Future Generation Computer Systems 27 (2011) 319–328

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

LOFAR Information System

K. Begeman [a], A.N. Belikov [a,*], D.R. Boxhoorn [a], F. Dijkstra [b], H. Holties [c], Z. Meyer-Zhao [b], G.A. Renting [c], E.A. Valentijn [a], W.-J. Vriend [a]

[a] Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747AB, Groningen, The Netherlands

[b] Donald Smits Center for Information Technology, University of Groningen, Nettelbosje 1, 9747AJ, Groningen, The Netherlands

[c] ASTRON, P.O. Box 2, 9900AA, Dwingeloo, The Netherlands

ARTICLE INFO

Article history:
Received 28 December 2009
Received in revised form
30 July 2010
Accepted 16 August 2010
Available online 21 August 2010

Keywords:
Information system
Grid
Scientific data processing

ABSTRACT

In this paper, we present a newly designed and implemented scientific information system for the LOFAR Long Term Archive. It is a distributed multi-tier storage and data processing system that allows a number of users to process Petabytes of data.

The LOFAR Information System is designed on the base of the Astro-WISE Information System. The system allows the use of the computing and storage resources of the BiGGrid combined with the metadata database along with the data access and data processing interfaces of Astro-WISE.

The architecture of the system, problems solved during the implementation and scientific use cases for the system are also described.

1. Introduction

Requirements from scientific communities nowadays put a number of challenges on scientific information systems. The origin of these challenges is the nature of the scientific community itself. Today, data processing for a single scientific case requires efforts from a number of scientists, programmers and engineers from many institutions spread over the globe. The main challenge is not the data volume itself, but an optimal integration of users, workflows and pipelines, and various computational and data storage resources. The scientific information system should be robust, integrate geographically distributed resources, and be federated in the sense of data storage and data processing facilities. Most importantly, the system should not simply store data, but be an integrated environment for data processing from the raw scientific data up to the science-ready end product. The system should allow sharing of the data between scientists to speed up data processing and should provide full data lineage and data provenance.

It is not enough just to handle Petabytes of data and provide fast access to each data item. The system should allow intensive data mining on the data without retrieval of data items themselves. Moreover, some science cases require access only to the metadata without actual retrieval of processed data.

In this paper we will introduce the LOFAR Information System which deals with the problems described above. The proposed solution is based on the existing data storage and processing system for wide-field astronomy (Astro-WISE[1] [1]) and extending this system for LOFAR [2].

The practical task for the developers of the LOFAR Information System is to combine three systems in one: Astro-WISE metadata storage and data processing, EGEE/BiGGrid data storage and computing nodes and the LOFAR project administration.

The development of the LOFAR Information System started in 2008 and is in its final stage. We will present its system components and functionalities, and discuss how the system fits the requirements of LOFAR and how it can be used for other scientific applications. We will start with the review of data storage and processing challenges of LOFAR in Section 2, and continue with the description of the implementation of the LOFAR Information System in Section 3. We will briefly review science data processing interfaces in Section 4 and science use cases for the system in Section 5. Section 6 gives the current status of the system and future development.

2. LOFAR

LOFAR,[2] the Low Frequency Array, is building a huge radio interferometer. The first building phase of LOFAR will deliver a

* Corresponding author. Tel.: +31 503633476.
E-mail addresses: kgb@astro.rug.nl (K. Begeman), belikov@astro.rug.nl (A.N. Belikov), danny@astro.rug.nl (D.R. Boxhoorn), f.dijkstra@rug.nl (F. Dijkstra), holties@astron.nl (H. Holties), Z.Zhao@rug.nl (Z. Meyer-Zhao), renting@astron.nl (G.A. Renting), valentyn@astro.rug.nl (E.A. Valentijn), wjvriend@astro.rug.nl (W.-J. Vriend).

1 http://www.astro-wise.org.
2 http://www.lofar.org.

0167-739X/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.future.2010.08.010

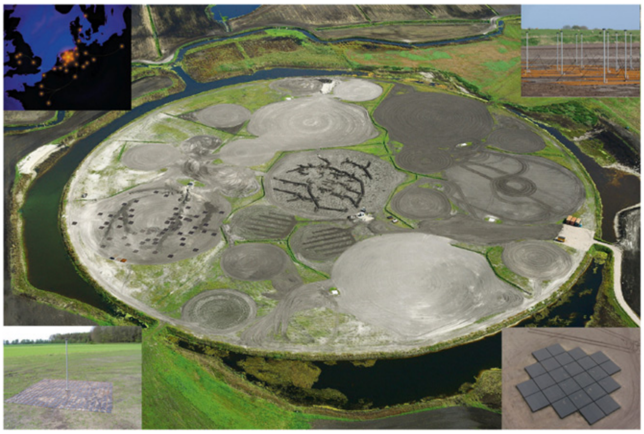


Fig. 1. The core of the LOFAR telescope (central photo) with low-frequency and high frequency antennas (lower corner pictures) and low-frequency antenna field (upper right corner). The final configuration of LOFAR antenna fields in the full deployment mode is shown in the picture in the left upper corner (with the core in the center and Dutch national and international stations).

compact core area (approx. 3 km in diameter, 18 stations) as well as 18 remote Dutch stations and 8 International stations. Each station will be equipped with up to 96 high band antennas and 96 low band antennas (see Fig. 1). Each station is connected to the Central Processing System (CEP) through a Wide Area Network (WAN) using a 10 GB ethernet. The LOFAR instrument performs an aperture synthesis at the Central Processing System using signals from each station. CEP consists of an IBM supercomputer of the type BlueGene/P, and a linux based cluster at the University of Groningen.

Thanks to its unprecedented sensitivity, Lofarprojects can explore the early stages of the Universe. For example, the Lofar Surveys project[3] will identify extremely distant radio galaxies using an empirical correlation between the radio spectral steepness and distance. Only Lofar has the sensitivity and field of view to efficiently pick out radio galaxies at redshifts $z > 6$. This first time inventory of extremely high redshift radio galaxies will:

- constrain the formation of supermassive black holes (the radio emission is powered by those black holes),
- yield a detailed view of the interstellar medium (the fuel of star formation which shows up as absorption of the radio emission),
- identify proto-clusters, as distant radio galaxies have been shown to pinpoint these.

Other areas of astronomical research that will benefit from the capabilities of the LOFAR instrument include pulsar and other transient phenomena, cosmic rays, solar science/space weather and cosmic magnetism. The Lofar Information System will deal with the data delivered by the Central Processing System and stored in the Long Term Archive (LTA). We will now review the LOFAR data processing model and, in the next section, the design for the Lofar Information System based on the Astro-WISE Information System.

