# OmegaCEN bi-annual report 2008-2009

## 1. The OmegaCEN data center

OmegaCEN is the national datacenter for wide-field imaging and the expertise center for astronomical information technology. It is situated at the Kapteyn Astronomical Institute of the University of Groningen (RUG) and its core staff is funded by Kapteyn and NOVA. OmegaCEN is partner in various optical and radio wide field surveys, conducted by international collaborations.  The largest optical survey is the Kilo Degree Survey (KiDS), a 1500 square degree survey through 4 optical filters (u,g,r,i) with the OmegaCAM wide-field imager at the VST (commissioning planned summer 2010). OmegaCAM will lead and coordinate the KiDS data flow operations. The center is headed by prof Valentijn and brings together astronomers and astronomical IT experts. About 15 scientists (staff, postdocs, PhD students and scientific programmers) work at OmegaCEN and jointly conduct astronomical research and R&D in astronomical information technology.  OmegaCEN is coordinator of Astro-WISE, a unique, advanced survey information system for astronomy. Astro-WISE connects in real-time databases, compute and storage grids available at national datacenters and satellite nodes in the Netherlands, Germany, Italy and France. Geographically-spread survey teams collaborate in this single virtual research environment. Here, they calibrate astronomical wide field imaging data up to the Petabyte regime, do research analysis and publish the results. The KiDS team is ready to use Astro-WISE for the KiDS survey processing.
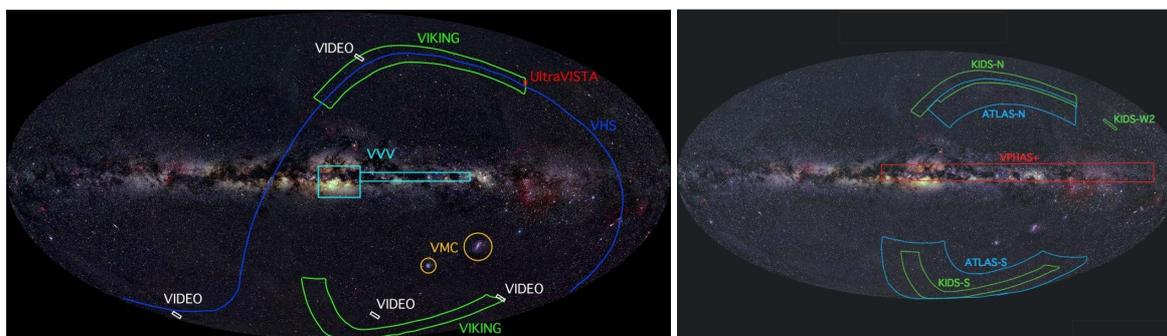


**Fig.1**: *Sky coverage of the VST (right) and VISTA (left) public surveys awarded by ESO. KiDS, VPHAS+, VIKING and ULTRAVISTA are all led or co-led by NOVA scientists.*

OmegaCEN also builds and operates an extensive hardware park.  This includes a database cluster (8 TB storage space and 128 GB internal memory) and 160 TB of file storage space connected to the users via web-portals. The datacenter also runs servers specially equipped for image stacking and on-line image cut-out services. Together with  the nodes in Munich (Pan-STARRS), Naples, Bonn and Nijmegen the distributed storage pool of Astro-WISE is near the Petabyte regime.
In 2008-2009 OmegaCEN further extended the collaboration with the RUG's Donald Smits Center for Information Technology (the computing center of the RuG), also providing  dedicated access to the High-Performance Compute Cluster which is upgraded in December 2009 to a 200 twelve-processor core cluster.

The extensive (web)services for database browsing, processing and publishing were improved. A new webservice, Quality-WISE, serves as a single portal for scientists to efficiently qualify survey data (see Verdoes Kleijn et al., 2009).  The Astro-WISE Grid was integrated with the EGEE Grid which enables the Astro-WISE applications to run on the EGEE Grid as a preparation for Lofar operations. This merging of Grid technologies has been published in a refereed computational science journal (Begeman et al., 2010).

