RI031675

EuroVO-DCA

The European Virtual Observatory Data Centre Alliance

COORDINATION ACTION

RESEARCH INFRASTRUCTURE

COMMUNICATION NETWORK DEVELOPMENT

D4 - Medium term strategic plan for EuroVO-DCA

Due date of deliverable: 31/11/2008

Actual submission date: 27/02/2009

Start date of project: 01/09/2006 Duration: 28 months

CNRS Final Version

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

AUTHORS

Editor: Françoise GENOVA

Authors : Françoise GENOVA, Mark G. ALLEN, and the EuroVO-DCA Board: CHRISTOPHE ARVISET, Paolo PADOVANI, Fabio PASIAN, Enrique SOLANO, Edwin VALENTIJN, Wolfgang VOGES, Michael G. WATSON

Comments from Andrey BELIKOV, Andrew LAWRENCE, Anita M.S. RICHARDS and Jonathan A. TEDDS are gratefully acknowledged.

EXECUTIVE SUMMARY

This document is the final deliverable of the Euro-VO Data Centre Alliance project (EuroVO-DCA), a Coordination Action of the Sixth Framework Programme. It outlines the medium term strategy for European data centres in the context of implementation of the astronomical Virtual Observatory. The medium term strategy builds on successful construction and use of knowledge infrastructure in astronomy and the experience gained from the EuroVO-DCA coordination action. Importantly, this strategy is interfaced with the establishment of a European-wide strategy for the entire discipline, and its potential impact on the on-going conceptualisation of the knowledge infrastructure.

Astronomy has traditionally been at the forefront of the implementation of scientific digital repositories, with a culture of widespread access to remotely available information. Data Centres have a long history, and provide a wide range of services to the astronomy research community. They archive and provide access to the data obtained by ground- and space-based observatories, and provide services and tools for making use of and analysing these data. In addition to observationally based services, there are also data centres that provide theoretical models and results of simulations, reference and bibliographic services, and services centred on a given scientific theme. The Virtual Observatory (VObs) is a recent new step, aiming at providing seamless access to the wealth of available distributed resources. It has been developed from the beginning as a fully international endeavour, with an international alliance of VObs projects (the International Virtual Observatory Alliance or IVOA) to define the domain-specific interoperability standards.

Data Centres are a major component of the Virtual Observatory, populating the system with data and services. One main aim of EuroVO-DCA was to characterize the population of European data centres, first by defining data centres in the VObs context, then by performing, for the fist time, a European-wide data centre census collecting information in a uniform way. The census provides a lively snapshot of the very diverse landscape of European astronomical data centres, which covers all scientific areas of astronomy, with a large variety of different approaches to delivery of data and services. A number of large data centres host multiple archives and services, but a significant number of small and less well resourced data centres provide important often specialised archives, services and tools. The census shows a high level of interest in Virtual Observatory methods, with many data centres already making some use of VObs access protocols, and indicating intent to implement IVOA standards. The census has been analyzed to derive requirements on the VObs projects, which should be taken into account to accommodate the high level of diversity in the data centres.

The EuroVO-DCA project has fully fulfilled its objectives, and beyond. The Data Centre Alliance is one of the rare operational examples of a discipline-wide community of fully distributed knowledge infrastructure providers, well identified via the census. EuroVO-DCA has also paved the way for the Seventh Framework Programme *Euro-VO Astronomical Infrastructure for Data Access* project.

Data centres and the VObs are fully recognized as an essential infrastructure for research in the global strategy for European astronomy recently established by the Astronet ERA-NET, and requirements for medium term sustainability have been derived from the practical experience of data centres and of the VObs. The main efforts for mid- and long-term sustainability will fall on the national funding agencies and on the intergovernmental organisations (ESA, ESO). There is also a role here on the medium term for EC, in particular in support to European and international cooperation efforts. This of course raises the question of sustainability of EC support, at least at the level of Coordination Actions.

Table of content

1. INTRODUCTION

This document is the 4th deliverable of the *Euro-VO Data Centre Alliance* (EuroVO-DCA) project, a Coordination Action of the Sixth Framework Programme. It shows how the future strategy of astronomical data centres is built on the already long history of data sharing in astronomy, which begun long before the advent of the World Wide Web, and took advantage of the new technical possibilities it offered.

The emergence and aims of the Virtual Observatory, and of its European incarnation, Euro-VO, are explained. The major role of data centres, and the Data Centre Alliance set up to support their take-up of the Virtual Observatory framework, are described. One main activity of EuroVO-DCA has been to characterize the population of European Data Centres by performing a census, and the results and implications of the census are described.

A summary of the lessons learnt at the practical level from EuroVO-DCA activities, and its potential impact on astronomy, and at a more general level in the conceptualization of the knowledge infrastructure, is provided, and the on-going transition towards operations through the Seventh Framework Programme project *Euro-VO Astronomical Infrastructure for Data Access* (EuroVO-AIDA) is briefly described.

