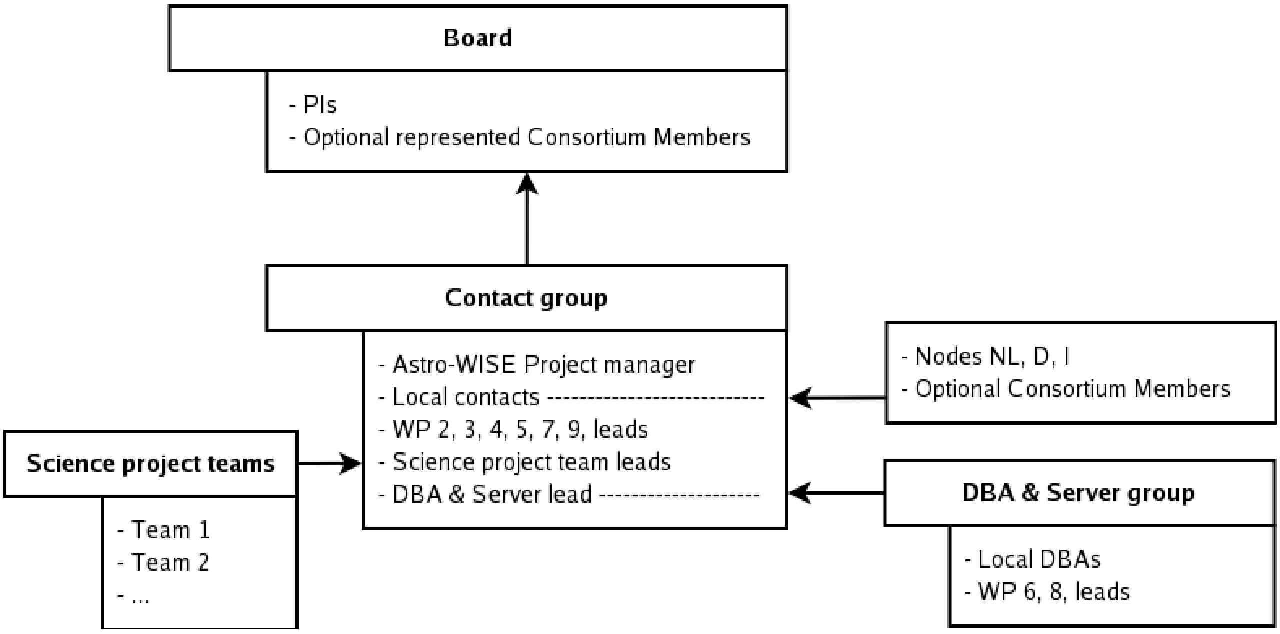
Astro-WISE services for Science projects

Science Target Project	Infrastructure acronym	Participants	Consortium Members	Customers/ Scientific Target partners
Scientific target 1: Tracing Dark matter / Dark energy in the Universe				
KIDS	OmegaCAM @VST	NOVA MPG INAF	-University of Bonn, D -University of Bochum, D -Astrophysical Institute Paris, IAP, F	-Institute of Astrophysics, Cambridge, UK -University of Edinburgh, Institute for Astronomy, UK
VIKING	VISTA	NOVA MPG INAF	-University of Bonn, D -University of Bochum, D -Astrophysical Institute Paris, F	-Institute of Astrophysics, Cambridge, UK -Queen Mary University London, UK -University of Edinburgh, Institute for Astronomy -Phys/ Dept, Bristol University
Pan-STARRS - 3π	Pan-STARRS	MPG	-Durham University, Extragalactic Astronomy & Cosmology Research Group UK -Max Planck Institute for Astronomy, Heidelberg, Germany	-University of Hawaii, Institute for Astronomy, USA -Johns Hopkins University, USA -Las Cumbres Observatory, USA -Harvard-Smithsonian Center for Astrophysics, USA -Queen's University, Astrophysics Research Center, Ireland -National Central University, Taiwan -University of Edinburgh, Institute for Astrophysics, UK
Scientific Target 2: Galaxy evolution: the origin of species of galaxies				
KIDS	See description under Science Target 1			
VIKING	See description under Science Target 1			
VESUVIO	OmegaCAM @VST	NOVA MPG INAF		-Netherlands Organization for Space Research- SRON , NL -Institute of Astrophysics, Innsbruck, Austria -University of Tasmania, Australia National Observatory, Athens, Greece -University of North Carolina, USA
Pan-STARRS MDS	Pan-STARRS	MPG	-Max Planck Institute for Astronomy, Heidelberg, Germany -Durham University, Extragalactic Astronomy & Cosmology Research Group UK	-University of Hawaii, Institute for Astronomy, USA -Johns Hopkins University, USA -Las Cumbres Observatory, USA -Harvard-Smithsonian Center for Astrophysics, USA -Queen's University, Astrophysics Research Center,