2.1. Lofar data processing

The LOFAR architecture can be described as consisting of the following components (see [3,4]):

1. Stations. A station selects the sky signals of interest for a particular observation out of the total sky. This process results in one or multiple station beams onto the sky. Additionally the station is able to store (raw) antenna data.
2. Wide Area Network (WAN). The WAN is responsible for the transparent transport of all the beam data (the beam signals on the sky) from the stations to the Central Processing Unit.
3. CEntral Processing System (CEP). The Central Processing System is responsible for the processing and combination of the data from all stations in such a way as required by the specific science application the user is interested in.
4. Scheduling, Administration and Specification (SAS). Given the specification, the main responsibility of SAS is to schedule and configure the system in the right mode.
5. Monitoring And Control (MAC). The main responsibility of MAC is to control the system (in real-time) based upon the actual configuration of that moment. Additionally, MAC facilitates the (real-time) monitoring of the present state of the system.
6. System Health Management (SHM). SHM is identified as an autonomous block to predict and act on failures of the hardware before it fails. Ultimately it should identify which system component is the cause of a failure.

The main responsibility for CEP is to correlate the station data and deliver a data product which can be further processed by the user. The Central Processing Unit is divided in four sections: an on-line section for processing of real-time data streams from the stations, a storage section collecting the processed data streams and making the resulting datasets available to the off-line processing section which stores the final dataproducts on the archive/export section from where they are transferred to the LTA. MAC is responsible for controlling CEP and allocates resources to particular observations, which can run in parallel. The Blue Gene/P (BG/P) racks collect the station data and correlate the stations with each other and consequently reduce the data. The resulting data streams from the on-line processing sections are collected in the temporary storage subsystem. The BG/P is connected to the stations through the WAN. The data sent by the stations are sent in logical packages, each containing a time–frequency window of a single voltage beam. Each I/O node of the BG/P will receive data from the stations and run a data handling application that will buffer the input data and synchronize its output stream with the other I/O nodes based on the timestamps contained in the data (see Fig. 2).

Large amounts of processing power and internal interconnection bandwidth are provided through the BG/P supercomputer; a peak processing power of 34 TFlops is available for the processing tasks. This processing power in combination with the I/O capabilities of BG/P allows for a correlator capable of handling 2926 baselines of 32 MHz bandwidth together with the channel filter. The storage system provides disk space for the collection of data streams and storage of complete observation datasets for off-line processing. This storage is intended for temporary usage (typically 5 days) until the final data products are generated and archived or the raw data itself is exported or archived. Access to data in the storage system is through storage clients that have access to the metadata and file locations.

Finally, a general purpose Linux cluster is used for the off-line processing. The off-line processing section offers general-purpose processing power and high bandwidth interconnections to the off-line processing applications. The largest part of this cluster is a "normal" Linux cluster computer optimized on cost per Flop.

The LOFAR systems are controlled via the SAS/MAC systems. Integrated with SAS are two applications that provide the main interface to LOFAR for astronomical users: the "NorthStar" proposal preparation and submission application and the "Management of Measurements" (MoM, see [5]) project administration application. The latter manages the overall observation project process. It administrates resource allocations and project resource access and controls the interfaces to the LOFAR Information System.

[3] http://www.lofar.org/astronomy/surveys-ksp/surveys-ksp.

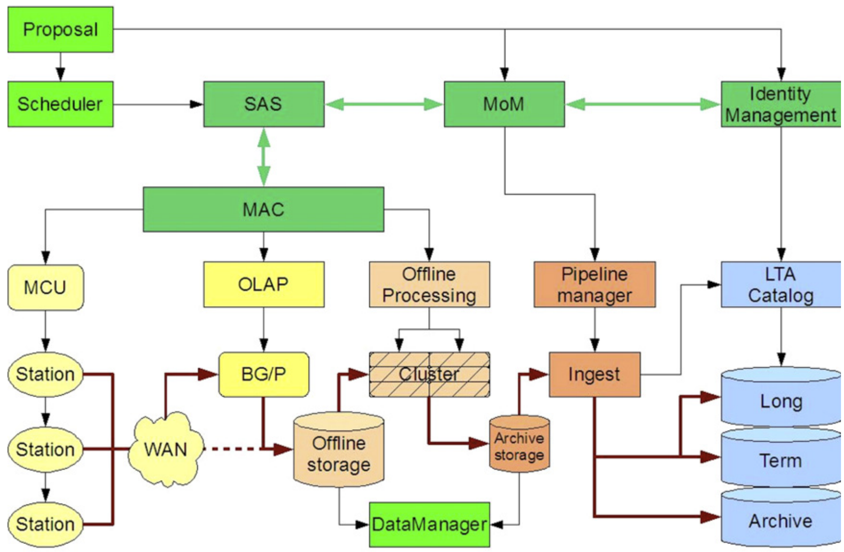


Fig. 2. LOFAR raw data processing and data ingest.

2.2. LOFAR wide area network

The LOFAR stations are connected to the Central Processing system through the LOFAR Wide Area Network (WAN, see [6]). The WAN consists mainly of dark fibre and active communication equipment (network switches, optical multiplexers and media converters) connecting each station with CEP. Connections from most Dutch stations are bundled at the concentrator node, close to the LOFAR core, from where data is transported over the LOFAR backbone fiber to Groningen. The WAN transports observation data and a relatively small fraction of the data streams due to monitoring and control data. The WAN is implemented using 10 GB Ethernet technology. The bulk data streams vary in bandwidth between 2 and 20 Gbps per station and are based on UDP.

From CEP, science grade LOFAR data products are transferred to LOFAR LTA sites over 10 Gbps connections that are currently dedicated for this purpose.

2.3. LOFAR data products

LOFAR provides a number of scientific applications, each with their own requirement with respect to the data products to be produced. The main LOFAR data products will be:

- UV data, a raw type of dataproduct that stores correlations of LOFAR station data streams. A single observation will produce a multi-terabyte size UV data set. UV data can be used for re-processing to generate, e.g., (improved) images.
- Images, a processed datatype that stores multidimensional "pixel" data (axis may be 2- or 3-D spatial as well as, e.g., frequency and polarisation). Images can have a size ranging from Gigabytes to Terabytes.
- Beamformed data, a type of dataproduct storing the combined (added) timeseries of one or more LOFAR stations. It is a type of data product used to study transient phenomena. Terabytes of data are potentially produced per hour.
- TBB data, a datatype storing full timeresolution (sampling frequency 200 MHz) dumps of antenna data (up to 96 per station) for a fraction of time up to a couple of seconds. These dumps are triggered by a cosmic event found in the LOFAR datastream or by special purpose detectors placed near the stations. A single dump of the full LOFAR array will generate Terabytes of data.

Besides the above dataproduct types there will be many derived and, in general, smaller sized dataproducts as well as catalogues of observation and event data. All these dataproducts and related processing parameters as well as links between different items are integrated in the LOFAR LTA data model, which is shown at Fig. 5.