Finally, the endorsement of the VObs strategy and accomplishments by the European global strategy exercise of the Astronet ERA-NET is indicated, as well as the practical requirements to ensure medium term sustainability of European data centres and the VObs.

2. DATA CENTRES IN ASTRONOMY

Astronomy has traditionally been at the forefront for the implementation of scientific digital repositories, with a continuum between raw and processed observational data stored in observatory archives, value-added databases, and results published in academic journals. Astronomers utilize these repositories at all stages of their scientific work, and a culture of open data access and of widespread usage of remotely available information is well established in the astronomy community. National and international ground- and space-based observatories produce terabytes of data, which are publicly available from observatory archives all around the world. Data centres and archives develop added-value services, such as high level ('science ready') data products, data access and analysis tools, and databases of results published in academic journals.

The implementation of astronomical data centres devoted to electronic information began early in the electronic age. For instance, the database containing the observations of the *International Ultraviolet Explorer* satellite of the European Space Agency[1], which operated between 1978 and 1996, has been a remarkable early precursor, to demonstrate the importance of public availability of data, allowing scientists to re-use it for scientific purposes which are often different from the original ones: roughly five times as many publications were based on archive data as those based on the original proposals (Wamsteker & Griffin, 1995). An early example of value-added database is SIMBAD, which compiles published information about astronomical objects (180 000

[1] The IUE satellite was a trilateral project with NASA and United Kingdom

queries/day on average in 2008). It has been created in 1983 by the merging of two data bases which had been developed at the beginning of the seventies and has been accessible remotely since then. SIMBAD has evolved significantly over the years, in phase with progress of astronomy and with technical advances, and is still now a major resource for the community. Since these beginnings, many agencies world-wide, in particular NASA in USA, have had a dedicated policy of long term support to data repositories and value-added databases. NASA requires organisations to make data publicly available as a condition of funding. Europe has set up several other successful archives of data obtained from satellites, such as the European Space Agency's ISO (infrared) and XMM-Newton (X-ray) observatories. Currently, Europe also hosts the world-leading archive of ground-based astronomical observations, the ESO archive, which collects observations carried out with ESO telescopes since the early 1990s, including the complete record of observations obtained with the Very Large Telescopes.

3. A WORLD-WIDE COMMUNITY OF DATA CENTRES: THE ASTRONOMICAL VIRTUAL OBSERVATORY ENDEAVOUR

The advent of the World Wide Web produced a revolution in the way astronomical data and information is distributed and used in the daily work of astronomers. Many data providers have implemented Web access to their repositories, and http links between them. This *data network*, which consists of huge volumes of highly distributed, heterogeneous data, opened up many new research possibilities for astronomers and greatly improved the efficiency of doing science. This enormous growth in volume and diversity of data does however present challenges for managing complexity which need to be met with a global vision for astronomical digital repositories.

Astronomy is at the leading edge for conceptualizing and implementing a coordinated approach to the deployment of digital repositories. The global vision comprises several key elements:

- An 'open' data policy – most observational data are publicly available after a proprietary period of one year in most cases;
- Agile and eager-to-evolve data repositories - contrary to the 'static' vision conveyed by the term 'repository' -, as long as the providers get the proper support and that the proposed evolution remains manageable;
- Coordination of efforts at the international level, with a bottom-up approach, in particular for the definition of interoperability standards. Great attention is paid to the participation of and feedback from data managers in the standardization process, and to a proper evaluation of new technologies, which have to be implemented not too early, since the goal is to build an operational system with some sustainability in mind, and not too late, to take the full profit of new possibilities;
- a fully science-driven approach: all evolutions are driven by the needs of the science community, and in turn the science community is motivated to use VObs-enabled data and tools because of the unique and powerful capabilities they provide.

One of the lessons learnt from implementing this data network is that in systems which facilitate access to data, proper definitions of data characterization and quality are required. Mechanisms for propagating this information are mandatory in order to allow high quality research using the data, and also to give the user enough confidence in the whole system – which means additional tasks on the data repositories.

The radical new step for providing a coordinated approach has been under way since the turn of the century with the rapid development of the Virtual Observatory (VObs) concept. This aims at providing seamless unified access to data holdings – all archives speaking the same 'language', accessed through uniform portals, and analysable by the same tools. This means giving all astronomers easy access to, and usage of, data gathered by the many astronomical sub-disciplines, well in pace with the rapid development of multi-wavelength astronomy, which now makes a significant fraction of the published papers.

VObs projects around the world have formed the International Virtual Observatory Alliance (IVOA, http://www.ivoa.net), which coordinates in particular the assessment of the VObs architecture and the development of domain-specific interoperability standards. The VObs goes one step further than giving access to distributed data repositories: the implementation of the interoperability standards (the *interoperability layer*) on the top of the data repositories, permits operations, such as data aggregation and combination, which are essential for the full scientific exploitation of the data infrastructure.