				Ireland -National Central University, Taiwan -University of Edinburgh, Institute for Astrophysics, UK
VST16	OmegaCAM @VST	INAF	-Max Planck Institute for Astronomy, Heidelberg, Germany -Argelander-Institut fur Astromie, Bonn University, Germany	-University of Naples, Italy -University of Nottingham, UK -Institute for Astrophysics, Innsbruck, Austria -University of Edinburgh, Institute for Astrophysics, UK -Astrophysical Institute, Potsdam, Germany
Coma-LS	ACS@HST	NOVA MPG	-Durham University, Extragalactic Astronomy & Cosmology Research Group, UK -Instituto de Astrofisica de Canarias, Spain	-Liverpool John Moores University, UK -Space Telescope Science Institute, USA -Herzberg Institute of Astrophysics, Canada -Institute of Astrophysics Cambridge, UK -Swinburne University, Australia -University of Florida, USA -Rochester Institute of Technology, USA -University of Texas, USA -University of Tokyo, Japan -Subaru Observatory, Japan/USA -Cardiff University, UK -University of Bristol, UK -Osservatorio di Padova, I -Queens University, Canada -University of Waterloo, Canada -Gemini Observatory, Chile -National Optical Astronomy Observatory, USA -NASA Goddard, USA -Johns Hopkins University, USA -University of Hawaii, Institute for Astronomy, Hawaii, USA -San Francisco State University, USA
SUDARE	OmegaCAM @VST	INAF	-INAF-OAP, Padova, Italy	-Institute of Astrophysics Cambridge, UK -INAF-OAT, Teramo, Italy

Scientific Target 3: Very Early Universe: formation of its first structures				
ULTRAVISTA	VISTA	NOVA	-Leiden Observatory, University of Leiden, NL -Niels Bohr Institute, University of Copenhagen, DK -Laboratoire d'Astrophysique de Marseille, F	-University of Edinburgh, Institute for Astronomy, UK
LOFAR-EoR	LOFAR	NOVA	-ASTRON, Dwingeloo, NL -Kapteyn Astronomical Institute, University of Groningen	-Max Planck Institute for Astrophysics- MPA, Garching bei Munchen, D
Scientific Target 4: Radio sky mapping				
LOFAR-Surveys	LOFAR	NOVA	-Leiden Observatory, University of Leiden, NL -ASTRON, Dwingeloo, NL	-University of Edinburgh, Institute for Astronomy, UK -Max-Planck-Institut für Radioastronomie, Bonn, Germany -Max Planck Institute for Astrophysics- MPA, Garching bei Munchen, D
Scientific Target 5: History of the Milky Way				
OmegaWHITE	OmegaCAM @VST	NOVA	-Astronomy Department, Radboud University of Nijmegen, NL	-Nordic Optical Telescope, Spain -University of Sheffield, UK -University of Warwick, UK -Harvard-Smithsonian Center for Astrophysics, USA
STREGA	OmegaCAM @VST	INAF		-INAF-OAR, Roma, Italy -INAF-OAT, Teramo, Italy -Pisa University, Italy -University of Bologna, Italy -University of Edinburgh, Institute for Astrophysics, UK
Scientific Target 6: Detection of planets outside Solar System				
OmegaTRANS	OmegaCAM @VST	MPG INAF	-Leiden Observatory, University of Leiden, NL	
PanPlanet	Pan-STARRS	MPG	-Max Planck Institute for Astronomy, Heidelberg, Germany	-University of Hawaii, Institute for Astronomy, USA -Johns Hopkins University, USA -Harvard-Smithsonian Center for Astrophysics, USA -University of Edinburgh, Institute for Astrophysics, UK
Scientific Target 7: Cultural Heritage Analysis				
PUBLIC SCROOGL	SCROOGLE@ASTROWISE	RUG-CIT NOVA	-Dept Artificial Intelligence, University of Groningen	- Nationaal Archief, The Hague