2.4. LOFAR data processing requirements

According to LOFAR Archive and Reprocessing Requirements [7], the data stored in LTA will rise from the current 100 TB to 800 TB in 2010 and 1.5 PB in 2011. Due to inflow of new observations and reprocessing of already archived data the total data volume will rise up to 20 PB in the next 5 years of observations. Because the research groups participating in LOFAR and the available storage and computing resources are distributed, the data must be stored in different geographical locations, while being accessible for retrieval and reprocessing by any of the Lofar users. An important requirement is an open nature of the archive and data processing system. In practice this means that new groups of users (new projects) can be added to the LOFAR Information System with associated storage space and processing resources. They should be able to integrate their resources into the system with minimal effort and resources spent on adaptation of the LOFAR Information System.

3. LOFAR long term archive

The development of the LOFAR Information System did not start from scratch. In fact, the LOFAR Information System is an extension of the Astro-WISE Information System that was initially developed to store and process data of the KIlo Degree Survey (KIDS), an optical survey that will be performed on the VLT (Very Large Telescope) Survey Telescope (VST) starting in 2010. The VST is a 2.61 m diameter imaging telescope installed on the Paranal Observatory in Chile. The instrument installed on the telescope is OmegaCAM, a large format (16 × 16 k) CCD camera which will produce up to 15 TB of raw image data per year.

The optical image data processing goes from the raw image (which is basically a snapshot of part of the sky in some wavelength range) to the final science-ready catalog which is the same part of the sky but in the form of measured positions of objects on the sky and their calibrated brightness. To come to this final

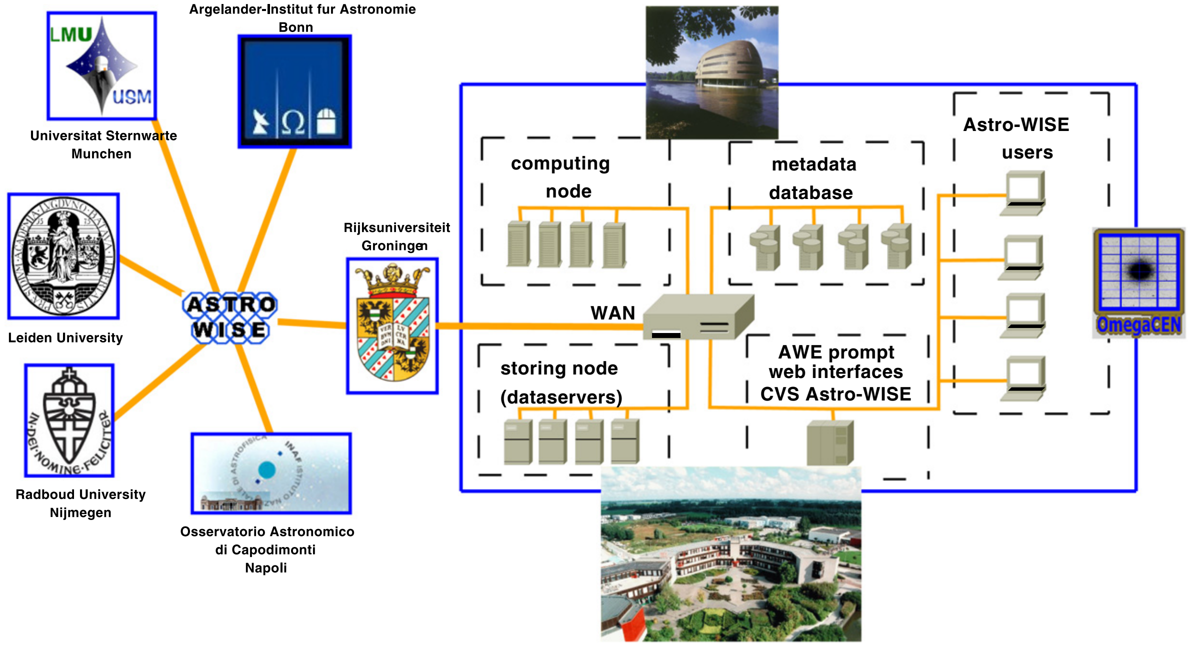


Fig. 3. Typical architecture of an Astro-WISE site in full mode (data storage node, computing node, web interfaces, CLI and version control system). The Rijksuniversiteit Groningen Astro-WISE node is shown (the computing node and metadata database are hosted by Donald Smits Center for Information Technology—CIT, top picture, the other elements are hosted by Kapteyn Astronomical Institute/OmegaCEN, bottom picture), sites of other participants have similar structure.

stage, a scientist must go through a number of steps and perform processing of large two-dimensional arrays (or three-dimensional in the case of radio astronomy). As we learned from LOFAR data processing requirements, the data volume in the case of LOFAR reaches Petabytes per year with each file (image or other data product) reaching Terabyte sizes.

The following sections review in detail changes which we had to introduce to adapt Astro-WISE for LOFAR and to build a new information system for LOFAR. We would like to underline that the LOFAR Information System inherits all the features of Astro-WISE we describe below.

3.1. Tier architecture

The Astro-WISE information system implements three main principles which come from the requirements to the scientific information system described above. These principles are:

1. Component based software engineering (CBSE). This is a modular approach to software development, where each module can be developed independently and wrapped in the base language of the system (Python, Java) to form a workflow in the system.
2. An object-oriented common data model used throughout the system. This means that each module, application and pipeline will deal with the unified data model for the whole cycle of the data processing from the raw data to the final data product.
3. Persistence of all the data model objects. Each data product in the data processing chain is described as an object of a certain class and saved in the archive of the specific project along with the parameters used for the data processing.

Astro-WISE is a distributed system with federated resources and a partially mirrored metadata database. Each Astro-WISE site is independent and can be deployed in various configurations. The components are:

1. Data storage nodes called dataservers that are used to store all data files created by the programs run by the users of the system or ingested by the users.
2. A metadata database to store a full description of each data file with all links and references to other objects.
3. Astro-WISE programming environment to provide a command line interface (CLI) to the system and a number of web interfaces which provide the users with an easy and fast access to the stored data and the data processing facilities.
4. Computing nodes for the data processing.
5. A version control system for developers to implement new modules, classes and libraries into the system.

An Astro-WISE site can consist of all the components described above (full node, Fig. 3) or some of the components down to the CLI only. In the latter case, users should specify which external data storage nodes and metadata database instances will be used by default.

