Table 1.3 c List of milestones

Milestones are control points where decisions are needed with regard to the next stage of the project. For example, a milestone may occur when a major result has been achieved, if its successful attainment is required for the next phase of work. Another example would be a point when the consortium must decide which of several technologies to adopt for further development.

Milestone number	Milestone name	Work package(s) involved	Expected date [1]	Means of verification[2]

[1] Measured in months from the project start date (month 1).

[2] Show how you will confirm that the milestone has been attained. Refer to indicators if appropriate. For example: a laboratory prototype completed and running flawlessly; software released and validated by a user group; field survey complete and data quality validated.

Table 1.3 d1: ALBiUS – Algorithms for Long Baseline interoperable User Software

Work package number		**Start date or starting event:**							
Work package title	Algorithms for Long Baseline interoperable User Software								
Activity Type[1]	RTD								
Participant number									
Participant short name	JIVE	ASTRON	Cambridge	ESO	MPIfR	NRAO	Oxford	Manchester	Bordeaux
Person-months per participant[2]**:**	60	30	24	30	12	24	20	36	24

Objectives

To develop key algorithms required for the successful exploitation of recent technological advances in radio astronomy, which have led to both an explosion in data rates and an expansion in the spectral window observable by sensitive interferometers. This includes providing new algorithms, developing existing algorithms to make them available for a new generation of instruments, and providing interoperability to allow better exploitation of the existing software packages. The latter will encourage a more unified approach to software development in radio astronomy across Europe and beyond.

Description of work

Background

Through the ALBUS project, RadioNet has made a concerted and sustained effort to substantially improve and develop the main user interface to one of its major products, the data of its interferometers. The project has initially concentrated on calibration algorithms and will continue with work on parallel processing for the remainder of the FP6 project. It has taken shape in the distribution of ParselTongue, a Python interface to classic AIPS. The project has successfully mobilised a group of radio-astronomers around Europe to work on interferometry data analysis in a distributed project. We propose to build on this collaboration and develop the synergy with the new generation of telescopes within FP7.

By 2009 the cm radio telescopes (e.g. eEVN and eMERLIN) will have completed the revolution in output data rates that provided the initial motivation for the ALBUS project. On the same timescale, the frequency coverage of interferometry instruments will be enhanced by at least two new major interferometer facilities: LOFAR and ALMA. Both of these initiatives are focusing on streaming processing of their data, in a system based on the original aips++ library. The enormous output

[1] Please indicate one activity per work package:
RTD = Joint Research Activity;; MGT = Management of the consortium; COORD: Coordination activity; SUPP: transnational access activity or service activity; .
[2] except human effort already included in the calculation of the access costs.

datarates of this new generation of instruments will provide an entirely new set of challenges for data processing, with automated processing and access to distributed computing environments becoming essential.

How will this look in 2010 from the user's perspective? The expert radio-astronomer may be looking at the new instruments and hoping to run experiments on these too, using tools and tricks established on cm instruments. Or perhaps she is wishing to use the beautiful new tools developed for LOFAR and ALMA on her cm-wave data too? One of the major obstacles facing the new software packages is a lack of trust in the newly implemented algorithms compared with the tried and tested versions available in the older packages. The lack of interoperability between the various packages results in a situation where astronomers are unwilling to invest time in learning a new package until they are convinced of the reliability of *all* of the algorithms required for their data analysis in that new package. Providing interoperability allows for a more piecemeal introduction to the new software, which in turn allows for confidence in the new software to be gradually developed.

The major aim of ALBiUS is to address these issues and to pursue a coherent and uniform approach to software use and development across all the major radio interferometer facilities. The assumption is that there will always be users who want to fine-tune the parameters of their calibration and imaging. The main goal is to salvage the existing algorithms in the era of ALMA and LOFAR, as well as for the SKA. At the same time we look at upgrading these algorithms for the new instruments.

Task 1 Interoperability

Portable Algorithms

We propose to carry out a pilot project to transparently process a dataset using software from multiple packages. There is scope to exploit the commonality in the Python user interface of AIPS, Casa, MIRIAD and GILDAS to implement this. The major issue is the variation in the data formats and calibration models used by the different packages. A simple approach is to convert between the different data formats using an intermediate data format which is already accessible to all the packages, such as UV FITS. An alternative approach would be to adopt HDF5 as the basis of an interoperable data format. The interoperable packages would then either have to be made HDF5-aware, or else be provided with a conversion routine from HDF5 to their native format. Using HDF5 has the added benefit of giving access to additional HDF5 tools not found in the main interferometry packages, in particular for 3-D data visualisation. This work will be carried out at ASTRON, JIVE, ESO.

