RI031675

EuroVO-DCA

The European Virtual Observatory Data Centre Alliance

COORDINATION ACTION

RESEARCH INFRASTRUCTURE

COMMUNICATION NETWORK DEVELOPMENT

D11 – TEG REPORT: FRAMEWORK FOR THE INCLUSION OF THEORY DATA AND SERVICES IN THE VOBS

Due date of deliverable: 31/10/2008

Actual submission date: 16/12/2008

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

MPG 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[1]:

Santi Cassisi, Miguel Cerviňo, Gerard Lemson, Pedro Osuna, Joop Schaye, Nicholas Walton, Hervé Wozniak.

Editors:

Gerard Lemson, Hervé Wozniak, Jens Zuther[2]

[1] The authors are all members of the Theory Experts Group of EuroVO-DCA. They are introduced in Appendix A.
[2] EuroVO-DCA and Max-Planck-Institut für extraterrestrische Physik

Table of Contents

Figures

Acknowledgments

This whitepaper is the final deliverable of EuroVO-DCA Work Package 4. It is authored by the Theory Experts Group (TEG), but it would not have come into existence without the efforts of many others.

We give special thanks to Ugo Becciani, Laurent Bourges, Igor Chillingarian, Claudio Gheller, Franck LePetit, Patrizia Manzato, Carlos Rodrigo, Enrique Solano, and Jens Zuther for their efforts in the EuroVO-DCA activities on theory. By implementing services and through useful discussions and feedback to our ideas they have played a fundamental part in ensuring that there was actually something to write about.

We also thank the EuroVO-DCA Board and Jai Won Kim, Anita Richards, Guy Rixon and Jonathan Tedds for careful reading of the document.

For further discussions we thank Jeremy Blaizot, Francois Bonnarel, Sebastien Derriere, Norman Gray, Mireille Louys and Rick Wagner. We thank all the attendees of the Euro-VO DCA theory workshop for their contributions and discussions. More generally, this document owes a lot to the discussions held in the frame of the IVOA Theory Interest Group.

1 Executive summary

This whitepaper is the final deliverable of Euro-VO DCA work package 4 on theory in the Virtual Observatory. Here we describe our views on how "theory" can be incorporated into the framework of the Virtual Observatory (VObs, for acronyms we refer the reader to Appendix C).

We introduce the effort and provide definitions and descriptions of the main concepts used in this whitepaper: Theory, VObs and International Virtual Observatory Alliance (IVOA).

We then introduce the effort itself: why are we interested in incorporating theory into the VObs, why should theorists be interested in participating in this effort, and what are we actually thinking about as constituting "theory" in the VObs context?

We give an overview of efforts in this realm basically since the start of the VObs, but especially since the formal introduction of the theory interest group in the IVOA, Jan 2004. We cannot be comprehensive of course in describing the work that has been done, both in the IVOA and in national or even smaller scale projects. Our aim is to give illustrative examples covering the range in sizes and types of physics. In this context it is important to state that the slight bias towards "European" projects is only because those are the ones we are most familiar with. Where relevant we have done our best to include relevant projects from the rest of the world.

We conclude by describing our vision of what really constitutes the "framework of the VObs" and how theory fits into this. We define the framework as a list of requirements on the type of activities and what one may describe as boundary conditions. An example of the latter is that we believe that scientists MUST participate in the VObs, both as producers and users. An example of the required activities is that standards must be defined to enable or at least facilitate interoperability. We can summarise this section in the following two main recommendations:

- Simulation specific data access protocols such as SimDB and SimDAP should be developed as a matter of urgency

- In order to encourage take up of Theory-VO services, the Euro-VO will rapidly develop prototype Theory-VO services allowing access to scientifically useful theory simulations and models.

Finally, we conclude that it is well possible to introduce theory in this VObs framework, and the effort is on-going. But issues remain to be resolved. The architecture of the VObs needs no adjustment to accommodate theory. A resource registry for discovering interesting resources is as relevant for theoretical data products and services as it is for observational ones. Data models are required (and can be constructed) for describing theoretical resources. The approach to data access protocols based on a discovery and a retrieval phase is suitable for theory as well. The IVOA query language is relevant for filtering the sometimes very large data sets. Services can in principle be deployed, protected and chained together in workflows according to the standards of the Grid and Web services working group.

It is therefore mainly in the details of this process that work remains to be done. Particularly the standardisation for data description, discovery and data access protocols require more work above and beyond what is currently available in the IVOA. The reasons for this are simply that the existing standards so far have mainly dealt with observational data products and theory data products can be very different and very diverse. Work along these lines is progressing though. The SimDB and SimDAP standards follow the IVOA approach for what we call cosmological simulations. Efforts have started in the IVOA to investigate their applicability to other types of simulations such as isochrones, stellar evolution etc. It seems likely that these can be served with only minor changes to the models and data access service specifications, but alternative approaches are being explored as well such as a simple self-described service protocol.

The question that remains is whether developments in this area will be sufficient for the astrophysical community to start participating. This may in the end be the greatest challenge, convincing theorist to make the extra effort of conforming to the standards and publishing their results in the VObs framework. Willingness for participation expressed by the respondents to the Census of European astronomical data centres is a very good starting point. Here is important to develop successful prototypes and to assess relatively simple manners to take those extra steps. It is in this field that projects such as the Euro-VO are of great importance.

2 Introduction

How does gravitation structure the Universe on scales spanning more than 6 orders of magnitude? What are the effects of Dark Energy and Dark Matter that appear to be the dominant, but invisible components of our Universe? How do galaxies form and evolve in a hierarchical scenario? What are the histories of the production and distribution of the metals in the Universe, within and between the galaxies? What is the impact of star formation and feedback processes on the large-scale galactic dynamics? How do molecular clouds collapse and form stars? What are the effects of turbulence and magnetic fields in the mixing of chemical elements in the Interstellar Medium? How do black hole accretion, jets and outflows operate? How do dust grains grow in planetary discs? How common are exo-planets?