One of the LOFAR requirements to the information system is to include external distributed storage capacities which are not controlled by the system itself (i.e., BiGGrid storage nodes). This requirement forced a review of the architecture of the Astro-WISE system and a subsequent switch to the tier approach. The scale of the data storage and computing for LOFAR reaches the level of the Large Hadron Collider experiment (LHC), and as in the case of LHC/EGEE [8] we adopted a tier architecture for the data storage and data processing grid (see Fig. 4). Three tiers were introduced:

- Tier 0 is the Central Processing System (CEP) which receives information, provides necessary computations, and delivers the data to the Long Term Archive;
- Tier 1 consists of a number of data storage nodes and computing nodes deployed in Groningen (Donald Smits Computing Center, CIT) or at any location where the research group participating in the project can provide a facility for data storage. A node of Tier 1 can be an Astro-WISE dataserver, a distributed filesystem node (GPFS, Lustre) or a BiGGrid site;
- Tier 2 is external to the LOFAR Information System and consists of users who would like to retrieve data from the system and/or process the data outside the system. Additionally, all external astronomers and communities such as the Virtual Observatory are considered as Tier 2 users.

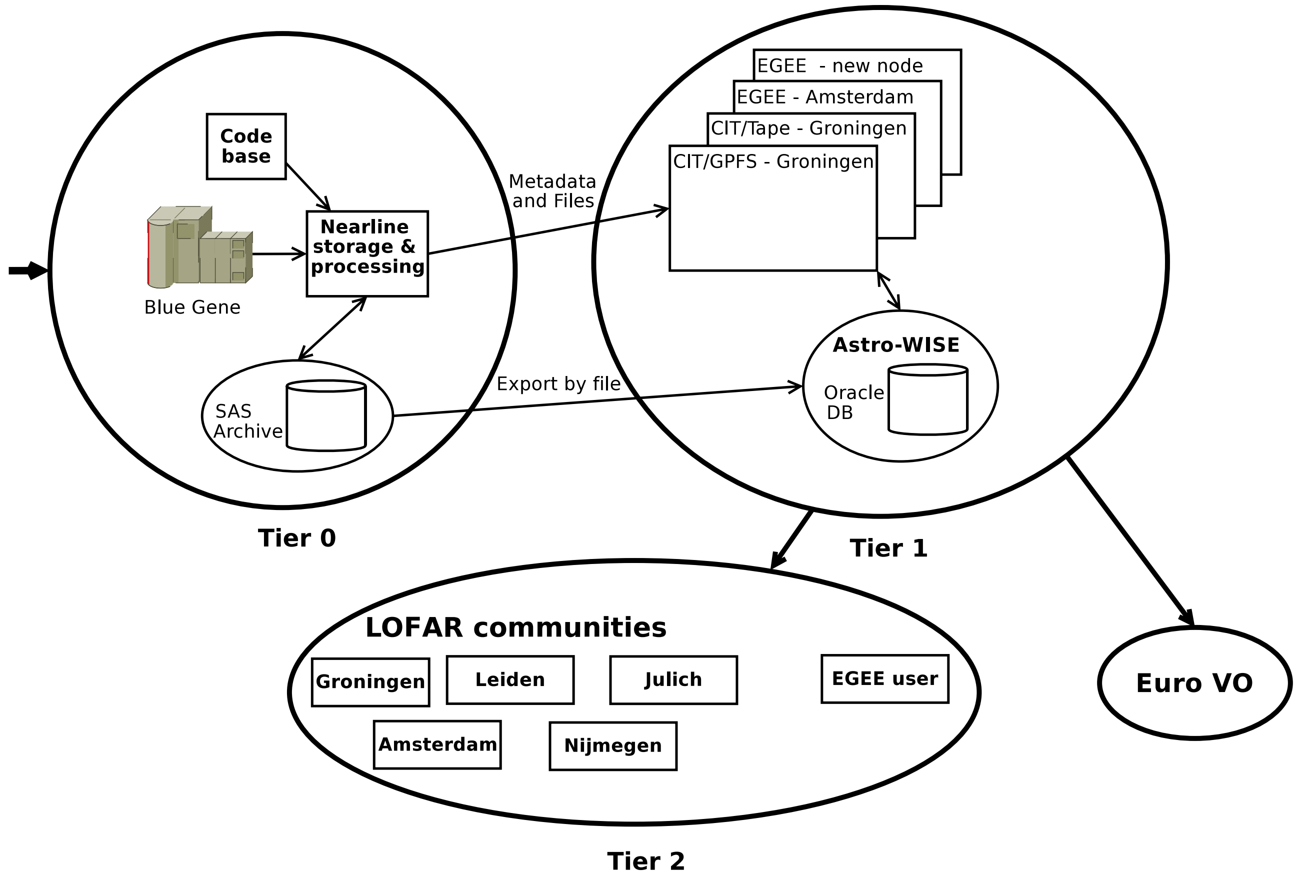


Fig. 4. Tier architecture adopted by Astro-WISE for LOFAR Long Term Archive.

The users can operate from Tier 2 to browse the metadata database (which is stored on a Tier 1), to start computations on computing elements at Tier 1 and to retrieve, update or ingest data on Tier 1.

In the case of the LOFAR Information System, we have similarities with the LHC Computing Grid architecture (tier structure of the system) and differences (Tier 0 and Tier 2 will not store data but ingest data into the system).

3.2. LOFAR data model

The data model implemented in the Astro-WISE Information System performs the data processing for optical images and can be adjusted for infra-red images. However, for the radio observations we had to develop and implement a completely new data model. The data model should cover all possible LOFAR use-cases, and include not only the description of radio images and other LOFAR data items, but also the processing parameters.

The LOFAR Long Term Archive (LTA) data model works from the data file it describes back to the measurement(s) that generated it. It does this by describing itself, all initial and intermediate data products that were needed to generate it, and all processes that were part of its generation. The data model describes entities of the LOFAR archive and relationships between them, as well as attributes for each entity. A "DataProduct" entity is a multi-Terabyte file with its own format (usually this is a multi-dimensional array of floating point values) which is ingested from Tier 0 or created by a pipeline running on Tier 1. Each entity has a full and complete metadata description done in Submission Information Package (SIP). SIP is an XML scheme[4] used to transfer metadata information from Tier 0 to Tier 1 and from Tier 2 to Tier 1. On Tier 1 the data from the XML file is ingested in the relational database (Oracle 11g RAC).

Concerning the processes, there is a fundamental distinction between the two types:

1. Measurements: these are the actual physical observations with the LOFAR and other instruments.
2. Pipelines: these are processes that take the products of measurements or other pipelines and do further processing.

In general this leads to an ER diagram (see Fig. 5) where "DataProduct" can appear as result of a pipeline executed on Tier 1 level (and then it will depend on N input DataProducts) or as a result of observations inserted to Tier 1 (and then it has a link to the corresponding Measurement). As for data products/files, several types are identified: Interferometer data, Transient event data, Tied array data and Images. In a similar fashion, measurement types are also identified, for example, Interferometer, Cosmic-Ray, Tied Array and Event triggered. Pipelines include the Imaging pipeline, the Cosmic-Ray pipeline, the Pulsar pipeline and the Solar pipeline.