Figure 1: **The International Virtual Observatory Alliance**, showing the participant VObs projects (clockwise from the top: Europe (represented by ESA and ESO), China, India, Canada, Spain, Italy, Armenia, France, Germany, Hungary, Japan, Korea, USA, Russia, UK, and Australia)

The VObs is not a monolithic system, but, like the Web, a set of standards, which make all the components of the system *interoperable* – data and metadata standards, agreed protocols and methods, standardised mix-and-match software components. These standards and software modules constitute the *VObs Framework*.

Several strands of work are needed to fully implement the Astronomical Virtual Observatory:

1. Development of standards and protocols, and their international agreement, coordinated by IVOA;
2. Construction of "glue" software components - portal, registry, workflow, user authentication, virtual storage;
3. Uptake by data centres, who need to "publish" to the system, i.e. to write VObs compliant data services connected to their holdings;
4. Construction of tools to effectively take advantage of seamless access to data;
5. Support to the scientific community in its uptake of the new framework.

Data centres have thus several key roles within the VObs, such as:

- to provide astronomers with easy, long term access to observation archives
- to develop expert data centres which provide services to the community (e.g. reference databases and value-added services; databases and services focussed on specific domains of expertise; "virtual instruments" which allow users to obtain high level data products and to solve specific questions)
- to improve data quality, e.g., by providing calibrated data and improving calibration procedures in data pipelines, or by selecting data, homogenizing it, improving it by critical comparison, for compilation databases
- to develop and implement visualisation and data analysis tools
- to publish their data in the VObs framework, enabling the community to build on top of them added value services and tools, thanks to the existence of VObs interoperability protocols
- to implement theory-related services, such as making simulation data available in the VObs, i.e. with the possibility to re-use them and to compare them to observational data, or interfacing the VObs with massive computations – here in particular the astronomy data and service grid meets the computational grid
- to provide outreach to their science community and to the general public

Each data centre has its own speciality in terms of wavelength domain, observational technique, service provided, knowledge of a specific user community, development of access tools or data management methods, and will thus be able to contribute to increasing the common knowledge and to the feedback to standards and tools developers.

4. NETWORKING EUROPEAN DATA CENTRES: THE EURO-VO DATA CENTRE ALLIANCE

Europe is one of the major actors of the international VObs endeavour, through the national VObs projects and those of the two intergovernmental organisations ESA and ESO. The European VObs activities are focussed through the European Virtual Observatory (Euro-VO) consortium, which is also a member of the IVOA. Euro-VO has 8 member organisations, comprising the European intergovernmental (ESO, ESA) and national research organisations and VObs initiatives.

Each partner potentially brings to the VObs systems different kinds of data centres with different management methods, and different technical expertise. Each national node and agency has its own organization, which depends on its priorities, which in turn depend on its funding sources, and on the way astronomy is organized in each country. In some countries, for instance, most VObs resources are provided through a dedicated project, in others, there is a national coordination of the resources, which are provided by individual institutes and laboratories ; in others, coordination is provided directly by a national funding agency ; or agencies in charge of telescopes are themselves committed to implement the VObs layer on top of their data archives. Many combinations of the above are possible but the common characteristic is that all partners are strongly committed to VObs development. Similarly the balance between the implementation of data centres and technical work on the VObs framework is different for the different partners.

Euro-VO consists of three interacting elements, addressing the different strands of work described above:

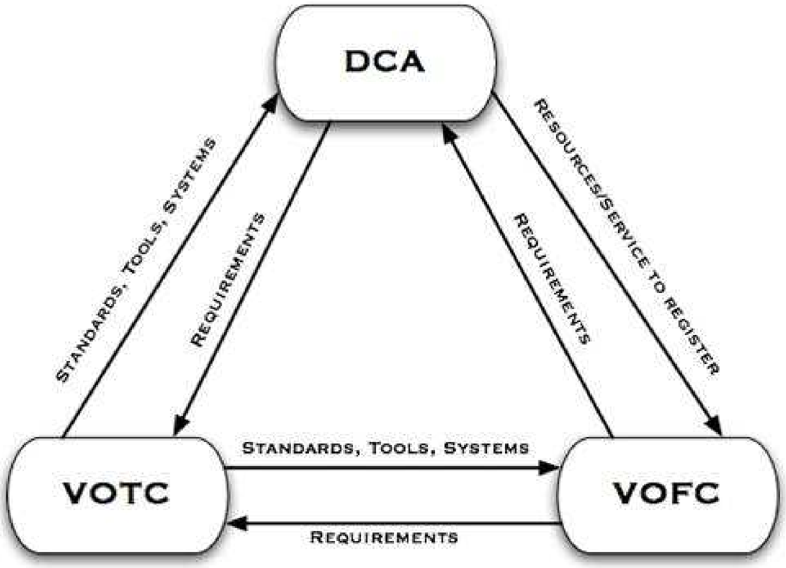

Figure 2: The Euro-VO project. The figure represents the three elements of the Euro-VO project

