University of Groningen

Scaling Astro-WISE to LOFAR long term archive. From KIDS to LOFAR

Begeman, K.; Belikov, A. N.; Boxhoorn, D.; Dijkstra, F.; Holties, H. A.; Renting, G. A.; Vermaas, N.; Vriend, W. -J.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

ORIGINAL ARTICLE

Scaling Astro-WISE to LOFAR long term archive
From KIDS to LOFAR

**K. Begeman · A. N. Belikov · D. Boxhoorn ·
F. Dijkstra · H. A. Holties · G. A. Renting ·
N. Vermaas · W.-J. Vriend**

Abstract The Astro-WISE information system was developed to handle data processing for the KIDS survey. In this paper we describe the adaptation of the WISE concept to allow scaling to support archives containing tens of petabytes of stored data and the changes we introduced to accommodate the system for the LOFAR Long Term Archive. With this we provide an example of how

K. Begeman · A. N. Belikov (✉) · D. Boxhoorn · W.-J. Vriend
Kapteyn Astronomical Institute, University of Groningen, Landleven 12,
9747AB, Groningen, The Netherlands
e-mail: belikov@astro.rug.nl

K. Begeman
e-mail: kgb@astro.rug.nl

D. Boxhoorn
e-mail: danny@astro.rug.nl

W.-J. Vriend
e-mail: wjvriend@astro.rug.nl

F. Dijkstra
Donald Smits Center for Information Technology, University of Groningen,
Nettelbosje 1, 9747AJ, Groningen, The Netherlands
e-mail: f.dijkstra@rug.nl

H. A. Holties · G. A. Renting · N. Vermaas
Netherlands Institute for Radio Astronomy (ASTRON), Oude Hoogeveensedijk 4,
7991 PD, Dwingeloo, The Netherlands

H. A. Holties
e-mail: holties@astron.nl

G. A. Renting
e-mail: renting@astron.nl

N. Vermaas
e-mail: vermaas@astron.nl

Astro-WISE technology can be adapted to support a wider range and scale of data.

Keywords Information system · Grid · Scientific data processing

1 Introduction

Astro-WISE[1] [1] is an astronomical scientific information system which was developed to host the data for the Kilo Degree Survey (KiDS[2]). Astro-WISE combines data storage and data processing allowing the user to carry out intensive and in-depth data mining along with partial or full data reprocessing on the survey data. Despite the fact that the "parent" instrument for Astro-WISE is a VST wide-field camera OmegaCAM [2], it is possible to ingest into Astro-WISE the data from a number of instruments and process them with the same pipeline. In case of non-image data (spectroscopic or radio data) a new information system based on Astro-WISE technology can be created by implementing a new data model and incorporation of new pipelines.

The data volume of the KIDS survey, including all intermediate data products, will not exceed 0.5 petabyte. Incorporation of additional data sources such as the VIKING survey will further increase the data volume. However, supporting multi petabyte data archives such as needed for the LOFAR Long Term Archive [4], which is built for the LOFAR telescope[3] [3] that has been designed and constructed by ASTRON [5, 6] and is operated by the International LOFAR Telescope foundation (ILT), requires a new approach to the data storage. Furthermore, inclusion of new non-optical data requires changes to the data model while the distributed inhomogeneous character of the LOFAR Long Term Archive also requires integration with new types of storage and processing facilities.

Essential difference with Astro-WISE is that the system which is responsible for the ingest—Central Processing System of LOFAR (CEP)—is not a part of the archive. CEP generates data for the archive and ingests it into the archive, but the archive has no control over any resources of CEP or access to the data stored by CEP. In the case of Astro-WISE the data are ingested by the user of Astro-WISE.

In this paper we will show the changes that were implemented to make it possible to increase the capability of the Astro-WISE information system for LOFAR [7, 8] to handle tens of petabytes.

---

[1]http://www.astro-wise.org

[2]http://www.strw.leidenuniv.nl/~kuijken/KIDS/

[3]http://www.lofar.org

2 Adaptation of Astro-WISE for LTA

Astro-WISE is an astronomical information system which was originally developed for the processing of optical wide-field images. Astro-WISE allows processing of the data from raw images all the way up to catalogs keeping all dependencies and processing parameters in the system. A detailed description of Astro-WISE can be found in [9].