3.3. Data and metadata storage

The original Astro-WISE approach to data storage is to use Astro-WISE dataservers. A dataserver is a front-end to underlying storage space which can use any operating system and any filesystem. The underlying storage space is blocked from any access other than via the dataserver interface. The dataserver provides store and retrieve operations for any types of files—from raw n-dimensional images to text files. A more detailed description of Astro-WISE dataservers is given in [9].

The users of the system do not know the actual location of a file on the underlying file system of dataservers and operate with

4 http://proposal.astron.nl/schemas/LTA-SIP.xsd.

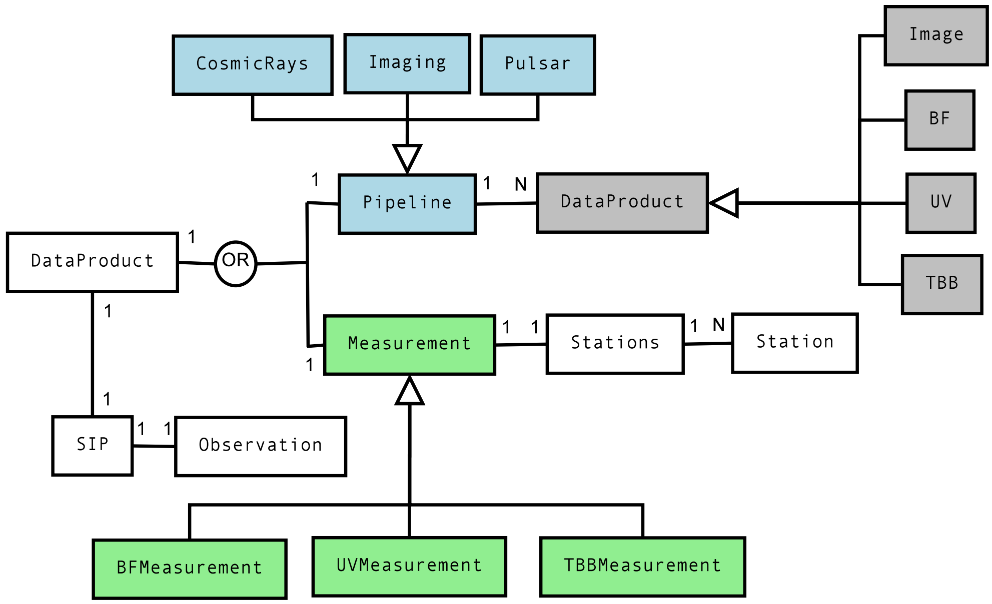


Fig. 5. LOFAR data model.

the URL of the file only. The URL has the form `http://<data server address>/<file name>`. Files are retrieved from dataservers with the use of the HTTP protocol, in the case of the LOFAR Information System, the SRM (Storage Resource Manager [10]) protocol is added to retrieve data from the BiGGrid storage.

Each data item is stored in the LOFAR Information System as a file that is registered in the metadata database with a unique filename. This unique filename must be used to retrieve the file from the system.

The data model realized in the system describes a file with a set of persistent parameters which are stored in the metadata database. Each data item (file) is represented by an object. Attributes representing the parameters used for the data processing are also part of this object.

The Persistent Object Hierarchy is the core of the data lineage and processing in the Astro-WISE environment. Each processing step, any changes to the file or metadata are done through the invocation of the methods on these objects, and these changes are saved through the persistent attributes of the object into the database.

As a result, the metadata database allows an intensive search on the attributes of an object or its dependencies. This is used to avoid partial reprocessing of the data if the object with desired processing parameters already exists and the user has access rights to this object. The detailed description of this feature of Astro-WISE is given in [11].

Access to the database is done through Python. Python is used for both the Data Definition Language and the Data Manipulation Language. The database interface maps class definitions that are marked as persistent in Python to the corresponding SQL data definition statements.

The introduction of BiGGrid storage in the LOFAR Information System requires changes in the current design of the Astro-WISE data storage. Astro-WISE stores files using the unique filename, users access files through one of the dataservers and retrieve it with the use of the HTTP protocol. The data stored on the Grid can be accessed and retrieved using the SRM protocol. In order to connect to the Grid data storage we have to add an additional interface and protocol to the existing system to deal with multi-protocol storage access.

In any data model realized in Astro-WISE, there is a class "DataObject" which defines how and where a file is stored. This class is coupled with the Storage Table in the metadata database, where all information relevant to the file is placed, including the URI of the file.

First of all, an interface to the BiGGrid is needed to retrieve and store data from/to the grid. The interface needs to be able to figure out whether a user has a valid proxy, and knows how to generate and renew a proxy given a certificate, and is able to store and retrieve files to or from the grid using certain protocols.

Second, if multiple protocols can be used to retrieve or store a file, the system needs to be able to provide an interface to allow users to select a protocol to use, or be able to retrieve or store a file using its default logic. For example, a user may prefer to retrieve or store a file only from or to the grid, or the user does not care where the file is retrieved or stored at all.

Finally, due to security reasons, the system needs to provide a mechanism which separates the access rights of the DataObject and its associated data files. This is required by LOFAR, where different groups or users may have different privileges to retrieve the associated files.

LOFAR put additional requirements on the data storage. These requirements can be summarized as:

- there can be multiple copies of the same file, i.e., the Storage Table can have multiple entries of the same object, stored in different locations (original Astro-WISE dataservers, BiGGrid nodes, etc.);
- any user can browse the metadata database by parameters but cannot retrieve the file if he has insufficient privileges to do it (in the original Astro-WISE system the user cannot see anything about the stored data item if he has insufficient privileges);
- the framework should be easily extended to support other storage systems by adding the corresponding interface (for Global File Systems, for example).

In the EGEE Grid/BiGGrid, the LFC (LCG file catalog) is a central component. It keeps track of the location of files over the Grid. All files are uniquely identified with a Grid Unique IDentifier (GUID). Multiple, human readable, logical file names (LFN) may point to this GUID. Each file may be stored at a number of physical locations. The URI of the files stored on the grid starts with `srm://`, therefore we use the term "srms" here to describe the URI of these files.