The approach of theoreticians attempting to shed some light on these fundamental questions is nowadays often numerical. And although analytical calculations still have an important place in theoretical astrophysics, within the context of this whitepaper the focus is on the results, "data", of such numerical computations Indeed, computational astrophysics has a long history, dating back to the computation of the first stellar structure models in the 1950s or the first evidence of chaos in gravitational systems in the 1960s. Since then astrophysics has always been a major driver of computational developments. Addressing astrophysical problems often requires the capability of the fastest supercomputers and the most recent state of the art algorithms.

The above astrophysical introductory questions – and many others not listed there – are only now coming within our ability to answer because of advances in computing and related information technology. Advanced computing has become essential to their future progress. Coupled with continuing improvements in processor speeds, converging advances in networking, software, visualization, data systems and collaboration platforms are changing the way research is accomplished.

This whitepaper discusses whether it is possible to apply the framework of the *Virtual Observatory* (VObs) to all the products made by means of theoretical models and/or by running numerical simulations, with the goal of making these accessible online in a digital format. Sometimes we will refer to this as the "theory VObs" (TVO), though it should not be concluded from this that the TVO should be in any way distinct of the main VObs.

Research in the subject of this whitepaper has been going on for a while now, and one goal is to represent the state of the TVO. It can also be seen as a follow-up to an earlier whitepaper, *Theory in the VO* ([1] , for references see Appendix D), which built on results of previous work to summarise the need for special treatment of theory in the VObs. Five years on from that whitepaper, it is a good time to make an assessment of what has happened, and draw conclusions for the future.

The layout of the whitepaper is as follows. First we define the various concepts. Most challenging is to describe the main subject of this study, *theory*. We will not attempt to define what constitutes theoretical astrophysics. But we will try to give at least a rough demarcation of *that* part of theory to which the VObs framework might be applied, and for which a special treatment might be warranted.

After that a short introduction to the VObs effort is given, again not aimed to be comprehensive, but to list those aspects we deem relevant for the further discussion. This part will emphasise the role of the International Virtual Observatory Alliance (IVOA), but will also list results of national VObs projects and the Euro-VO.

The next section gives the motivation for this whitepaper: why do we think theory and VObs should be brought together, and what are the issues that must be addressed.

The following sections describe the "state of the art" in this field. We describe standardisation efforts in the IVOA and implementations of VObs-like services by national VObs projects and other groups. We also describe results from EuroVO-DCA activities, in particular the data centre census (see section 5.3.2), which paid explicitly attention to theory activities, and the theory workshop organised by WP4. The emphasis in this section is naturally somewhat biased towards projects from the partners in EuroVO-DCA. We do expressly *not* mean to imply that theory activities are not going on elsewhere, we are simply less familiar with them.

We conclude that indeed it is possible to implement VObs-like services for theory resources and that they are useful for the scientific community at large. It is also possible to extend the approach to defining standards in the IVOA towards theory specific resources. At the same time it will be shown that in spite of the progress over the past years, work remains to be done to give theory its rightful place in the VObs effort, and to give the VObs its deserved place in the day-to-day activities of theoretical astrophysicists. We will give recommendations to the various parties involved in the process on how to proceed.

We see two main audiences for these results. The first audience consists of scientists, astronomers and astrophysicists. This includes theorists interested in publishing their data in VObs-compliant manner, and others, theorists *and* observers wishing to use these resources. The second audience consists of what we will call *VObs engineers*. This includes anyone working on the VObs framework, both those active in the IVOA, defining standards for publication and use of astrophysical data and other resources, and those implementing these standards as online services. The latter tasks are often coordinated by national VObs projects and we will have some special recommendations to those as they have the special task of being the natural intermediary between the more abstract IVOA efforts and the scientists in their "constituency".

3 Definitions

3.1 Theory

The field of theoretical astrophysics is very large and we will not even attempt to give a comprehensive definition of it. Since we are interested in the applicability of the VObs framework to this field, we will try to give some rough demarcation of what types of theoretical activities are of interest.

First of all we are mainly interested in computational astrophysics, as this is the part of theory that produces results ("data") that are open for handling in the VObs framework. Within the past two decades, this field has seen various major changes. The field of cosmological simulations for example has changed from being dominated by individual researchers writing and running their own "private" simulations on their personal workstation, to large projects that gather experts from different fields and that use the huge computational power available through parallel machines or GRID technologies. And though in the fields of stellar evolution, atmosphere models, or photo-ionization codes the collaborations are still relatively small, most of these have started sharing their results, and even codes, with the rest of the community.

Many recent advances in our understanding of the formation and evolution of the Universe have been made by projects managed by large international consortia. Many other 'extreme' computational projects are on-going at various European supercomputing centres, which attempt breakthroughs in many current topics including our understanding of the formation and death of stars and the origin of planetary systems such as our solar system. Smaller projects are still very active since they contain the seeds of future challenges. Common to all grand challenge problems is the huge dynamic range in parameter space, in time scales and spatial scales in three dimensions, and the large number of physical processes and timescales that need to be taken into account. Computational astrophysics is definitively an interdisciplinary field requesting the most sophisticated and advanced codes and resources.

Twenty years ago, important results could be obtained by an isolated theoretician with only one numerical simulation that could be published in a few graphs. Nowadays, many simulations of complex systems produce huge amounts of output that deserve to be published in *series* of papers.

Besides such 'extreme' simulations, another modern approach is to run thousand of small simulations to explore a parameter space or to fit some observational data. Whatever the technique, the outcome of a modern theoretical work is difficult to manipulate, analyse, extract and publish. Due to the regular increase in computational power that is quicker than the increase in network bandwidth it can be anticipated that the results of many future simulations might have to be stored in the Computing Centre where they were computed so that any analysis should be done remotely.

3.2 Virtual Observatory (VObs)

Virtual Research is a concept for a system to allow transparent distributed data access available worldwide. This allows scientist to discover, access, analyze, and combine data from heterogeneous data collections in a user-friendly manner. This can be applied to a variety of fields. In the context of Astronomical databases, archives and services, this has been called the Virtual Observatory (VObs).