Astro-WISE is a distributed system with an infrastructure which is based on the principle of deploying independent nodes. Each node can contain a data storage (dataserver), metadata storage (relational DBMS), processing elements (usually an HPC cluster) and numerous user services. All these components are optional for each site, as long as at least one of them is available in the system built from all Astro-WISE sites. Components of all Astro-WISE nodes build the three layers of Astro-WISE: a metadata layer (all relational DBMSes are sharing the same data model, all metadata storage components are fully or partially mirrored), a data layer (all dataservers are connected and create a single data storage network) and a data processing layer (the user can choose to submit jobs to any processing element of Astro-WISE).

Astro-WISE implements a common data model shared by all users and combines component-based software engineering (CBSE) with an object-oriented data model. The CBSE allows to implement each step in the data processing as a separate and independent module. The module (pipeline recipe) processes the data retrieving it from the system and ingesting back into the system newly generated data products. At the same time all data products are described by the same object-oriented data model so that the user can trace all dependencies

Fig. 1 Tier architecture for the LOFAR Long Term Archive. Tier 0 is CEP, Tier 1 is the core of the LOFAR LTA and Tier 2 are LTA users
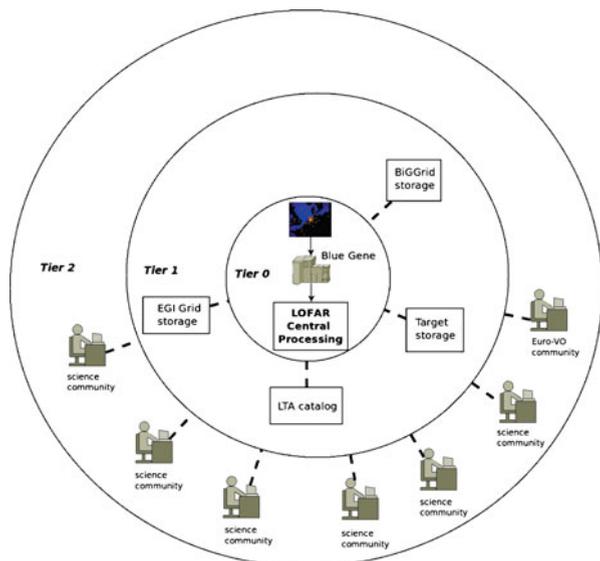
between data products. This approach allows to reprocess any data on any step of the data processing chain.

As we described above Astro-WISE is built from independent nodes, nevertheless, all these nodes create a single information system. The user can access any node in the system and send his job to any computing element of the system. The data ingest can be performed on any node with the data becoming available to the user on any node if the user has proper access rights. The adaptation of Astro-WISE system for the LOFAR Long Term Archive required changes in the architecture of the system for the following reasons:

– the ingest of the data is done from an external system which does not belong to LOFAR Long Term Archive—the Central Processing System (CEP). Although CEP can be geographically very close to LOFAR LTA storage it has its own system of Authorization & Authentication (A&A), separate hardware and software infrastructures. LOFAR LTA users must not access data storage at CEP, at the same time CEP should be able to access resources of LTA to ingest the data. In the original Astro-WISE all ingest operations are managed from Astro-WISE and by Astro-WISE users;
– for astronomers, the LOFAR LTA is an integral part of the LOFAR telescope. This requires, for example, that accounts, authorizations, and allocations are to be shared between all LOFAR services, including the LTA.
– The data storage is to be distributed across storage elements of different nature—the storage should incorporate Grid storage elements as well as GPFS based fileservers;
– the LOFAR LTA allows users to retrieve the data for external processing, i.e., new data products can be generated outside the LTA system and ingested into the archive with corresponding data provenance.

After reviewing all requirements described above, we changed the architecture for the LOFAR LTA from the original federated architecture with independent nodes to a tiered architecture (Fig. 1). Three tiers were introduced:

– Tier 0 is the Central Processing System which collects data from antennas, performs pre-processing, and delivers the resulting data products to the Long Term Archive. Tier 0 only provides temporary storage for LOFAR data products;
– Tier 1 consists of a number of internationally distributed data storage nodes and computing nodes . A node of Tier 1 can be an Astro-WISE dataserver, a distributed filesystem node (GPFS) or a Grid node;
– Tier 2 is external to the LOFAR LTA and consists of all systems that are used to retrieve data from the LTA and/or to process data outside the LTA. Communities such as the Virtual Observatory are considered to be Tier 2 users.