Organizational structure of Astro-WISE



Astro-WISE's organization connects in a simple way the three basic levels at which institutes are involved:

i) Consortium members who operate national data centers and run various Astro-WISE servers (all represented by PIs in the board), currently OmegaCEN/NOVA, CIT/University of Groningen, MPE/MPG, OAC/INAF, Terapix/IAP and ESO.

ii) Consortium members who operate satellite nodes and/or actively participate in the development and employment of the system (optionally represented in the board).

iii) Registered user communities who participate in one of AstroWISE's enabled surveys or projects.

Yearly workshops with the Consortium and users are held at the Lorentz Center in Leiden. Presentations given at the workshops can be viewed at www.astro-wise.org/agenda.shtml.

Five annual reports of Astro-WISE are published at www.astro-wise.org/doc_nova.shtml.

First "Science exploitation" papers using Astro-WISE

D. Carter et al. 2007, The HST/ACS Coma Cluster Survey I- submitted to Ap. J.

Heraudeau, Ph. and Valentijn, E.A., 2005, An optical survey of the ELAIS-S2 field. Data reduction with Astro-Wise *SF2A-2005:* Semaine de l'Astrophysique Francaise, 717

Kovac, 2007, *Ph D Thesis*, University Groningen, Searching for the lowest mass galaxies- Optical and HI properties of galaxies detected in the WSRT CVn survey

Verdoes Kleijn, G., et al. 2007, The Secondary Standards Programme for OmegaCAM at the VST in "The Future of Photometric, Spectrophotometric, and polarimetric standardization", *ASP* Edited by C. Sterken. San Francisco: Astronomical Society of the Pacific, p.103

Sikkema et al., 2006 , Globular cluster systems of six shell galaxies Astron. & Astrophys. *458, 53*

Sikkema, G., Carter, D., Peletier, R. F., Balcells, M., Del Burgo, C., Valentijn, E. A., 2007 HST/ACS observations of shell galaxies: inner shells, shell colours and dust. Astron & Astrophys 467, 1011S,

Zant, van der, T., Schomaker, L. and Valentijn, E.A., 2008, Large Scale Parallel Document Image Processing in procs of the conference "Document Recognition and Retrieval XV", accepted

Dr J. Wallage, mayor of the city of Groningen has an Astro-WISE demonstration at OmegaCEN, Kapteyn Institute, July 2006

Partners and participating funding agencies in the Netherlands

ASTRON, Netherlands foundation for research in astronomy, Dwingeloo
Astronomy Department, Radboud University of Nijmegen
Artificial Intelligence Department, University of Groningen
ATOS-Origin, Groningen
Center for Mathematics and Computer Science, CWI, Amsterdam
Donald Smit Center of Information Technology, University of Groningen
Dep. of Mathematics and Computing Science, University of Groningen
Kapteyn Astronomical Institute, University of Groningen
Leiden Observatory, University of Leiden
National Research school for Astronomy, NOVA
Netherlands Organization for Scientific Research, NWO, The Hague
OmegaCEN, Kapteyn Institute, University of Groningen
Oracle Nederland BV
Samenwerkingsverband Noord Nederland, SNN, Groningen
University of Groningen