- **Euro-VO Data Centre Alliance** (DCA): a network of the European data centres, which populates the system with data, provides the physical storage and computational fabrics, and using VObs technologies, publishes data, metadata and services to the Euro-VO;
- **Euro-VO Technology Centre** (VOTC), a distributed organization, coordinating a set of research and development projects on the advancement of VObs technology, systems and tools;
- **Euro-VO Facility Centre** (VOFC), which provides the Euro-VO with a persistent, centralized registry for resources, standards and certification mechanisms as well as community support for VObs technology take-up and dissemination and scientific program support using VObs technologies and resources. The VOFC provides a "public face" to the Euro-VO.

The **Euro-VO Data Centre Alliance** project (**EuroVO-DCA**) has coordinated the first integration of the European data centres in the VObs framework. EuroVO-DCA (RI031675) is a Coordination action of the Sixth Framework Programme *Communication Network Development* framework, which ran for 28 months from September 1st, 2006 to December 31st, 2008. Its objective was "*Technology take-up and full VObs compliant data and resource provision by astronomical data centres in Europe*". The project had several facets: dissemination of knowledge and good practice among the data centres, gathering their feedback, preparing the inclusion of theoretical services in the VObs, exploring the relations with the computational Grid, and fostering participation from data centres from other European countries. Feedback from implementation and input on standards has been conveyed to the VOTC and to the IVOA.

This assessment of a medium term strategy for European astronomical data centres is a major product (Deliverable D4) of the EuroVO-DCA project.

5. ANALYSING THE POPULATION OF EUROPEAN DATA CENTRES: THE EUROVO-DCA CENSUS

5.1. Data Centres in the Virtual Observatory

One of the goals of the EuroVO-DCA project was to identify the population of European astronomical data centres, which were the target of its activities. The first step was to define data centres in the VObs context, and we opted for an inclusive definition built on key concepts, as follows:

> *Data Centres are an essential component of the Virtual Observatory, publishing data, metadata and services, and providing the physical storage and computational fabrics. The VObs development is a strong incentive to share data and knowledge, and many teams are willing to provide data and services in their domains of expertise. 'Classical' Data Centres, such as ground-and space-based observatory archives, and generalist Data Centres, are key providers of added-value services and tools. More and more teams are willing to join with value-added services and tools in specific domains, and VObs 'Data Centres' work in very different contexts - national or international Agencies, scientific laboratories - , and are highly diverse in size and objectives, from small and specific to large and general. Common keywords are the willingness to* **provide a service to the community***, provision of* **added-value** *built on expertise, some kind of* **sustainability***, and concern for* **quality***. Lessons learnt from the long term history of astronomical Data Centres show that when beginning these activities, critical parameters are in particular* **having a critical mass** *adapted to the goals, and ensuring* **medium-term sustainability***, which requires at least a strong support from the local authorities. An important factor to win community support, which is indispensable to secure funding, is to find a* **national and/or international niche***.*
>
> *Many types of contribution are possible: data archives, with a particular emphasis put on 'science ready' data; added-value data bases, services; tools, software suites and algorithms, for instance for data visualisation, data analysis and data mining; thematic services to help solving a well-defined science question; full data analysis or research environments. New types of services are emerging, with in particular theoretical services, providing modelling results, or matching models with observations.*

5.2. The Census

A preliminary census was built at the beginning of the project, using information provided by the partners, and was available at the end of the kick-off phase. It also included, from this early stage, data centres from European countries which are not among Euro-VO partners, and allowed a first assessment of the data centre definition described in the previous section.

With our inclusive definition of what constitutes a Data Centre, the census was organised into five questionnaires, established under the responsibility of the Project Scientist and the Project Manager, with input from all project bodies. It aimed at capturing information about the broad scientific areas dealt with by data centres, as well as the nature of their data holdings. The respondents were asked to identify their data centre in a first questionnaire (*Introduction and identification of the Data Centre*), and then invited to fill as many forms as needed to describe their data holdings and services.

One of the goals was to identify data centres in the emerging field of theory, and two specific questionnaires were available for this purpose. Four forms were available:

- Observational archives and data products (questionnaire 2)
- Services/Tools/Software suites (questionnaire 3)
- Theoretical archives (questionnaire 4)
- Theory services (questionnaire 5)

The census questionnaires and results are described in Deliverable 5 of EuroVO-DCA, **Census of European Data Centres**. The census forms are accessible at http://cds.u-strasbg.fr/twikiDCA/bin/view/EuroVODCA/DCACensusQuestionnaires.