A VObs is thus a collection of interoperating data archives and software tools which utilize the internet to form a scientific research environment in which astronomical research programs can be conducted. In much the same way as a real observatory consists of telescopes, each with a collection of unique astronomical instruments, the VObs consists of a collection of data centres each with unique collections of astronomical data, software systems and processing.

3.3 National Virtual Observatories

Sixteen VObs projects are now funded through national and international programs (cf. next section), and all projects work together under the IVOA (see next section) to share expertise and best practices and develop common standards and infrastructures for the VO. Several additional countries, some with emerging astronomical communities, have already claimed their interest in the VO and could join soon the IVOA.

There is no typical organisation for national VO projects. Some of them are mainly light organisation aiming to coordinate their national community or contact points for the IVOA whereas others could act as funding agencies, hire their own manpower and perform VO–oriented software development and distribute. Data centres are active participants in national VO projects.

Two European agencies, the European Space Agency (ESA) and the European Southern Observatory (ESO), are also very active in the VObs, in a way similar to National organizations.

ESA-VO (http://esavo.esa.int/) main goal is to have the ESA Space Astronomy Centre (ESAC) as the VO node for all European space based astronomy by mainly ensuring all ESAC astronomy archives are accessible through the VO framework. ESA develops VO science applications and VO tools allowing ESA and other data providers to publish their scientific data holdings through the VO.

ESO has been a key player in the VO arena from the beginning. Its VO activities have been managed until mid-2008 by the Virtual Observatory Systems (VOS) Department of the Data Management and Operations Division (http://www.eso.org/org/dmd/vos/), whose mission was also to make the ESO Archive into a powerful scientific resource for the community. VOS has also been working towards making all ESO data VO-compliant and creating and ingesting ESO science-ready data products. As of June 1, 2008, VO activities at ESO are managed by the Virtual Observatory Project Office, which provides VO input to the whole of ESO in general and to the newly established Archive and Data Products Departments in particular.

3.4 International Virtual Observatory Alliance (IVOA)

The International Virtual Observatory Alliance (IVOA[3]) was formed in June 2002 with a mission to "facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory." The IVOA now comprises 16 VO projects from Armenia, Australia, Canada, China, Europe, France, Germany, Hungary, India, Italy, Japan, Korea, Russia, Spain, the United Kingdom, and the United States. Membership is open to other national and international projects according to the IVOA Guidelines for Participation.

[3] http://www.ivoa.net

Figure 1 The national VO-Projects participating in the IVOA

The work of the IVOA focuses on the development of standards. Working Groups are constituted with cross-project membership in those areas where key interoperability standards and technologies have to be defined and agreed upon. The Working Groups (WGs) develop standards using a process modelled after the World Wide Web Consortium, in which Working Drafts progress to Proposed Recommendations and finally to Recommendations. Recommendations will ultimately be endorsed by the Virtual Observatory Working Group of Commission 5 (Astronomical Data) of the International Astronomical Union. The IVOA also has Interest Groups (as for theory) that discuss experiences using VO technologies and provide feedback to the Working Groups.

Senior representatives from each national VObs project form the IVOA Executive Committee. A chair is chosen from among the representatives and serves a 1.5 year term, preceded by a 1.5 year term as deputy chair. The Executive Committee meets 3-4 times a year to discuss goals, priorities, and strategies.

The IVOA holds two Interoperability Workshops each year: a week-long meeting in the spring, typically May, and another meeting in the fall that is can be coordinated either with the annual ADASS conference or with a regional VObs project meeting. These meetings are opportunities for the Working Groups and Interest Groups to have face-to-face discussions and for the more difficult technical questions to be resolved.

The tasks of the IVOA are distributed over different working groups (WGs). Those relevant for this whitepaper are introduced here with the text of their charter. Later in this whitepaper we describe some of the standards defined by these WGs and their relevance for theory.

3.4.1 Applications[4]

The IVOA Applications Working Group is concerned primarily with the software tools that Astronomers use to access VO data and services for doing Astronomy.

The VO is enabling new ways of doing Astronomy. Interoperability between Astronomy data, services and software empowers astronomers to combine these for scientific discovery and analysis. VO Applications may take many forms such as GUI desktop applications, software libraries, web interfaces to services, or other innovative 'portals' to the VO. Legacy Astronomy software is becoming VO enabled and new and novel VO applications are being created. The IVOA Applications Working Group provides a means for VO Applications development and implementation to be closely linked to the standards development in the IVOA, and where necessary to propose and develop standards for VO Applications to interoperate.

The role of the Applications Working Group is to:

- *Provide a forum for announcement and discussion of VO Applications*
- *Provide feedback to IVOA on the implementation of interoperability standards in VO applications*
- *Identify missing or desirable technical capabilities for VO applications*
- *Identify missing or desirable components in terms of scientific usability*
- *Propose and develop standards specific to VO Astronomy-user-Applications*

3.4.2 Data Access Layer (DAL)[5]

The task of the DAL working group is to define and formulate VO standards for remote data access. Client data analysis software will use these services to access data via the VO framework; data providers will implement these services to publish data to the VO. The DAL working group will define the scope of the DAL standards, outline a process by which DAL standards are defined, and generate the initial version 1.0 of the DAL standard. This standard will provide guidance to data centres and survey projects when designing VO compliant interfaces. It will allow them to justify the allocation of resources for its implementation and maintenance. Once the work on Version 1.0 is accomplished the working group will coordinate future development of the standard.

The DAL working groups has defined various standards for accessing data sets, in particular images (Simple Image Access Protocol, SIAP[6]), spectra (Simple Spectra Access Protocol, SSAP[7]) and source catalogues (Simple Cone Search, SCS[8]).