OmegaCEN is the national representative for European Virtual Observatory (EURO-VO) programmes. It is partner in the EU Framework Programme 6 (FP6) EURO-VO Data Center Alliance (EURO-VO DCA) and the FP7 EURO-VO Astronomical Infrastructure for Data Access (EURO-VO AIDA). OmegaCEN team members are members of its Board, Workpackage Management Team and Internal Science Team. In the framework of these programs OmegaCEN developed the infrastucture to access data residing in Astro-WISE through the Virtual Observatory. To access the Virtual

Observatory resources from within Astro-WISE a package was built using the Plastic/SAMP communication protocol (PhD Buddelmeijer). This is part of OmegaCEN's R&D track to build a query-driven visualization infrastructure for very large, multi-dimensional datasets.

The Astro-WISE system is ready to be filled with 100's of Terabyte of imaging data by the teams of public and GTO surveys of VST/OmegaCAM and VISTA/VIRCAM. The Astro-WISE astrometric pipeline is thoroughly tested (Astro-WISE Astrometry Report, McFarland et al., 2008). The experience with processing and quality control of an extensive subset of CFHT Legacy Survey, in the summer 2009, was used to write a detailed OmegaCAM Surveys Operations Plan. This Operations Plan will be applied to the survey operations for the KiDS 1500 square degree survey and its VISTA near-IR complement, the VIKING survey (5 filters: Z,Y,J,H,$K_s$).

## Astronomical research

Astronomical research using the Astro-WISE system in the Netherlands was carried out by Valentijn, McFarland, Belikov, Verdoes Kleijn, Peletier, den Brok, Sikkema, Buddelmeijer, Nelemans and Spooren. In Groningen it focused on the area of galaxy evolution as a function of environment: in the Coma cluster using the HST/ACS Coma Legacy Survey by PhD Den Brok (Carter et al., 2008, Hammer et al., 2009, Hoyos et al., 2010); in nearby superclusters and field using a WFI survey by PhD Sikkema ("The influence of the environment on the evolution of galaxies", PhD defended March 2009) and of large volumes of the nearby universe using UKIDSS/SDSS catalogs (PhD Buddelmeijer). Further details can be found in the relevant sections of this bi-annual report.
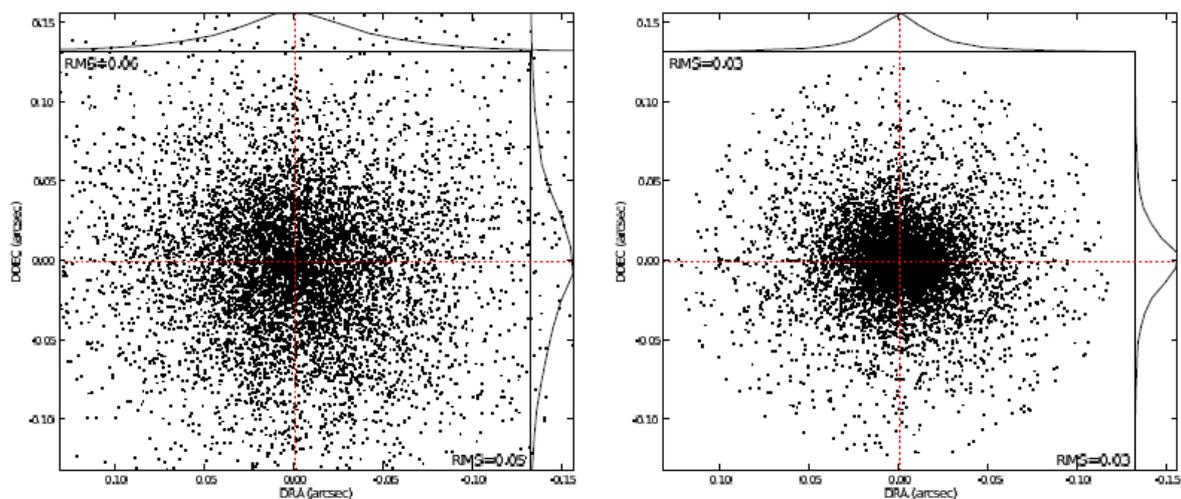