5.3. Summary of the Census results

The reader is referred to D5 for a detailed presentation of the census results. A summary is given here and in the following section.

The response of the astronomy data centre community to the census was very positive with a high level of participation: 69 data centres answered the census, and a total of 235 individual forms were filled (134 for questionnaire 2, 64 for questionnaire 3, 24 for questionnaire 4, 10 for questionnaire 5). The data centres in the census are located in 16 different countries, and two of the data centres (hosted by ESA and ESO) are international organisations. The distribution of the data centres across Europe appears to be affected by the different levels of VObs awareness in different countries, and also by the different levels of organisation of data centres in different countries.

The census shows as expected a wide diversity in the scale, content and function of astronomy data centres, from very large data centres responsible for the archives of major facilities, to small teams which provide data or services related to a specific scientific theme, with also some which curate data and provide reference services. The data centres cover essentially all scientific areas of astronomy as the instruments and missions are relevant to a wide range of scientific research. The largest archives in terms of data volume reported in the census are those of modern wide field instruments which have volumes in the order of 100s of Terabytes. Other similarly vast archives are expected for the ESA Gaia mission, ESO VISTA surveys, and other next generation instruments.

Overall the quality of the responses and the relatively high participation in the census gives us confidence that the results provide a useful characterization of the European Data Centre community. It appears that the VObs development has pushed towards open data access and the establishment of new, often small data centres willing to share their expertise with the scientific community

6. IMPLICATIONS FOR THE VIRTUAL OBSERVATORY AND FOR THE DATA CENTRES

This census identifies the community of European data centres who will most likely participate in, and lead, the up-take of VObs methods as Euro-VO moves into an operational phase. These data centres are already making use of the current IVOA standards, and report a very high level of intent to use IVOA standards. The census strengthens the identification of these data centres as a community, and provides Euro-VO with the basis for addressing their common needs.

6.1. Implications for the Virtual Observatory

As explained in the Census document, the information collected in this census will be used in the planning of future Euro-VO assistance programs to help data centres publish their data and services to the VObs. The information is also useful for the development of the VObs standards and will provide valuable feedback to the IVOA. Here we highlight a number of immediate implications for the VObs in general, that can be drawn from the astronomy data centre community response to this census, and which constitute important aspects of the medium term strategy:

- There is a high level of interest in the VObs, with many data centres already using VObs access protocols and aware of registering VObs resources in a VObs registry. Many data centres indicated the intent to implement VObs standards, and this includes the larger data centres and those holding the major archives and services. Some of the data centres are already organised around VObs technologies. It will be of great benefit to continue the close coordination of the data centres and the VObs development, so that the priorities for interoperability standards are in step with the requirements of the data centres and the needs of the astronomy community.

- Most data centres report that the current interfaces to their archives and services are web interfaces. VObs-enabled services should be accessible through standard web browsers, as well as using client applications. In addition, the up-take of VObs technologies in data centres will be facilitated by ensuring that implementation of VObs standards is not significantly more difficult than implementing a web interface. Moreover it is clear that data centres on average do not have significant resources for these purposes so it will be important to make it easy to implement VObs standards. This could be facilitated by the provision of tools enabling more easily this up-take.

- Most data centres usually already have their data and metadata organized and characterized in their own way to fulfil their project's and users' purposes. Adopting the VObs framework must be like implementing a VObs layer on top of their existing data and metada infrastructure, enabling data centres to publish their data holdings into the VObs on top of their existing organization. As above, this could be facilitated by the provision of tools enabling the implementation of this VObs layer on top of existing infrastructure.

- Data centres should be motivated to publish their data into the VObs framework realizing that the VObs will enable greater access and usage of their data, increasing then their scientific return on their initial investment.

- The data centre responses to the census reveal a wide range of data volume, data types, service functions and interfaces, and covering all areas of astronomy research. It will be important for the VObs in general to have strategies to accommodate this kind of diversity in the VObs. Firstly this requires identifying the most common elements, such as the fact that most data types fall within broad definitions of images, catalogues and spectra. Also that many observationally based data centres support data queries in terms of sky coordinates, and that the physical units used describe quantities can be expressed in a common (or convertible) system. These common aspects have driven much of the VObs standards development to date. To better manage the detailed descriptions of data and services requires hierarchical descriptions such those being explored within the data modelling efforts of the IVOA. The VObs in general will need to maintain an appropriate balance between simplicity and complexity, so that the detailed capabilities of data centres can be accommodated within the VObs framework, alongside of relatively simple interfaces.

- In terms of common data types and requirements on future developments, the census indicates that effort will be needed to address data types in the areas of photometry and time series which followed images, catalogues and spectra as the most common data types indicated by data centres. Theory data archives often have different types of data products compared to observational archives. The full list of data types collected in the census will be of use to IVOA efforts for data modelling, and development of standards for simulation/theory data.