---

[4] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaApplications
[5] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaDAL
[6] http://www.ivoa.net/Documents/latest/SIA.html
[7] http://www.ivoa.net/Documents/latest/SSA.html
[8] http://www.ivoa.net/Documents/latest/ConeSearch.html

3.4.3 Data Modelling (DM)[9]

The role of the Data Modelling group is to provide a framework for the description of metadata attached to observed or simulated data. The activity of the Data Model WG activity focuses on logical relationships between these metadata, examines how an astronomer wants to retrieve, process and interpret astronomical data, and provides an architecture to handle them. What is defined in this WG can then be re-used in the protocols defined by the DAL WG or in VO aware applications.

3.4.4 Grid and Web Services (GWS)[10]

The aim of the GWS WG is to define the use of Grid technologies and web services within the VO context and to investigate, specify, and implement required standards in this area. This group was formed from a merger of the Web Services group and the Grid group, ordered at the IVOA Executive meeting held during the IAU General Assembly in 2003.

3.4.5 Resource Registry[11]

This working group defines the structure and interface to an *IVOA Registry*. Such a registry " *… will allow an astronomer to be able to locate, get details of, and make use of, any resource located anywhere in the IVO space, i.e. in any Virtual Observatory. The IVOA will define the protocols and standards whereby different registry services are able to interoperate and thereby realise this goal.*"

3.4.6 Semantics[12]

The IVOA Semantics Working Group will explore technology in the area of semantics with the aim of producing new standards that aid the interoperability of VO systems. The Semantics Working Group is concerned with the meaning or the interpretation of words, sentences, or other language forms in the context of astronomy. This includes standard descriptions of astrophysical objects, data types, concepts, events, or of any other phenomena in astronomy. The WG covers the study of relationships between words, symbols and concepts, as well as the meaning of such representations (ontology). The WG covers use of natural language in astronomy, including queries, translations, and internationalization of interfaces.

3.4.7 VO Query Language (VOQL)[13]

The IVOA Query Language group will be in charge of defining a universal Query Language to be used by applications accessing distributed data within the Virtual Observatory framework.

3.4.8 VOTable[14]

The VOTable format is an XML standard for the interchange of data represented as a set of tables. In this context, a table is an unordered set of rows, each of a

---

[9] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaDataModel
[10] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaGridAndWebServices
[11] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaResReg
[12] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaSemantics
[13] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaVOQL
[14] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaVOTable

uniform format, as specified in the table metadata. Each row in a table is a sequence of table cells, and each of these contains either a primitive data type, or an array of such primitives. VOTable [2] is derived from the Astrores format [3] , itself modelled on the FITS Table format [4] ; VOTable was designed to be closer to the FITS Binary Table format.[15]

3.4.9 Theory Interest Group (TIG)[16]

During the IVOA executive meeting of January 2004 in Garching, Germany, the IVOA Theory Interest Group was formed with the goal of ensuring that theoretical data and services are taken into account in the IVOA standards process.

By its charter, the IVOA Theory Interest Group intends to:

- *Provide a forum for discussing theory specific issues in a VO context.*
- *Contribute to other IVOA working groups to ensure that theory specific requirements are included.*
- *Incorporate standard approaches defined in these groups when designing and implementing services on theoretical archives.*
- *Define standard services relevant for theoretical archives.*
- *Promote development of services for comparing theoretical results to observations and vice versa.*
- *Define relevant milestones and assign specific tasks to interested parties.*

3.5 Euro-VO DCA

The Euro-VO Data Centre Alliance project (EuroVO-DCA) is a Coordination Action funded in the framework of the Sixth Framework Programme e-Infrastructure *Communication Network Development* initiative. The top level objective of the Euro-VO Data Centre Alliance (DCA) is to coordinate European Data Centres in forming a co-operating community enhancing the European astronomical e-Infrastructure and, thereby, maximising the scientific utilisation of the rich astronomical on-line resources distributed all over Europe. The project will enable the identification and promotion of requirements from programs of strategic European and national interest that require VObs technologies and services. The implementation and sharing of standards and methods will make it possible to achieve a production-level European-wide e-Infrastructure.

The objectives of EuroVO-DCA can then be summarized in 6 key points, corresponding to the project work packages[17]:

- Co-ordinate the national and European Agencies VO initiatives, to implement networking of European data centres (WP2),

[15] http://www.ivoa.net/Documents/REC/VOTable/VOTable-20040811.html
[16] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaTheory
[17] http://www.euro-vo.org/pub/dca/workpackages.html

- disseminate knowledge and good practice about the VO technical framework (WP3).
- organise feedback from implementation of interoperability standards (WP3-2),
- prepare the inclusion of theoretical astronomy in the VObs framework (WP4),
- seek coordination with national and international projects for computational Grids (WP5),
- help data centres from beyond the partners' countries to participate in the VO endeavour (WP6).

4 Theory and the VObs

Why do we think it is of interest to introduce theory into the VObs? What are the issues involved? How will we set about addressing those? Here we try to motivate this whitepaper, point out the differences between theory and observational astronomy that make this effort non-trivial, and categorise the different types of efforts involved.

4.1 Why publish theory in the VObs?

Many of the motivations for publishing theory in the VObs are the same as for observations. The VObs is supposed to facilitate the access to data and services by providing standardised interfaces, so the question as to why publish theory will be partly answered by the general question why one would want to publish ones scientific results in the first place.

Publishing your results online...

- gives your colleagues access to the *data* described in your publication, thus providing an extra motivation to extend your work, which leads to new science and increases the publication's impact;
- enables scientists not directly involved in your project to verify your results, which in general helps to find mistakes, thus improving your results;
- provides others with a benchmark to which to compare their own results, both for reasons of reproducibility and for checking their research;

- is increasingly mandated by funding agencies[18];
- can serve as showcase for future proposals;
- may facilitate the refereeing process, by giving referees access to the data (possibly not yet public);
- … .

Publishing your results in the VObs…

- makes your results available in a standardised manner, facilitating their discovery and greatly increasing their reusability;
- enforces a good practice to follow VObs-like standardisation: it forces you to think carefully about your own results, which improves reusability even for you;
- is the proper thing to do if you also want to use the VObs for your own research, "what goes around, comes around";
- may not give obvious benefits to you, but you may agree that it is good for people to think about how others should publish their scientific results in a homogeneous manner, so that you and others have an easier job interpreting and using these.
- …

In any case, having your results seen and reused by others will increase the *impact* of your research.