**Fig.2:** *Astro-WISE Global Astrometry at work. Improvement from the local (chip-by-chip) astrometric solution to the global (all chips and point-ings at once) astrometric solution. The two-dimensional RMS of the source position residuals from astrometrically corrected frames improves from 0.077 arcsec for the local solutions (left panel) to 0.041 arcsec for the global solution (right panel).*

## R&D Astronomical Information Technology

The goal at OmegaCEN in this research area is to enable astronomers to efficiently run their mining algorithms, specialized processes and to publish results for a Petabyte-sized ocean of data. All steps are executed while keeping track of the processing configuration and propagation of data.
Research is performed on Extreme Data Lineage in a distributed environment (Nuffic-PhD Mwebaze). The aim is to exploit data lineage in scientific pipelines to the sub-image level and eventually couple applied methods and programs to the data in a database.
A review of the usage of data lineage in scientific processing and OmegaCEN's approach was published in a refereed computational science journal (Mwebaze, Boxhoorn & Valentijn, 2009).

Query-driven visualization of very large catalogues was developed by PhD Buddelmeijer in collaboration with the Scientific Visualization group of the Computing Science department (PhD

Ferdosi).The goal is to give within the visualization process direct access to multi-dimensional 10-100Tbyte datasets, which current visualization techniques cannot handle 'stand-alone'. It is leading research in a new domain called 'query driven visualization' which is currently emerging. A first prototype for query-driven visualization was designed and implemented in Astro-WISE.

In 2009, Sikkema started a postdoc position funded by EURO-VO AIDA to cross-correlate large astronomical catalogues. It combines on-line catalogs using both VO and Astro-WISE data-access infrastructures. To ensure scalability up to very large datasets the associations are done in the Astro-WISE database using dedicated association algorithms. A pilot version has been completed in 2009 and is being used in science cases. This development will lead to a VO-compliant web service for cross-correlation of large catalogues such as produced by the KiDS and VIKING surveys.

The unique e-science aspects of the system triggered other parties to use it as a platform as well. In particular, the novel concept of 'entire backward chaining' (the linking of data products to the input raw data) has drawn attention as it provides a backbone for the infrastructure for advanced distributed e-science. This development has culminated in a large spin-off programme in 2009: the Target programme. This programme is coordinated by Valentijn. Target is a collaboration between 5 science/research groups (OmegaCEN, Astron/LOFAR, RUG's Artificial Intelligence (RuG), RUG's Donald Smits Center for Information Technology and University Medical Centre Groningen) and 5 high-tech business enterprises (including IBM and Oracle). The Target partners drew up an innovative 32 Million Euro 4 year program. The Ministry of Economic Affairs, Samenwerkingsverband Noord Nederland (SNN), Sensor Universe, the provinces of Groningen and Drenthe and the city of Groningen have awarded the programme with grants over 16 million Euro. The rest is fully co-financed by the partners. Scientific research groups and businesses jointly develop and improve complex and scalable information systems aimed at very large datasets for scientific and commercial usage. The programme has been presented at numerous occasions, ranging from Sensor Universe Bestuurlijk platform, SSN programme committee, XLDB meeting (Lyon), Nikhef (Amsterdam), Akkoord van Groningen (mayor and city council) to privately to Prime Minister, Balkenende.



*Fig.3:* meeting with the board of the University at the awarding of the grants. From left to right: prof. dr. S. (Sibrand) Poppema (President University of Groningen), Prof.dr. Edwin Valentijn (Head OmegaCEN and project coordinator Target), drs. R.F. (Robert) Janz (General manager Donald Smits Center for Information Technology), prof. dr. K. (Koos) Duppen (Member of the Executive Board of the University).