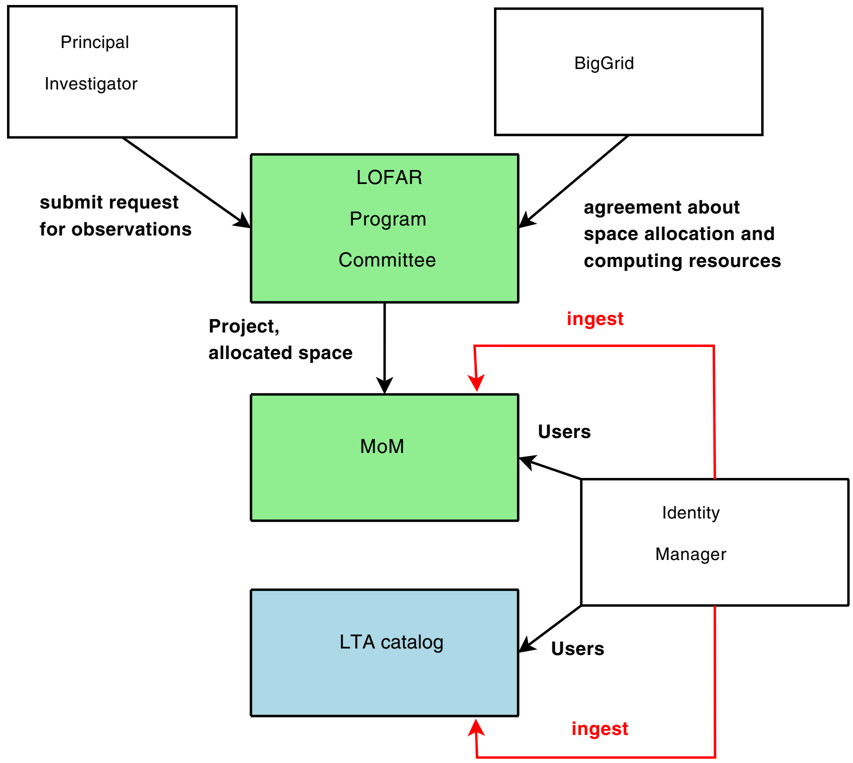


Fig. 6. LOFAR authorization and authentication system.

The advantages of storing srms in the Storage Table instead of using the LFC to retrieve them are:

- multiple srms of a file can be stored in the table, which makes it easier to check the number of copies of the file and to manage multiple copies of the same file;
- srm commands can be used to store and retrieve data from the grid directly; using these, we avoid the dependencies on the LFC, therefore making the system more robust.

The detailed design of the framework and how it is connected to the rest of the system is described in [9].

All the I/O of the processes makes use of the database. This includes all information about the processing of the data and the processing history. The database stores all persistent objects, attributes and raw data either as fully integrated objects or as descriptors. Only data arrays are stored outside the database in images and other data files. However, their unique filenames are stored in the database. Data can therefore only be manipulated through interaction with the database. A query to the database will provide all information related to the processing history and to all related files, attributes and objects. Thus, the system provides the users with transparent access to all stages of the data processing and thereby allows the data to be re-processed and the results to be fed back into the system. We utilize inbuilt security and authorization features of the database to enable sharing data in the system without breaking the dependency chain. The authorization is implemented at the level of the metadata database, that is, any user logging into the system through any interface (webservice, CLI) is checked in the database for his password, privileges, user group and access rights. In the case of LOFAR this involves additional challenge as there are three authorization systems which should be integrated: the original Astro-WISE authentication and authorization system (A&A), LOFAR A&A and EGEE Grid/BiGGrid A&A.

3.4. Authentication and authorization

In these three systems LOFAR A&A is the master one, all users and their privileges should be defined by LOFAR, exported to Astro-WISE A&A and coupled with Grid A&A mechanism (Virtual Organization and users in VO).

There is a system at Tier 0, called MoM (Management of Measures), where each LOFAR project is defined. The Principle Investigator (PI) of the project applies for the observations with LOFAR. If the proposal is accepted by the LOFAR Observatory, the description of the project and the information about the allocated storage space is ingested in MoM. The PI also assigns users (already registered in the LOFAR observatory) to the project.

The Novell Identity Manager (IdM) instance at Tier 1 creates an account for any users registered in the LOFAR observatory. These accounts are propagated to MoM on Tier 0 and to the metadata database (LTA catalog) on Tier 1 (see Fig. 6). Authentication with the database does not require a direct connection to IdM and/or MoM. If any of the subsystems are unavailable, the other subsystems will continue working and they will be synchronised automatically. There are staging tables in the metadata database at Tier 1 which are identical to the corresponding user, project, role and the resource classes as defined in the IdM. Selected columns from these tables are copied to the metadata database and optionally a database account is created (if this is requested by the PI, i.e., if this user is allowed to ingest data into the system). Passwords are propagated in the form of SHA1 or MD5 hashes.

Resources in the system are managed by the PI of each LOFAR project. They assign the storage space and computing resources for users in the project or revoke the user's privileges. They obtain information about occupied storage space from the Storage Table of the metadata database.

3.5. Data processing

Data processing in the LOFAR Information System is based on the Astro-WISE system which contains a sophisticated distributed processing system. This fully scalable system was designed to overcome the huge information avalanche in the wide-field astronomical imaging surveys. The system allows users to process raw science data up to the final science-ready data product maintaining a history of dependencies that can be followed from the end data product back to the raw observational data.

The distributed processing unit (DPU) sends jobs to parallel clusters or a single machine. It can automatically split up a processing job into parallel parts that can be run on different nodes. It allows users to run their own Python code. The system runs on OpenPBS[5] and OSCAR[6] or under its own queue management software. A DPU supports synchronization of jobs and can transfer parts of a sequence of jobs to other DPUs.

To implement data processing for LOFAR we had to enable DPU access to the GRID (BiGGrid) compute nodes. The main challenges were how to submit a job, how to monitor it and how to do the grid-authentication in a user-friendly way. Another problem is that the proper data processing software (which includes an access layer to the metadata database) must be installed on the assigned BiGGrid compute nodes. To do this, pre-compiled, self-extracting packages were made from the Astro-WISE software, which will be downloaded to the BiGGrid-nodes and executed before running the actual job. At BiGGrid sites which allow us to install the necessary software locally, this step is of course not needed.

The authentication problem was solved by using a MyProxy Upload Tool[7] [12] which should be run by the user once every two weeks at most. These proxies are protected by the Astro-WISE username and password. Whenever a user submits a job to the EGEE/BiGGrid DPU, the DPU retrieves the user credentials secured with the Astro-WISE username and password from the MyProxy server.

After the DPU has obtained a proxy from the MyProxy server, the required VOMS credentials are added to the proxy on the fly.

Unlike Astro-WISE job submission, the LOFAR Information System will use a robot certificate to submit jobs. The users will be able to launch a job on BiGGrid computing nodes without obtaining a personal certificate from a BiGGrid authority, i.e., all LOFAR users will share the same robot certificate on Tier 1. LOFAR will control the access to the computational resources at Tier 1 via the LOFAR Information System's authorization system.

3.6. Failure management

The robustness of the system is one of the core requirements of the LOFAR Long Term Archive. This robustness cannot be provided without a scenario of failure management which covers all sites involved in the information system. The problem of any distributed system of this scale is that the failure management is done on each site of the system independently. As a result, we have multiple systems of backup/recover which are acting on the base of the policy enforced by the LOFAR Consortium.