6.2. Implications for the contents and sustainability of data centres

The majority of data centres report that they provide along with raw, or lower level data products, some science ready data products. The goal of building the VObs as a framework for multi-archive science relies heavily on the scientific usability and interoperability of the products provided by data centres. As such the emphasis on data centres providing 'science ready' data must be increased, which puts an additional, but very worthwhile, requirement on the data providers.

Most of the data centres expect their archives to have very long lifetimes. While some data centres indicated that they contain collections of photographic plates, essentially all the data described in this census is electronic, with a strong correlation of data centre establishment date and evolution of the internet. As explained above, there is also some effect on the number and nature of data centres related to the development of the VObs which has triggered some efforts to share and publish data and services, even in small research teams. Long term preservation of such data may become more relevant to VObs community. In general, the sustainability of data centres will need to be addressed. The census collected brief information on the human resources supporting the Data Centres. While the numbers reported here vary in terms of the range of functions this includes, we note that many data centres are supported by less than 1 FTE. For larger data centres significant support has to be found in the long term.

7. EUROVO-DCA IMPACT

7.1. Lessons learnt

In the European view of the VObs, the Data Centre Alliance is one of the three 'pillars' of the VObs. The project has succeeded in assessing the DCA relations with the Technology Centre and the Facility Centre, in preparation for the operational phase. The importance of gathering requirements from feedback and implementation has been demonstrated, as well as the need for tools and tutorials, and the high impact of well planned workshops.

The combination of top-down and bottom-up approach has proven to be efficient to create a European-wide community of data centres, beyond the project partners. This also provides a forum for discussion and the sharing of best practices.

Moreover, the project has deepened the links and collaboration between the national and agency VObs project, which increases strongly the global impact of European activities. The EuroVO-DCA project is an excellent example of the possible impact, at the European, national and international levels, of EC-funded Coordination Actions – it also shows their limits since technical and service work were also needed to bring the VObs to its operational phase.

7.2. Impact on astronomy

EuroVO-DCA successful activity in helping data centres to take-up the VObs framework will trigger integration of more data in the VObs. We expect this to have a major impact on the daily research work of astronomers by giving them seamless access to the wealth of data and information available.

As explained above, EuroVO-DCA has also built a community of data providers, which will likely facilitate their work in particular by the sharing of good practice. The census and participation of recently formed data centres in the project activities also showed that the emergence and dissemination of the VObs concept has led many teams to decide to join, and to share their knowledge with the community by providing on-line data and services, and by interfacing them with the VObs framework.

Feedback gathered by the project has also helped the VObs teams to improve the relevance to user needs and usability of the VObs framework.

At a more global level, the remarkable success of the activities towards astronomical communities of non-partner countries has demonstrated that the VObs is, as expected, a very powerful integration tool for the scientific community, with a high impact in particular in the new EU members and in Associate Countries.

The impact of the VObs for astronomy has been well recognized by the European strategy exercise of the Astronet ERA-NET. The Astronet Infrastructure Roadmap recommends that data from all telescopes should be made available, and included in the VObs framework. This will be discussed in more details in the concluding section of this document.

7.3. Impact on the knowledge infrastructure conceptualisation

The community of data centres created by EuroVO-DCA can be seen as an operational example of community of knowledge infrastructure providers. This community is characterized by the very diverse range of data centres which emerged from the census exercise, with large and small teams working in very different contexts, from large agencies to small research laboratories, covering all fields of astronomy, connected together via the VObs. We consider this as one possible vision of the knowledge infrastructure, which puts into action real sharing of knowledge among a whole community. The fact that the astronomical VObs has been implemented from the beginning as an international endeavour produces a very appealing world-wide, discipline-wide data and service grid. This also exemplifies the major role of data and service providers, which are in charge of building the content and ensuring its quality. Sufficient resources will need to be provided to develop and maintain these activities which strongly contribute to the efficiency of research. The pragmatic definition of a data centre in the VObs context (recalled in Section 5.1. of this document) may find useful applications in other contexts.

The project has been keen to share the lessons learnt by participating actively in EC-organised 'Concertation' events, in the meetings of several other EC-funded projects, and by collaborating with OGF-Europe. An action on the assessment of the astronomy disciplinary model is on-going within the follow-up FP7 EuroVO-AIDA project, as explained in the next section.

At a more technical level, several other disciplines are assessing the possibility to reuse part of the IVOA interoperability standards. Contacts are going on in particular with groups in astroparticle physics, planetary studies, solar studies, atomic and molecular physics. Contacts with the climatology METAFOR project have been actively sought.

8. TOWARDS THE OPERATIONAL PHASE OF THE EUROPEAN ASTRONOMICAL KNOWLEDGE INFRASTRUCTURE: FROM EUROVO-DCA TO THE EURO-VO ASTRONOMICAL INFRASTRUCTURE FOR DATA ACCESS (EURO-VO AIDA)

EuroVO-DCA has fulfilled the objective of creating and coordinating a community of European Data Centres. It has also consolidated the cooperation between European VObs projects, and has disseminated the knowledge of the VObs in non partner countries, in which data centres are eager to contribute, and new national VObs projects are emerging.