Target is participating as a funded partner in the long term EU program "Enabling Grids for E-Science in Europe (EGEE)" and its continuation (EGI) and hosts the Low Frequency Array (LOFAR) application of the EGEE-3 Astronomy and Astrophysics cluster. Within the Target programme OmegaCEN designed and built the Long Term Archive for LOFAR. The prototype was delivered in December 2008 (Valentijn & Belikov 2009) and the archive in December 2009. A first rehearsal, filling

the Long Term Archive while off-loading the Lofar central processor was conducted successfully in november 2009. Within the Target framework, OmegaCEN led the data-handling study for the proposed ESA Class-M mission EUCLID. The Euclid Science Study team, of which Valentijn is a member, delivered the Euclid study report for ESA's down-selection procedure.
OmegaCEN is also member of the consortium for the MICADO instrument, which is proposed for the E-ELT, and delivered the first design for its data flow system (Phase A review held December 2009).

An international panel, chaired by O'Mullane (ESA), reviewed OmegaCEN in the fall of 2008. Its assessment was very positive. They pointed at two critical areas: understaffing of the center (now counterbalanced by the Target programme) and the lack of an operation plan for VST data handling (now produced and delivered - see above).

**Workshops and education**

OmegaCEN was (co-)organizer of two workshops on astronomical surveys and information technology at the Lorentz Centre in Leiden: "From exoplanets to galaxy clusters: science with Astro-WISE" in April 2008 and "I-science for Astronomy" in October 2008.
OmegaCEN's co-organized the lecture course "Virtual Observations" in Sep-Nov 2008 and gave a 3 day Astro-WISE tutorial at Leiden Observatory in August 2008.

**Publications 2008 & 2009**

**Refereed journal publications**
1. Begeman, K., Belikov, A.N., Boxhoorn, D., Dijkstra, F., Valentijn, E.A., Vriend, W.-J., Zhao, Z., 2010, Journal of Grid Computing, in press "Merging Grid Technologies: Astro-WISE and EGEE"
2. Begeman K., Belikov A.N., Boxhoorn D.R., Dijkstra F., Holties H., Renting G.A., Valentijn E.A., Vriend W.-J., Zhao Z., 2009, in FGCS (Future Generation Computer Systems), submitted, "LOFAR Information System"
3. Boulanger F., et al., 2009, Experimental Astronomy, 23, 277, "The molecular hydrogen explorer H2EX"
4. Carter, D., et al. 2008, ApJS, 176, 424, "The HST/ACS Coma Cluster Survey: I - Survey Objectives and Design"
5. Hammer, D., Verdoes Kleijn, G., den Brok, M., et al., 2009 , submitted, "The HST/ACS Coma Cluster Survey: II - Data Description and Source catalog"
6. Hoyos, C., den Brok, M., Verdoes Kleijn, G., et al., 2010, in preparation, "The HST/ACS Coma Cluster Survey: III - Catalog of Structural Parameters of Galaxies using single-Sersic Fits"
7. J. Mwebaze, D. Boxhoorn, E.Valentijn, 2009, Proceedings of the 2009 International Conference on Network-Based Information Systems (12th NBiS), Indianapolis, 475,"Astro-WISE: Tracing and Using Lineage for Scientific Data Processing"
8. van der Zant, T., Schomaker, L., Valentijn, E.A., 2008, in Document Recognition and Retrieval XV, SPIE Conf. Series, 6815, 27, "Large scale parallel document image processing"

**Other publications**
1. McFarland, et al., 2008, The Astro-WISE Astrometry Report, http://www.astro-wise.org/Public/Astro-WISEAstrometryReport.pdf
2. Sikkema, 2009, PhD, The influence of the environment on the evolution of galaxies
3. Spooren, O., 2009, MD thesis, "AM CVn survey in Astro-WISE"
4. Valentijn E.A., Belikov A.N., 2009, "LOFAR Information System Design", Proceedings of Euro-VO DCA workshop "Grid and the Virtual Observatory", Mem.Soc.Astr.It. 80, 509
5. Verdoes Kleijn G., Belikov A., McFarland J., Valentijn E., 2009, "Bringing Image Data Quality to the Virtual Observatory: an Astro-WISE example", In ESA Conf. Proc. "Multi-wavelength Astronomy and Virtual Observatory", 155

**More information**
OmegaCEN: www.astro.rug.nl/~omegacen
Astro-WISE: www.astro-wise.org
Target: www.rug.nl/target