Unlike the "classic" tier architecture of the LHC Computing Grid, the LOFAR Long Term archive does not manage or include data stored on the Tier 0 and Tier 2 levels.

The adaptation of the Astro-WISE architecture to a new data can require changes in the architecture of the information system itself, as we will see below. We will review changes introduced to three layers of Astro-WISE: metadata layer, data layer and data processing layer.

2.1 Metadata layer

Astro-WISE separates the bulk of the data stored in files from the metadata description of these files. The level of detail of the metadata stored in the database can vary. The required parameters for metadata are a unique identifier of the data item, the name of the file containing the data item, and the owner of the data. For the user, each data product is an object with a number of persistent parameters which are saved in the metadata database. The implementation of a new data model is not much of a technical challenge but requires the definition of the data model itself taking into account not only the data processing chain but also additional descriptors, including scientific ones. For example, it is important to implement for each data product a number of persistent parameters which will review the quality of the data item and provide this information to the user.

The implementaion of a new system for authorization and authentication is more of a technical challenge. The original Astro-WISE authorization scheme is based on the built-in security and authorization features of the metadata database. The authorization of the user in the system determines the rights of the user to perform a number of operations: browsing (parts of) the metadata database, retrieval of data items, processing of data items, ingesting of the processed data, etc. All interactions with the data in the archive are done through the interaction with the metadata database.

For the LOFAR Long Term Archive, the authorizations of a user are determined by a system which is external to the archive. The central LOFAR A&A system is the master one, which defines all users and their privileges and exports them to Astro-WISE A&A.

The MoM (Management of Measurements) system at Tier 0 contains the definition of each project, all LOFAR users that are associated with the projects, and all resources (observing time as well as LTA resources such as long term storage capacity) that have been allocated to the projects. Each LOFAR project has a Principle Investigator (PI), a Contact Author (who may or may not be the PI), and optionally one or more Co-Investigators (CoIs).

A Novell Identity Manager[4] (IdM) instance is used to synchronize accounts, authorizations and allocations between MoM on Tier 0 and the metadata database (LTA catalog) on Tier 1. 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

---

[4]http://www.novell.com/products/identitymanager

synchronized automatically once the connection is re-established. 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 hashes.

2.2 Data layer

The data layer of Astro-WISE is implemented by a network of dataservers—special web services which process a request from the user with the name of the file, search for the file in the underlying disk space and send the file back to the user (see [9] for more details).

Despite the fact that the dataservers network is highly scalable, it requires the storage systems to be adapted to Astro-WISE, while for the LOFAR LTA it is necessary to include a number of storage elements which are not part of the Astro-WISE system. In the case of LOFAR LTA these are principally Grid storage nodes based on dCache which provide GridFTP and Storage Resource Management (SRM) interfaces [10]. The integration of Astro-WISE system and Grid infrastructure is described in [8] and [11]. This integration is used for LTA storage and allows to keep multiple copies of the same data item on a number of storage locations.

The LOFAR Long Term archive data ingest system is a connector between Tier 0 and Tier 1 and allows to move the data from the temporary storage on Tier 0 (CEP) to permanent storage in the archive.

The data ingest from CEP into the archive includes the transfer of files and an ingest of the metadata description of the data contained in these files into the metadata database of the archive. A LOFAR LTA data item consists of file(s) and metadata. The metadata are transfered in a single metadata file generated by the MoM system.

An operator initiates the data ingest transaction which starts by requesting Tier 1 for the special identifier of the ingest session called a Storage Ticket. A Storage Ticket holds the state of the ingest transaction on both Tier 0 and Tier 1 sides. During the ingest the state of the Storage Ticket is updated on both sides. Both on Tier 0 and Tier 1 the Storage Ticket is used to monitor the ingest process to spot ongoing, failed and successfully ingested data items.