On top of all of this, publishing the results of your theoretical research to the VObs…

- allows others to compare observations to models, facilitating their interpretation, or enabling more sophisticated predictions (survey planning, exposure time calculator++, ...);
- allows one to see physical processes in action;
- allows fellow theorists to compare and test their analytical models on state of the art simulations;
- …

Even if these lists may not convince all astrophysicists to participate in the VObs framework, plenty of them do or are interested in doing so, ensuring us an audience and a core of potential data and service providers.

[18] For example the Max-Planck Society in Germany has mandated that all data products produced by its scientists should be made available online for at least 10 years after the publication date.

4.2 Challenges for publishing theory

The main challenge for the VObs to be able to support theoretical data is that in general they are very different from the observational data products that have so far been mainly handled. Though it is often possible to create synthetic images, spectra and source catalogues, these are generally the results of *post-processing* of more fundamental simulation results which should also be handled. And there is one big difference between publishing these *virtual observations* and the real ones: they do not observe the same universe. Hence there are no common coordinate systems to use in protocols to find familiar objects; there are in fact no common objects to be identified. So whereas *in principle* publishing mock observations should be as straightforward as making real observational data VObs compliant, *in practice* specific standards must be developed, or existing standards must be improved to ensure a proper handling by the VObs.

In any case, reducing the contribution of theory to the production of mock images or spectra would limit the possible impact of publishing theory in the VObs. Indeed, the virtue of the theoretical results is to provide an in-depth understanding of the physical and chemical processes at play in Nature, most of the relationships between the various components of a simulation, and undoubtedly the most interesting ones, being unobservable. Good examples are the structure of stellar interiors, the dark matter and dark energy component included in most N-body simulations, or the ionization structure in photoionised regions.

Hence we believe that it is still true what was claimed already in [1] , namely that theory requires special attention within the VObs framework. The focus on observational astronomical datasets and services of the VObs in general and the IVOA in particular has had the consequence that many of the current IVOA data models and data access standards are irrelevant for theoretical data products.

The most important parameters for the query protocols for discovering interesting datasets contain positions on the sky, the IVOA query language contains definitions for regions on the sky, data models characterise the spatial, temporal and/or wavelength extent of observational datasets, registry resources can indicate their footprint on the sky. In the few cases where theory resources been taken into account in the construction of standards, in general this has not happened in great detail yet (see section 5.1.5).

It is good to investigate this issue in some more detail, as it will indicate also the problems that must be solved when supporting theory with standardisation. One of its causes is that the ideas for the VObs originated in the observational community, and people from that community were the first developers and obviously their requirements lead to the first standards. But one can argue that the observational community is *a priori* more suitable for the VObs framework anyway:

1. *Simple observables*. Observations described using small number of parameters (space/time/wavelength/flux/polarisation). Hence standardisation is relatively simple, at least wrt. the data part. It is in the definition of appropriate metadata descriptions that most effort is spent.

2. *Common sky*. When looking at some part of the sky, one can expect to see the same objects as other observations of that same part of the sky.

3. *Small set of observatories* (of the order of a few hundreds). These are reused by many scientists; so many different scientists have very similar data products to begin with. Instead of re-observing, might first look at archive. Not much work for these to agree on common standard, since FITS has been a common format for astronomical data for many years.

4. *Archiving obviously useful*. Even 100 years old observations may be of use to present day astronomers. Consequently data centres existed before the VObs concept was explicitly announced, and they contained expertise of use to VObs development.

5. *Large, relatively homogeneous community*. Most astronomers are observers, all of whom are interested in the resources available. Moreover, though clearly there are large differences of specialisation in different wavelength regimes, to first order science-ready observations can be compared to each other by anyone, if only by overlaying one image on top of another. Consequently use cases abound for interoperability.

Though one may argue about the detailed validity of these points, we believe they are definitely true when compared to the analogous situation in theory:

1. *Complex "observables"*. Anything that can be imagined can be modelled and simulated. Consequently much more complex to standardise.

2. *No common worlds*. In general, computer simulations start from random initial conditions and in the general case it is impossible to identify common objects in different simulations. Hence any possible interoperability between different services must necessarily follow a different pattern.

3. *Every computer is a laboratory*. Every computer can be used to produce data from some simulation run on it. And everyone who knows how to program computers can create a simulation package for relatively little costs. Hence the potential for heterogeneity are very much larger than in the observational world. In some cases *ad hoc* standards are developing, but this is mainly true for the more complex types of simulations. For example cosmological simulations, of which there are only a few tens software packages that are freely available and reused.

4. *Moore's law*. Old simulations can be redone after a few years using cheaper, faster resources. So special reasons must be found why one would spend a lot of attention (and resources) to archiving one's results.

5. *Very diverse, self-reliant community*. The freedom one has in modelling naturally leads to great diversity in models. Furthermore many theorists/simulators will rather write their own codes than use someone else's results. Alternative use cases have to be found that are of sufficiently general appeal to warrant creating standards for.

In spite of these complexities, it is still felt worthwhile to spend efforts on theory. In the next section we will discuss what type of theory is of interest

4.3 What should be published and how?

We broadly identify three aspects to the publication of theory resources to the VObs. First is whether the subject, *what* is being published, makes any specific requirements on the publication process. The following is what specific interoperability use cases are supported, and finally what requirements the implementation of the standards has to fulfil. Of the first of these, we identified two different types of simulations, those in space-time coordinates and modelling and simulations of physical/chemical processes, which likely will require different approaches to their publication. We identify two types of interoperability, namely between different theory resources, and the other where observational and theory resources are combined. These are discussed in the next four subsections.