Distributed ParselTongue

ParselTongue was developed by the ALBUS project as a Python interface to AIPS. This is the main vehicle for making the algorithms developed in ALBUS available to the user community. It has also proved an extremely effective tool for pipeline data processing, and has been incorporated into the production environment of the EVN and MERLIN arrays. In addition, it is used by an increasing number of astronomers for processing datasets which are either large in size or consist of repeated observations, each requiring similar processing. We seek to develop the current ParselTongue functionality in the context of interoperability with other packages. Given the continuing dependence on AIPS of a large part of the radio astronomy community, even in this era of ever increasing data volumes, there is useful scope for enhancing ParselTongue to allow better exploitation of AIPS on a distributed computing environment. Some initial progress has already been made in ALBUS on creating an infrastructure that allows data distribution for parallel processing, but there are many problems which require more atomic procedures than currently available in the AIPS suite, which could be addressed. This work will be carried out at JIVE.

Task 2 - Calibration algorithms

Global Fringe Fitting

Global fringe fitting is a crucial algorithm for eMERLIN, eEVN, EVLA, extended (European) LOFAR baselines and the highest ALMA frequencies where an efficient means of determining residuals to the correlator model of the station-based delays is required. This algorithm is currently only implemented in AIPS, but will have to be incorporated into the new software packages if they are to be useful for these arrays. In addition, the current algorithm produces results which are difficult to interpret in an automated way due to peculiarities of the weighting scheme and resulting anomalies in the reported signal to noise ratio. Alternative methods to the non-linear least squares fit (Levenberg-Marquardt algorithm) used in AIPS are also likely to produce more robust results, less susceptible to converging on local minima. Similar methods are likely to be applicable to improve traditional self-calibration, the reliability and robustness of which will be of crucial concern to ALMA. We propose to evaluate the current fringe-fitting algorithm and provide an improved version for Casa. This work will be carried out at NRAO, JIVE, ESO.

Image Plane Calibration

In ALBUS, methods were developed for calibrating the effects of ionospheric and tropospheric fluctuations on long baseline interferometry data. These methods have been implemented in AIPS with some success. A limitation, however, is that the AIPS calibration model assumes that a single calibration factor is applicable across the instantaneous field of view of the interferometer. This assumption does not hold for the large fields of view and/or large fractions of the primary beam which will be observed by LOFAR, ALMA and APERTIF and which are becoming available for the EVN, eMERLIN and EVLA as the correlator capabilities of those instruments are enhanced. Polarization imaging over wide fields of view presents additional challenges for all these instruments and the polarization response of the dipole arrays used by LOFAR will have a particularly strong direction dependence.

These image plane effects can be corrected using some of the more modern calibration packages which are currently in development, but the calibration algorithms will require some development to incorporate the required direction dependence. We would also like to address some related calibration issues, such as correcting for station primary beams, and using mosaicing techniques to produce images covering areas of the sky much larger than is achievable with a single pointing, the latter being particularly important for ALMA with its relatively small instantaneous field of view.

High spectral and temporal resolution is required to achieve large fields of view as well as to allow high quality spectroscopy and the study of transient phenomena (both astronomical sources and contaminants such as RFI). This implies very large data sets: 100's of GBytes for EVN and several TBytes for LOFAR (see Table 1 for an estimate of the output data rate of several instruments). Efficient processing of such large data sets requires access to a distributed computing environment and we would therefore seek to modify the existing algorithms to enable this. This will likely also require the development of support for distributed computing in the chosen calibration packages which currently have only limited support for parallel processing.

Solving this issue will be important for LOFAR and ALMA, but suitable test datasets from GMRT, Westerbork and the EVN already exist. This work will be carried out at Cambridge, Oxford, Manchester, ASTRON, NRAO, ESO.