Before the data is transfered, the destination location on Tier 1 nodes must be defined. This will primarily depend on the storage resource that has been allocated to the project the data item belongs to. Depending on the type of a storage node, the proper protocol will be used to transfer the files. The current storage locations are located in Amsterdam (SARA[5], Grid nodes),

---

[5]http://www.sara.nl
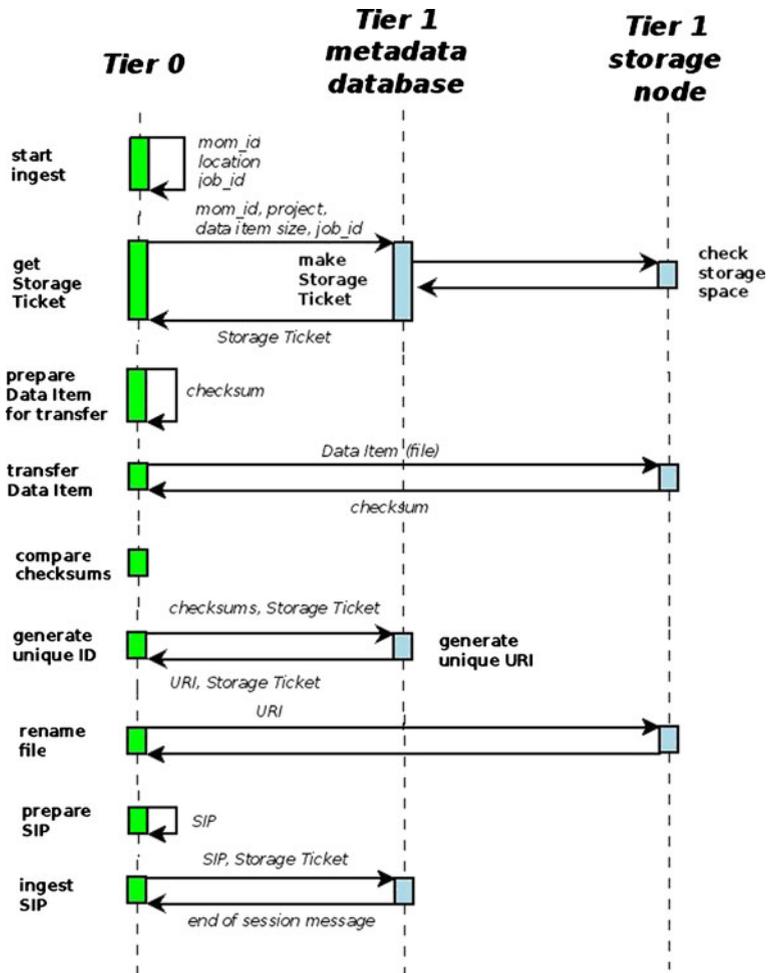
Fig. 2 Sequence diagram for a LOFAR LTA ingest transaction. Interactions between Tier 0 (CEP), Tier 1 metadata database and Tier 1 designated storage node are shown

Jülich (Forschungszentrum Jülich[6], Grid nodes) and Groningen (Target[7] [12], GPFS nodes).

Figure 2 shows the ingest transaction in detail. The ingest of metadata into the LTA database is done by using an XML-RPC server. The metadata is encoded in an XML file called the SIP (Submission Information Package). The structure of the SIP is specified by a specially designed SIP XML Schema Definition format. Both Tier 0 and Tier 1 are using the same XML schema and

---

can check the SIP by using the XML schema. This maintains consistency across the LOFAR system when changes in the data model are introduced either on Tier 0 or on Tier 1.

When the data files have been transferred to the designated node, the metadata from the SIP file is ingested into the metadata database of Tier 1. The Storage Ticket is archived and the ingest session is finished.
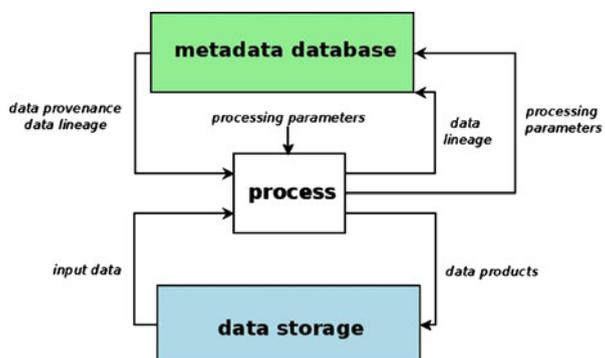
2.3 Data processing layer

The core principle of the data processing in the Astro-WISE information system is that the data in the archive can be processed and reprocessed keeping all data products, dependencies and processing parameters in the archive. Figure 3 shows a typical process in Astro-WISE which is initiated by user. The process requests the metadata database for all necessary data files, data files are retrieved from the data storage (dataservers), the process creates data products using processing parameters specified by the user and ingests these data product into the information system (data products submitted to the dataserver and metadata for newly created data is committed to the metadata database along with the processing parameters). In this sense the ingest procedure described above (see Fig. 2) is performed as a typical Astro-WISE process.