4.3.1 Simulations in space-time coordinates

The main characterising feature of this type of simulations is that a volume of 3D space is explicitly modelled and evolved forward in time. The numerical codes widely used are called N-body, Adaptive Mesh Refinement (AMR), Smooth Particle Hydrodynamics (SPH), and so on. The subject of these simulations can range from cosmology and large scale structure of the universe down to planet formation. In the past we have referred to this type of simulations as "cosmological", meaning, simulating "part of the cosmos" (on whatever scale!). Sometimes we use this phrase in this whitepaper as well. The data products resulting from these simulations can be very large, easily exceeding the largest observational catalogues. Examples of the types of data products are shown in Figure 2.

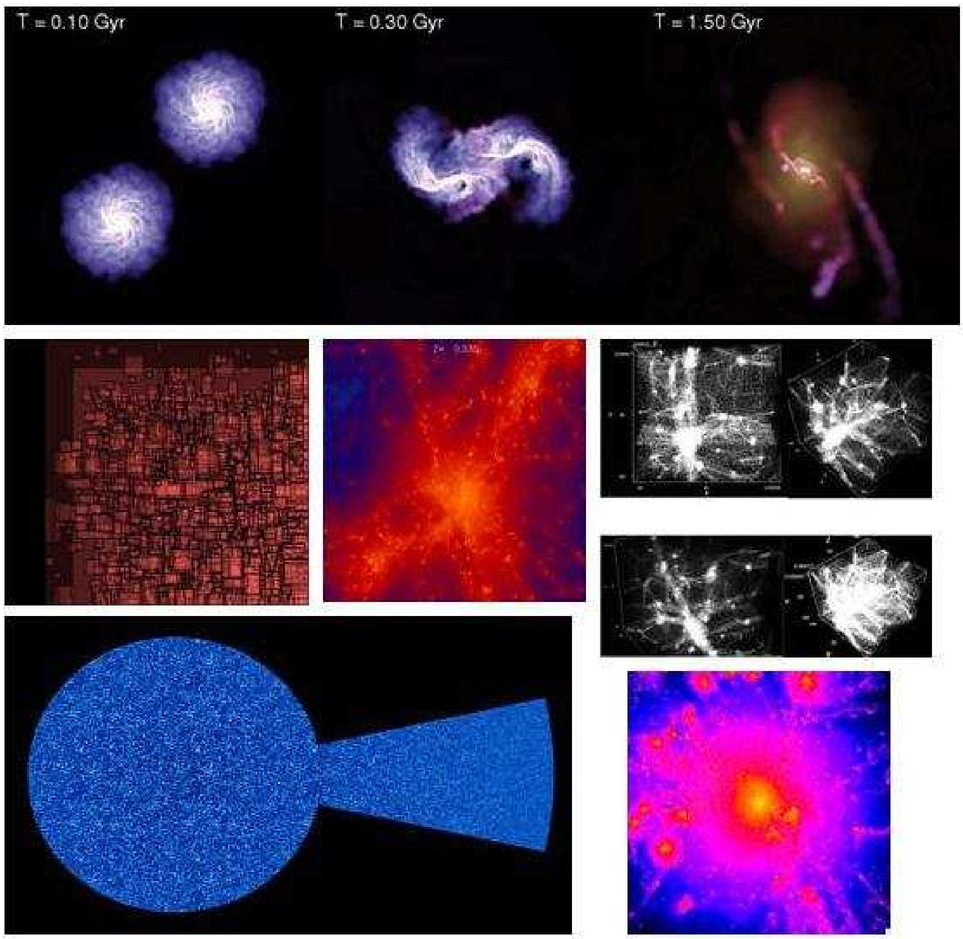

Figure 2 Examples of data products of cosmological simulations, obtained from presentations of the DCA WP4 workshop by Stefano Borgani, Patrizia Manzato, Joop Schaye, Volker Springel, Rick Wagner.

4.3.2 Modelling and simulations of physical/chemical processes

This type of simulations concerns models of astrophysical processes that do not necessary involve space coordinates. They are often nicknamed "micro-physics simulations" in the following. Examples are stellar atmosphere models (both in terms of structure of the atmosphere and emitted fluxes), stellar evolutionary tracks, stellar structures, luminosity functions and isochrones, synthesis models of stellar populations, photo-ionization models, chemical processes in interstellar clouds like cloud cooling and fragmentation, interstellar medium, global galaxy models and chemical evolution models.

Important features distinguishing this class of simulations from cosmological simulations are the size and type of products. In general the results of these simulations are much smaller and more easily handled, and because of this in general come in large grids, resulting from the study of an often high-dimensional parameter space. Illustrations of some typical data products of these types of simulations are shown in Figure 3.