Calibration of Astrometric Source Positions

Having accurate source positions is essential for many radio astronomy applications, among which is high-resolution interferometric imaging of weak targets. At present, such measurements are carried out with two different techniques: wide-angle astrometry based on group delays (for major calibrator sources distributed on a 5^o grid on the sky) and narrow-angle astrometry based on phase-referenced measurements (for sources in the inter-space between the major calibrators). We seek to develop innovative algorithms and observing techniques that will combine the two approaches to produce an improved, denser, and unified celestial reference frame comprising all sources (whether observed with wide- or narrow-angle astrometry).

Ultimately, these new methods should serve as a basis for conducting deep astrometric surveys with instruments such as the EVN, eMERLIN or the IVS2010 network. Algorithms such as those used in traditional geodesy to adjust positions of second-order geodetic markers into a first-order grid of markers may be adapted to realize such a global adjustment. The work will include studies of potential algorithms, simulations to test these algorithms, and analysis of test data acquired with the EVN and eMERLIN to validate observing strategies. This work will be carried out at Bordeaux.

Task 3 - Tools for Large Datasets

Automated Data Quality Control

The new generation of interferometers currently coming on line (LOFAR, eEVN, eMERLIN, EVLA, ALMA) will have vastly increased output data rates compared with current instruments. Traditional methods for data inspection will become impractical for these new instruments. It is therefore necessary to develop new techniques, such as subspace decomposition, to allow automated identification (and either correction or excision) of artefacts resulting from poor calibration, poor atmospheric or ionospheric conditions or other problems.

Some methods to mitigate RFI are known to be very effective but are not yet implemented at many radio observatories. Sophisticated approaches for arrays like the SKA and focal plane arrays are still to be fully explored. We would like to investigate the usefulness of subspace techniques and other existing algorithms and to make them available in interferometry, where appropriate. We also seek to develop further RFI excision methods for focal plane arrays (e.g. APERTIF), perhaps using cross correlation between elements to derive template spectra of the RFI for subtraction from the astronomical beams. The most useful algorithms would then be implemented in one (or more) of the mainstream interferometry analysis packages. This work will be carried out at MPIfR, Manchester, ASTRON, Cambridge, Oxford.

Source Parameterization

One of the primary scientific deliverables of LOFAR and APERTIF will be huge catalogues of all the detected sources. This so-called Global Sky Model will also provide the basis for the calibration, as a good sky model will be required to make tractable the complex calibration needs of these instruments. Most sources will be unresolved, so simple point-source models can be used, but a significant fraction will be (sometimes very) extended. The vast size of the Global Sky Model means that new and innovative methods must be developed to

1. automatically extract from the data an accurate description of the sources in a computationally efficient way.
2. store the source descriptions in a form which is compact and yet retains sufficient information to allow an accurate reconstruction of the objects described in it.

Traditional methods for doing this have involved describing sources in terms of collections of point sources (CLEAN components) or as collections of elliptical gaussians. However, these methods are neither very compact nor can they be easily extended to incorporate a complex time or frequency dependence. They are also limited in the dynamic range that they can achieve, which would prove a limiting factor for instruments such as LOFAR or APERTIF. A possible solution is to use techniques based on shapelets or pixons.

Such techniques can describe extended sources with relatively few parameters to high accuracy and, given their continuous nature, do not suffer from the problems caused by the discreteness of CLEAN-component models and will be able to achieve much higher dynamic range. We propose to develop an optimal means of producing a Global Sky Model based on these techniques.

In order to derive maximum value from the Global Sky Model, it would be beneficial to develop this in the context of the Virtual Observatory and also to provide a link back to the experiment preparation stage via a proposal tool such as NorthStar. We will seek to develop this connection, defining the necessary interfaces to enable access to the information required for optimal experiment planning. This work will be carried out at ASTRON, Cambridge and Oxford.

Source Modelling

Spectral line observations, particularly of complex chemicals such as will be observed with ALMA, pose a particular problem for interpretation of the source parameters extracted from the data cubes. This is further complicated by the large data volumes that ALMA will produce. We propose to develop a standardised optimization interface to the existing spectroscopic modelling codes to enable automated modelling of the physical parameters of large numbers of sources and the comparison of these with theoretical models. The modelling codes will be developed to be compatible with this interface which will be documented and distributed to facilitate development of new algorithms. This work will be carried out at ESO.

Deliverables (brief description and month of delivery)

RFI mitigation software – month 18

New implementation of Global Fringe Fitting algorithm – month 24

Final report on calibration of pilot experiment using interoperability framework – month 30