The pipeline should not depend on the location of the storage element or processing element which is used. To ensure uniformity among the LTA locations standards must be agreed on. The different LTA sites must agree on:

–   versioning of the LOFAR software and related libraries;
–   installment of the LOFAR software and related libraries.

All software (both developed for LOFAR and external packages) are to be identified by versions. The versioning is to be applied consistently during the installation of software so that all sites participating in the LOFAR data processing can support the same versions of a pipeline. The version of the

Fig. 3 Typical process in Astro-WISE system. Data and metadata are exchanged with metadata database and data storage

pipeline is stored as one of processing parameters linked to the resulting data product.

An important part of the LOFAR Long Term archive development is an integration of LOFAR pipelines with Astro-WISE Data Processing Units (DPU, see [9]). The DPU is an Astro-WISE front-end to the processing facilities. The user submits jobs to the DPU and interacts with the processing clusters via the DPU. The introduction of the DPU creates a uniform way for the user to interact with different processing elements. The DPU has a single interface for accepting jobs, but can dispatch these jobs to various compute clusters. The jobs are programmed in Python and form wrappers around (LOFAR or other) binaries. On the compute side, a DPU job-runner retrieves the needed data, generates the configuration files, executes the pipeline and stores the resulting data set. After the processing is finished the DPU transfers the log of the processing run back to the user. Using the Astro-WISE DPU the user has one interface towards all processing facilities of the LOFAR LTA.

Currently LOFAR pipelines are being tested both in the Grid and Target (HPC cluster) environment.

3 Conclusion

The LOFAR Long Term archive is an example of the adaptation of the original Astro-WISE design to the additional requirements from the LOFAR instrument. The case of the LOFAR LTA shows the high flexibility of Astro-WISE, which allows to include new elements keeping features of the original information system.

The LOFAR Long Term Archive is a key project for Target[8]—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.

---

[8]http://www.rug.nl/target

References

1. Valentijn, E., et al.: Astro-WISE: chaining to the Universe. In: Shaw, R.A., Hill, F., Bell, D.J. (eds.) Proc. of ADASS XVI, ASP Conf. Ser., vol. 376, p. 491 (2007)
2. Iwert, O., et al.: The OmegaCAM 16K × 16K CCD detector system for the ESO VLT Survey Telescope (VST). In: Dorn, D.A., Holland, A.D. (eds.) Proc. of the SPIE, vol. 6276, p. 62760A (2006)
3. Röttgering, H.: LOFAR, a new low frequency radio telescope. New Astron. Rev. **47**, 405 (2003)
4. Holties, H.A., et al.: The LOFAR long term archive. Exp. Astron. (2012, to be submitted)
5. de Vos, M., Gunst, A.W., Nijboer, R.: The LOFAR telescope: system architecture and signal processing. In: Proc. of the IEEE, vol. 97(8), p. 1431 (2009)
6. Gunst, A.W., Bentum, M.J.: Signal processing aspects of the low frequency array. In: Proc. ICSPC 2007, p. 600 (2007)
7. Valentijn, E., Belikov, A.: LOFAR information system design. In: Proc. of Euro-VO data center alliance workshop: grid and virtual observatory. Mem. Soc. Astro. It., vol. 80, p. 509 (2009)
8. Begeman, K., et al.: LOFAR information system. Future Gener. Comput. Syst. **27**(3), 319 (2011)
9. Begeman, K., et al.: Merging Grid technologies: Astro-WISE and EGEE. J. Grid Computing **8**, 199 (2010)
10. Fuhrmann, P.: dCache, the overview. White Paper, dCache.org. http://www.dcache.org/manuals/dcache-whitepaper-light.pdf (2011)
11. Begeman, K., et al.: Astro-WISE and Grid. Exp. Astron. (2012). doi:10.1007/s10686-012-9293-2
12. Belikov, A.N., et al.: Information systems playground—the target infrastructure. Exp. Astron. (2012). doi:10.1007/s10686-011-9262-1