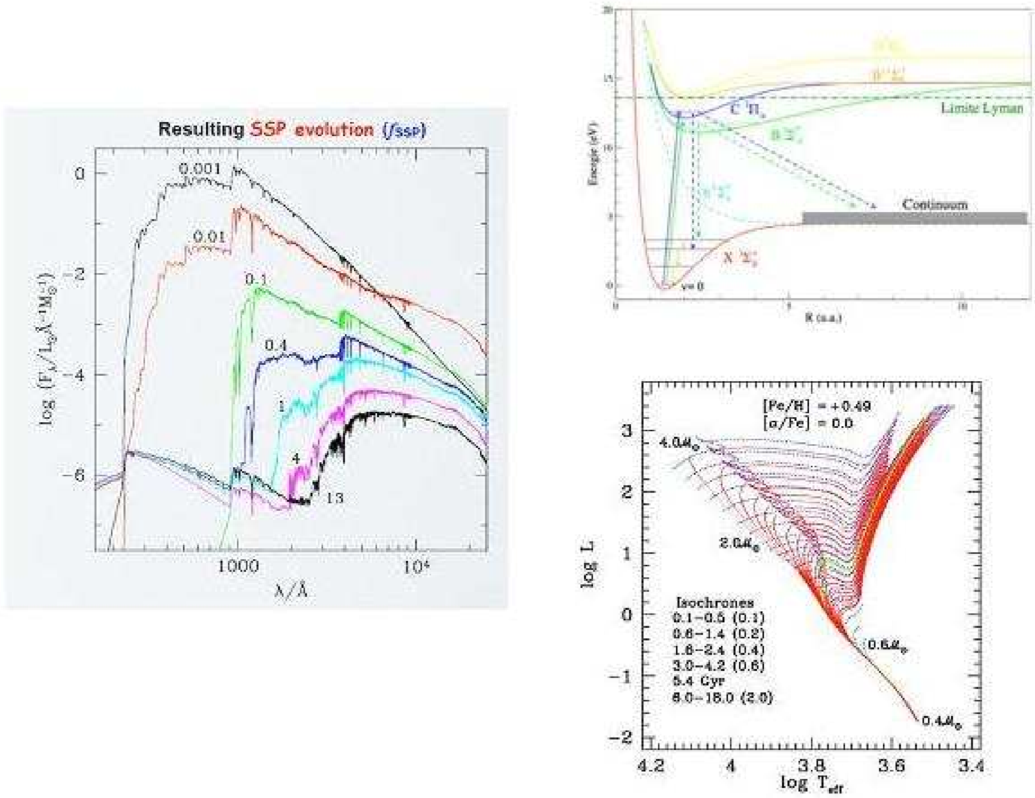

Figure 3 Examples of results of micro-physics simulations, obtained from presentations during the DCA WP4 workshop by Santi Cassisi, Stephane Charlot, Franck LePetit

4.3.3 Theory- theory interoperability

An important promise of the VObs is that it facilitates, or even enables interoperability between distributed resources. But what does interoperability mean when theory resources are involved? As we saw above, efforts at achieving interoperability in the VObs had been restricted mainly to observational resources. An important reason for this is that scientific use cases for such interoperability are easy to imagine: when looking at the same part of the sky with different telescopes and in different wavelengths one may expect to obtain information about the same objects. If the corresponding data products are published online in a standardised fashion, it will be straightforward to write applications that can provide a multi-wavelength view of these sources.

As mentioned in the section 4.2, there are some key points which facilitate interoperability in the observational context. First is the assumption (fact) that we all see the same sky, with the same objects in it no matter which instrument we point at it. Second, the resulting data products are very similar, because the relevant observables are very limited and always the same: space, time, wavelength and luminous intensity. Third is that we can build on existing standards for representing these data products, for example FITS for images. Important here is that FITS is not only a data format standard, but especially that it contains standards for describing the meta-data.

It is much less straightforward to imagine use cases for interoperability in a purely theoretical setting. Consequently it is more difficult to define appropriate standards to support interoperability of theoretical resources. The reasons can be seen from noticing that the points which facilitate this effort in the observational context do in general not hold for simulations: there is no common world to observe (simulate), the objects under investigation come with a number of observables only limited by the simulator's imagination and capabilities of implementation, and standardisation of data products is consequently less well evolved, if existent at all.

4.3.4 Theory-observational interoperability

One of the potentially very promising results of theory VObs projects is that they might narrow, if not bridge, the gap between observers and theorists. How to do so, where each discipline has its own and generally very different technical requirements, is still an open question, though some proposals have been made and efforts are under way to implement them. One example is the concept of the "virtual telescope"; an example of this are the FITS files produced by the X-MAS code stored inside the ITVO database as a start implementation. Theorists might attempt to provide services that allow users to observe their simulations and produce synthetic images or spectra that can be directly compared to their observational counterparts. Other examples are services that use theoretical models to interpret observations, such as tools to "invert" galaxy spectra to predict their stellar populations[19]. Currently, some analysis tools and services making use of theoretical models are present in VObs applications, VOSpec for example, but the theoretical models that are used are hard coded in the applications, and not obtained form other VObs services.

However, in many cases numerical simulations can be performed only by using various theoretical "ingredients" provided by research groups working in different scientific fields. Just as an example, it is worth mentioning that stellar model computations can be performed only when accurate theoretical predictions about the equation of state, radiative opacity, boundary conditions, etc. are available. The same evidence is true for population synthesis tools, as well. Some illustrations relevant to interoperability, both theory-theory and theory-observational are shown in Figure 4.

[19] http://cds.u-strasbg.fr/twikiDCA/pub/EuroVODCA/WCAWP4Workshop/Matching_Data_and_Models_in_the_VO.pdf