The backup and recovery of data is done at each site separately. Each Grid site involves its own system to backup the data. On the CIT node of Tier 1 (Groningen) RAID6 arrays are used and tape for backup.

The metadata database can be mirrored (or partially mirrored as it is done in the case of Astro-WISE) on any sites participating in Tier 1, but one copy of the metadata database always resides at the Groningen node of Tier 1. The metadata database and all its mirrors are backed up continuously with full online backups being made weekly. At Tier 1 in Groningen the metadata database is hosted by a database cluster distributed over two locations to provide maximum uptime. In the case of a disaster, experience with the Astro-WISE Information System has shown that complete disaster recovery of a 4 TB metadata database from backup can be done within one day, with a minimum of two hours. As well for each LOFAR use case (for example, data ingest) we have a scenario for dealing with errors.

The metadata database is responsible for the integrity and sustainability of the system and contains links to all the data stored in the system. The DPU monitors the status of jobs on a compute cluster or on the Grid. It knows when a job is queued, running, succeeded or failed. Running jobs inform DPU as they are still active. The detailed status of a job is available via a web page which the DPU creates on-the-fly.

4. Science data processing interfaces

The main language used in the system is Python, but each user can develop his own application or use an existing application which should be wrapped into Python. Usually, users develop pipelines or workflows using existing blocks with the help of predefined Python libraries and classes. Users can also change an existing data model if necessary or implement a new one. The Command Line Interface of the LOFAR Information System, the LWE (LOFAR-WISE Environment), can be installed on a site without any other components (data server and/or metadata database). Basically the LWE prompt is a link to a local Python installation with the corresponding libraries and environments. Apart from the LWE prompt, the LOFAR Information System will support a range of web interfaces. This will make it possible for a user to work with data stored in the information system, without the LWE prompt, but use the world wide web only. The following web services are in a developing phase:

- dbviewer[8] —the metadata database interface allowing users to browse and query all attributes of all persistent data classes stored in the system;
- LTA web service[9]—a quick search web interface, the input is limited to a few query parameters, the output only shows the major data products and measurements.

In the future there will be more web services built based on the existing Astro-WISE services:

A web service to start, configure and monitor standard pipelines (similar to Astro-WISE target processor[10]).
A web service to convert data products to viewable images in the browser and generate graphs (similar to Astro-WISE cut-out service[11]).

5. Use cases

We selected two of many LOFAR use cases to show the most prominent features of using the LOFAR Information System for data processing. These cases underline the advantages of the system and show how time and resources can be saved by using an information system instead of individual data processing by the users.

Surveys. One of the goals of LOFAR is to make a survey of the sky at low radio frequencies. During observations and subsequent processing, images of the observed sky are created from the

5 http://www.openpbs.org.
6 http://svn.oscar.openclustergroup.org/trac/oscar.
7 http://myproxy.egee.astro-wise.org.
8 http://rugtest5.service.rug.nl:9002.
9 http://lofar.astro-wise.org.
10 http://process.astro-wise.org.
11 http://imageview.astro-wise.org.

measured interferometer data at Tier 0. Basically, this is done through a discrete Fourier transformation. What makes this process complex is a repeated loop of self-calibration and source detection. This happens because in an interferometer the aperture is only sparsely filled, so knowledge of the observed objects and the system is needed to calibrate the data.

The calibration happens on several parameters, including antenna location, ionosphere behavior, human generated interference, systematic offsets, system errors, celestial object location and flux, system clock, etc. The initial models of the system, celestial objects and ionosphere will be improved over time as more of the sky is surveyed. One of the uses of the Tier 1 storage and processing will therefore be that later in time older data will be re-calibrated and imaged, as then the models of the various systems will have been improved and one will be able to generate images with more accuracy and a better signal to noise ratio.

The lineage of data items will allow the user to relaunch the pre-defined pipeline and trace all changes in the surveys due to the arrival of the new data.

Epoch of Reionisation. The LOFAR Epoch of Reionisation (EoR) project[12] will detect directly the extremely subtle radio signal (the redshifted 21 cm line emission) from gas around the EoR. To achieve this, a similar process to the one described above for Surveys happens. Over time, data will be collected and initially processed at Tier 0. Once more information about the various systems is known, the data will need to be reprocessed at Tier 1 to improve accuracy and the signal to noise ratio. This reprocessing will be done on a regular basis (once per half year) during the whole period of the observations with LOFAR.

At the end of July 2010 the core of both pipelines—the imaging pipeline—has been installed both in the Grid and the Target (CIT site at Groningen) environments and is in the testing phase.

6. Current status and future development

The core elements of the LOFAR Information System are in place and ready to be used. Tier 0 at Groningen with the BlueGene/P and the offline storage for the processed data is already in use since 2009. The LOFAR A&A system is tested and in place.

Tier 1 has by the end of July 2010 2 sites: Target/CIT (Groningen) and a BiGGrid[13] site managed by SARA[14] (Amsterdam). In total the BiGGrid central facilities (including Amsterdam and Groningen) currently have 6750 cores and about 5 PB of storage. Part of this will be available to LOFAR. In CIT (Groningen) the LOFAR Information System can use 1.5 PB of the Target storage (with an ability to be extended up to 10 PB in the future). The compute cluster at the University of Groningen with 3280 cores, more than 50 TB of global scratch space and 237 GB scratch disk space locally for each node in the cluster can be used as the Target compute facility. Target can also make use of the Big Grid compute facilities. Regarding the Groningen Grid facility, plans have been made for an extension in the near future. The up-to-date status of the Target storage and the computing facilities can be inquired at the Target website.[15] Not all of these facilities will be used by LOFAR, but the LOFAR Information System will be the most massive user of the Target infrastructure for the next 5 years.

LOFAR pipelines can be run at Tier 2, i.e., a LOFAR user can access data stored in the system and upload the processed data back to the system. The final goal is to implement an approach where several pipelines are running completely inside the system, and users are allowed to change parameters of these pre-defined pipelines (this approach is implemented in the parent Astro-WISE system[16]).

Nevertheless, despite all the experience achieved with the integration of pipelines in the Astro-WISE information system, there are a number of additional challenges which the LOFAR Information System poses, one of them is the development of the subsystem, which is capable of handling jobs on the Grid external to the original Astro-WISE Grid, especially for the handling of error messages and lost jobs (see, for example, [13] for the description of the problems and possible solutions).

The LOFAR Information System is a key project for Target[17]—the expertise center built by the University of Groningen, Astron, IBM and Oracle. Target develops information systems for a number of scientific communities including LOFAR. The center started working in 2009 and the LOFAR Information System is Target's first delivery.