The project has been closely interlinked, as expected, with the VObs development in Europe and at the international level. Feedback from EuroVO-DCA has been transmitted to infrastructure developers, IVOA standard developers, and other relevant software engineering staff within the VO-TECH project, other European VObs technology groups, and IVOA interoperability Working Groups. The **Framework for inclusion of theory data in the VObs** is a major contribution to the emerging standards in this domain, and the census shows that a significant community of data centres is willing to contribute also in this new field.

With respect to interfaces with generic infrastructures, the project has also succeeded in coordinating VObs with computational grid projects. This includes the creation of the "Astronomy & Astrophysics" cluster included in EGEE-III, a close collaboration for the dissemination of knowledge about the computational Grid among the astronomy community, and successful development of pilot applications to interface the VObs tools with the Grid environment to make numerical simulations and with grid storage systems.

The project has also been present and active in OGF, and has built a good working relation with OGF-Europe.

Last, but not least, the EuroVO-DCA partner community has submitted a successful proposal to the first Infrastructure Call of the Seventh Framework programme in the *Scientific Digital Repositories* framework, which demonstrates the impact of the astronomy disciplinary model in this domain. This project, *Euro-VO Astronomical Infrastructure for Data Access* (EuroVO-AIDA), covers all aspects of Euro-VO, the Data Centre Alliance, the VO Facility Centre, and the VO Technology Centre, and combines networking, service and research activities in an *Integrated Infrastructure Initiative* (I3). EuroVO-AIDA began on 1st February 2008, for a duration of 28 months, and its work programme has been carefully articulated with those of EuroVO-DCA and of the FP6 technological Design Study VO-TECH. The I3 context gives a new dimension to EuroVO-DCA-type coordination activities, by integrating them fully with all other Euro-VO activities. The Census of Euro-VO Data Centres produced by EuroVO-DCA is a very valuable resource and will be maintained by EuroVO-AIDA.

A more global assessment of the astronomical model for the knowledge infrastructure, including the three aspects, Data Centre Alliance, Facility Centre and Technology Centre, is one mid-project deliverable of EuroVO-AIDA. It will of course take into account the evaluation made in this document for the medium term strategy of the Data Centre Alliance.

9. CONCLUSION: EUROPEAN DATA CENTRES IN THE EUROPEAN ASTRONOMICAL LANDSCAPE

9.1. The Astronet European strategic exercise for astronomy

To ensure sustainability, it is critical that the VObs is taken into account in the global strategy of the discipline. The Astronet ERA-NET has recently produced such a global strategy for Astronomy at the European level. The "Science Vision for European astronomy" document was released in September 2007. The press release mentions the VObs and data archiving. The follow-up medium term Infrastructure Roadmap has been produced in 2008 (taking into account the ESFRI Roadmap and its evolution).

Virtual Observatory was included in Panel D "Theory, computing facilities, and networks, Virtual Observatory". A very detailed questionnaire sent to the European astronomy communities shows that a majority of projects is ready to be VObs compliant. The Astronet Roadmap recommendations are strongly supportive of the VObs endeavour, and consider the provision of data archives and the VObs are important parts of the global infrastructure of astronomy. Among the recommendations:

- Provision of a public VObs compliant archive should be an integral part of the planning for any new facility. We recommend data centres provide science ready data.
- Providers of astronomical tools should make them VObs-compliant so they can easily talk to other VObs tools and can be accessed within the VObs environment.
- The development of the VObs should be coordinated with evolution of the generic e-infrastructure, and that evolution should reflect the domain-specific needs of astronomy.

It is a real success for a 'young' project such as the VObs to be recognized as an important component of the global disciplinary infrastructure. This relies on the strength of the astronomical VObs concepts, and on the VObs project accomplishments at the international, European and national levels. Next generation instruments like Gaia, ALMA, the ELT and SKA represent a massive investment of European resources, and the VObs is vital if the astronomical community as a whole is to reap the benefits.

9.2. Medium term sustainability of Euro-VO

The inclusion of Euro-VO as a genuine component of the global astronomical infrastructure in turn means that the sustainability of its constituents, and thus community support, has to be ensured on the medium term. This has different aspects :

- From the **user point of view**, the VObs must remain science driven. Data services and on-line tools are already part of astronomers everyday life, but the VObs provides in addition seamless access to on-line data and services. Among the main user requirements, is the necessity that the VObs layer be transparent for users, who do not want to know about the standards and the technological aspects. Another strong requirement on the VObs is that data must be properly described. All astronomers want good information about provenance (instrument, observation conditions, data processing, etc). Most astronomers also want science ready data, especially for the data from other sub-disciplines of which they are not specialists, but that they want to re-use – data reusability is a main aim of the VObs.