Acknowledgement

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

References

[1] E. Valentijn, et al., Astro-WISE: chaining to the universe, in: R.A. Shaw, F. Hill, D.J. Bell (Eds.), Proc. of ADASS XVI, in: ASP Conf. Ser., vol. 376, 2007, p. 491.
[2] E. Valentijn, A. Belikov, LOFAR information system design, in: Proc. of Euro-VO Data Center Alliance Workshop: Grid and Virtual Observatory, Mem. Soc. Astro. It., vol. 80, 2009, p. 509.
[3] A.W. Gunst, M.J. Bentum, Signal processing aspects of the low frequency array, in: Proc. ICSPC 2007, 2007, p. 600.
[4] M. de Vos, A.W. Gunst, R. Nijboer, The LOFAR telescope: system architecture and signal processing, Proceedings of the IEEE 97 (8) (2009) 1431.
[5] A.P. Schoenmakers, Measurements set version 2 for LOFAR, 2006.http://www.lofar.org/wiki/lib/exe/fetch.php?media=public:documents:ms2_description_for_lofar_v05.pdf.
[6] D.H.P. Maat, R.B. Gloudemans, LOFAR-WAN architectural design document, 2007. http://www.lofar.org/wiki/lib/exe/fetch.php?media=public:documents:20_lofar-wan_add.pdf.
[7] C. Vogt, M. Wise, LOFAR archive and reprocessing requirements, 2008. http://www.lofar.org/wiki/lib/exe/fetch.php?media=public:documents:lofar_archive_requirements_rev0.9.pdf.
[8] G.A. Stewart, D. Cameron, G.A. Cowan, G. McCance, Storage and data management in EGEE, in: Proc. Fifth Australasian Symposium on Grid Computing and e-Research, AusGrid 2007, Ballarat, Australia, CRPIT, vol. 68, 2007.
[9] K. Begeman, et al., Merging grid technologies: astro-WISE and EGEE, Journal of Grid Computing 8 (2010) 199.
[10] A. Sim, A. Shoshani, et al., The storage resource manager specification, version 2.2. http://www.ogf.org/documents/GFD.129.pdf.
[11] J. Mwebase, D. Boxhoorn, E. Valentijn, Astro-WISE: tracing and using lineage for scientific data processing, in: Proc. of the 2009 International Conference on Network-Based Information Systems, 2009, p. 475.
[12] MyProxy software. http://www.ngs.ac.uk/tools/certwizard.
[13] M. Rahman, R. Ranjan, R. Buyya, Cooperative and decentralized workflow scheduling in global grids, Future Generation Computer Systems 26 (2010) 753.

K. Begeman obtained Ph.D. in Radioastronomy from Groningen University in 1987. Ever since he is working as an expert in astronomical data processing in Kapteyn Astronomical Institute and is a leading developer of Astro-WISE Information System.

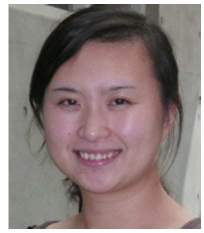
12 http://www.astro.rug.nl/~LofarEoR.
13 http://www.BiGGrid.nl.
14 http://www.sara.nl.
15 http://www.rug.nl/target/infrastructuur.
16 http://process.astro-wise.org.
17 http://www.rug.nl/target.

A.N. Belikov was graduated from Moscow State University in 1996. He obtained Ph.D. in Astronomy in 1999 from Astro-Space Center of Lebedev Physics Institute and Ph.D. in Computer Sciences from Mannheim University in 2005. Since 2007 he is working at OmegaCEN, Kapteyn Astronomical Institute for Euro-VO DCA/AIDA projects porting Astro-WISE information system to the Virtual Observatory and developing LOFAR Long Term Archive.

D.R. Boxhoorn started to work in SRON, Groningen in 1990 developing IRAS data reduction for the next 3 years. In 1994 he obtained M.Sc. in Astronomy from Groningen University and spent next 5 years working on ISO-SWS data reduction and database. From 1999 he is employed by OmegaCEN, Kapteyn Astronomical Institute and is OmegaCEN DBA and data processing expert.

F. Dijkstra obtained a Ph.D. in theoretical chemistry from Utrecht University in 2000. After his Ph.D. he moved to the area of high performance computing and Grid, further exploiting the knowledge gained during his Ph.D. about software optimisation and parallelisation and on running demanding simulations on different computer platforms. Currently he is working as Grid expert at the Donald Smits Centre for Information Technology of the University of Groningen in the framework of Target project, working on improving the connections between Target and other Grids. He has also been involved with the EU project EGEE for several years, first at SARA in Amsterdam and later in Groningen. His main interests are the operations of the local Grid infrastructure and using his knowledge about the Grid infrastructure to get new applications running on the Grid.

H. Holties is System Engineer in the Radio Observatory department of ASTRON. He received his Ph.D. in Physics for research on fusion plasmas through numerical simulations. Since 2002 he has been responsible for the software and ICT systems of the Westerbork Synthesis Radio Telescope (WSRT) and has led software development projects aimed at beginning to end support of astronomical observation programs. From 2007 he has become actively involved in the software development of the LOFAR telescope and he is currently responsible for the design and implementation of the LOFAR Long Term Archive.



Z. Meyer-Zhao is a Grid expert working at the Center for Information Technology (CIT) at the University of Groningen. She received her first MSc degree in Applied Computer Science in 2005. One year later, she received her second MSc degree in Computer Science specialized on Artificial Intelligence. Currently, her work is mainly focused on the integration of the Astro-WISE system with the Grid, to enable storing and processing of huge quantities of scientific data and developing LOFAR Information System.

G.A. Renting was graduated from Groningen University in 2000 with M.Sc. in Applied Physics, with main subject on Material Science, (polymer semiconductors), minor subjects on Astrophysics and Software Engineering. For next five years he was employed by YP Engineering developing SCADA and telemetry systems for water management and industrial automation and by electricity load forcaster Electrabel modeling electricity usage patterns. From 2005 he is working as software engineer at ASTRON. Currently he is working on radioastronomy science pipeline processing and large scale archive systems for both the LOFAR and WSRT telescopes.

E.A. Valentijn is head of OmegaCEN and is Professor of an Astronomical Information Technology at the University of Groningen. He obtained Ph.D. in Radioastronomy in 1979 from Leiden University. From 1998 prof. Valentijn is NOVA programme manager of OmegaCEN and coordinator of Astro-WISE. Valentijn leads the Target, a multi-disciplinary programme building science e-infrastuctures of very large data storage and processing.

W.-J. Vriend finished his Astronomy study in 2000. For next 4 years he worked for Sogeti, an IT company delivering IT expertise to financial, telecom and industry companies in the Netherlands. From 2005 W.-J. Vriend is working for OmegaCEN as a software engineer and developer implementing web services, interfaces and pipelines for Astro-WISE and LOFAR Information System.