- Requirements **from the data centres** have been detailed above. The definition of the 'proper' (i.e. both usable and useful) standards is a complex task, which works when the relevant people from the community (archive providers, data centre staff) are mobilized, and when requirements and feedback from users are also carefully taken into account. This means 'enough' complexity in the standards to start with, not too much to remain usable by implementers and understandable by astronomers, with incremental addition of more complex requirements – a delicate balance, to maintain on the medium term.

The VObs projects must get support at different levels to ensure medium term sustainability:

- International cooperation and agreement on the interoperability standards is a must, so support must be given to the IVOA collaboration activities

- European coordination and cooperation has been and will remain critical: the VObs projects learn from each other and build a global strategy, and duplication of efforts is limited. This helps national activities to be better focussed and to have a much higher impact.

- R&D and technological work on the evolution of standards must continue, to take into account evolving requirements, including as explained above feedback from users and implementers, and the progressive implementation of more complexity, but also scientific and technical evolution, for instance the specific needs of the new large space and ground based projects from the Astronet and ESFRI Roadmaps.

- Support to users and data centres through the ESA/ESO Facility Centre and through the national projects.

The VObs will be fully part of the astronomical infrastructures on the medium term as required only if the following conditions are fulfilled:

- Archives from current projects are maintained on the long term, even after project completion
- New projects provide data pipelines, data archiving, data distribution, and VObs compliance – which they seem to understand and to be ready to do from the Astronet census
- Funding of data centres is provided by the relevant authorities – which can be international, as for ESA or ESO, national and/or local depending on the context, size and objective of the data centres

The VObs itself requires support at different levels for the operational phase :

- Support of national and agency (ESA, ESO) authorities to the VObs projects which will continue to maintain and evolve the VObs framework, and to support the data centres in their uptake of the VObs
- Commitment of the national and agency projects to international cooperation within IVOA (which is a best effort alliance), which require that they get funding for that purpose
- Support to European collaboration has also to be found on the medium term.

The main efforts for mid- and long-term sustainability will fall on the national funding agencies and on the intergovernmental agencies ESA and ESO. There is also a role here on the medium term for EC, in particular in support to European and international cooperation efforts. This of course raises the question of sustainability of EC support, at least at the level of Coordination Actions.

ANNEX 1 – ACRONYM LIST

ALMA	Atacama Large Millimeter/submillimeter Array
ArVO	Armenian Virtual Observatory
AstroGrid	UK Virtual Observatory project
China-VO	Chinese Virtual Observatory
CNRS	Centre National de la Recherche Scientifique
CVO	Canadian Virtual Observatory
D#	Deliverable number
DCA	Data Centre Alliance
EC	European Commission
EGEE	Enabling Grids for E-science in Europe (EC-funded project)
ELT	Extremely Large Telescope (selected in the ESFRI roadmap)
ERA-NET	European Research Area - Network
ESA	European Space Agency
ESFRI	European Strategy Forum on Research Infrastructures
ESO	European Southern Observatory
Euro-VO	European Virtual Observatory
EuroVO-AIDA	Euro-VO Astronomical Infrastructure for Data Access (EC funded, FP7 Call "Scientific Digital Repositories")
EuroVO-DCA	Euro-VO Data Centre Alliance (EC Funded, FP6 Call eInfrastructure Communication Network Development)
FP	Framework Programme
FTE	Full Time Equivalent
GAVO	German Astrophysical Virtual Observatory
HVO	Hungarian Virtual Observatory
I3	Integrated Infrastructure Initiative
ISO	Infrared Space Observatory
IVOA	International Virtual Observatory Alliance
JVO	Japanese Virtual Observatory
METAFOR	Common META data FOR Climate Modelling Digital Repositories
NASA	National Aeronautics and Space Administration
NVO	National Virtual Observatory (USA)
OGF	Open Grid Forum
OV France	Observatoire Virtuel France (French Virtual Observatory)

R&D	Research & Development
RVO	Russian Virtual Observatory
SIMBAD	Set of Identification, Measurement and Bibliography for Astronomical Data
SKA	Square Kilometre Array (selected in the ESFRI Roadmap)
SVO	Spanish Virtual Observatory
UK	United Kingdom
USA	Unites States of America
VISTA	Visible and Infra-red Survey Telescope for Astronomy
VO-I	Indian Virtual Observatory
VObs or VO	Virtual Observatory
Vobs.it	Italian Virtual Observatory
VOFC	Euro-VO Facility Centre
VOTC	Euro-VO Technology Centre
VO-TECH	The European Virtual Observatory - VO Technology Centre (EC-funded project, Infrastructure Design Study, 2005-2009)
XMM-Newton	X-ray Multi-Mirror Mission – Newton