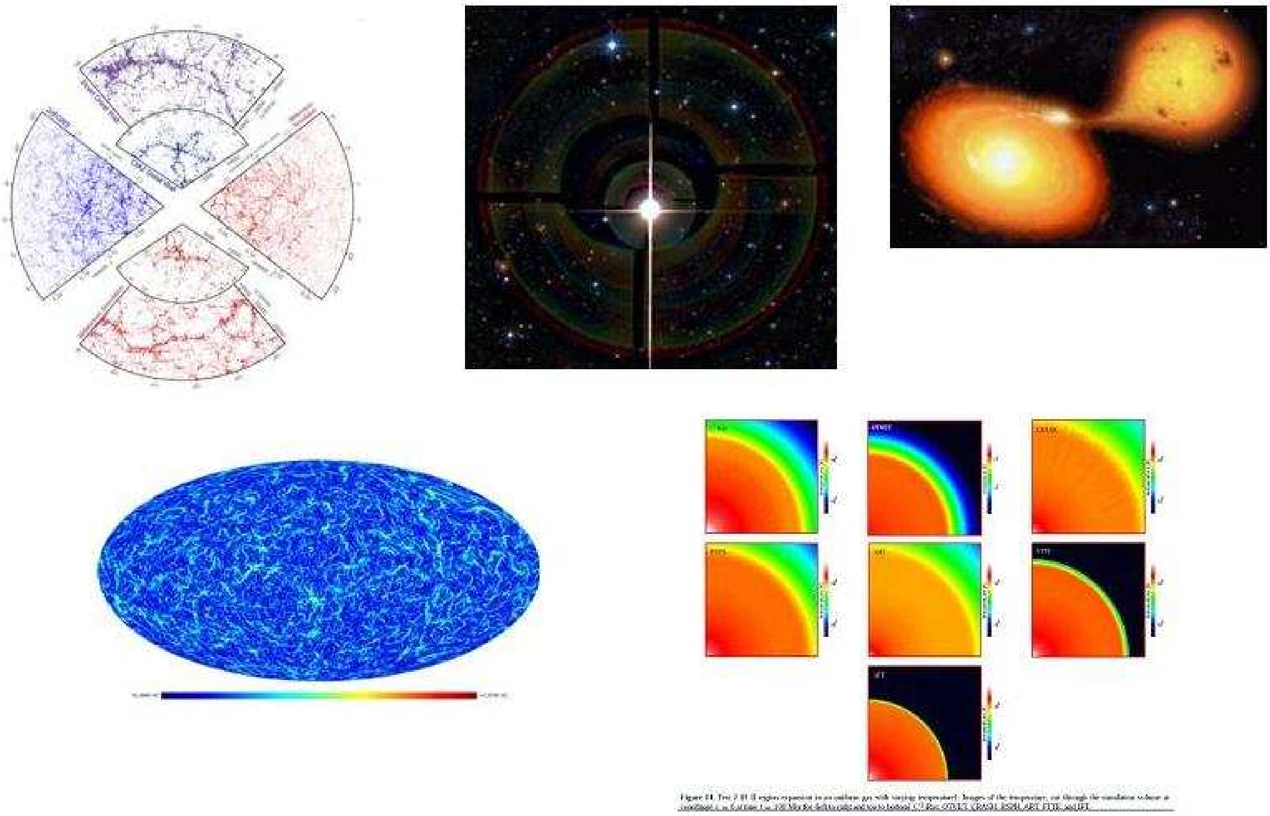

Figure 4 Illustrating theory-based interoperability obtained from the presentations during the DCA WP4 workshop by Emanuel Bertin, Ilian Iliev, Volker Springel, Peter Teuben

5 The current state of affairs

Here we give an overview of the state of the "Theory VObs" at the time of writing of this whitepaper, i.e. at the end of the Euro-VO DCA project, December 2008. This includes activities by scientists, by national VObs's, the Euro-VO and the IVOA.

5.1 IVOA

The IVOA is the standard body for the VObs. Its activities are therefore the most relevant when describing a *framework* for the inclusion of theory into the VObs. The Euro-VO DCA on its own can not define this. For this reason most EuroVO-DCA activities on theory have been performed within the context of the IVOA. This includes participation in IVOA interoperability meetings, definition of standards, joint meetings. Here we describe the current status of theory in the IVOA, with a short history of its development.

5.1.1 Whitepaper *Theory in the Virtual Observatory*

This paper [1] was the first attempt to discuss the introduction of theoretical results and data into ongoing VObs and in particular IVOA activities. The whitepaper had two main target audiences. The first was the community of theorists who either wish to take the initial steps necessary to publish their results online, or who want to make their existing online presence "VO compatible", in both cases ensuring consistency with well defined standards. For this audience, the authors attempted to describe existing VO efforts, with particular emphasis on the standardization efforts embodied by the International Virtual Observatory Alliance (IVOA). The second audience comprises developers working in the various VOs, especially those involved in the IVOA. For this group, the authors described how theoretical archives and related services imply interesting new requirements on these efforts.

This whitepaper was the direct motivation for the creation of a specialized "theory interest group" dealing with theory data in the IVOA. The document has formed the basis of a charter for this working group described hereafter.

5.1.2 Theory interest group (TIG)

Organisation

Since its formation the TIG has been the most active interest group in the IVOA with many well motivated participants. Since a workshop in Cambridge, UK Feb 2006 it has been actively working on two specifications, SimDB and SimDAP, and in the IVOA interoperability meeting in Baltimore, October 2008 a proposal for a third, S3 was made. These are discussed below in some more detail. A lot of discussions have recently been devoted on how an interest group might go about defining standards. According to the IVOA's rules [5] only working groups can do so. The current solution is that a single working group must ultimately be responsible for the promotion of a standard, but that the interest group can organise the effort. In the mentioned efforts SimDB in particular will require input from a number of working groups: DM, DAL, Registry, Semantics and it is not clear at the current time which working group will finally propose the standard. SimDAP will be moved to the DAL group.

Use cases

During the interoperability meeting in Kyoto, 2005, it was decided that science use cases will be defined by various members of the theory group. They have been used as a starting point to extract requirements for the various IVOA working groups that are of a theoretical nature. We give hereafter a short abstract of each use case to derive the main drivers for future VO development for Theory. Details can be found at: http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaTheoryUseCases

Most of these use cases deal with the comparison of some observational datasets with one or several theoretical models pre-computed or to be computed on the fly. The observation–theory cross–comparison appears to be a natural use case which must be handled by the VO.

Structure of X-Ray Clusters: A user has obtained and analysed a detailed X-Ray image of a galaxy cluster and derived temperature and pressure maps from them. Using locally developed statistical techniques small scale sub-structure is investigated and an interpretation in terms of a turbulent intra-cluster medium is proposed. To investigate whether this interpretation is realistic a similar analysis is desired on simulated images. This requires availability of hydro simulations of galaxy clusters with the appropriate physics included and services that can create mock temperature and pressure images from this simulation.

Virtual Telescope Configuration: This use case describes a distributed workflow where an astronomer aims to reproduce an observation by combining appropriate simulations, visualisation services and virtual telescopes altogether mimicking the real observational configuration. This real observational configuration is assumed to consist of a source, emitting photons, possible foregrounds affecting the photon properties/distribution and a telescope.

Synthesis Models in VO: A user wants to compare the results given by different grids of evolutionary synthesis models (produced by different developers/groups) with observational data.

Multiwavelength Analysis of Interstellar Clouds: An observer gets reduced spectra at different wavelengths towards a molecular cloud and tries to match these spectra by a physics-chemistry model and a radiative transfer model to get physical conditions in the cloud. He/she searches and finds relevant codes (a PhotoDissociation Region – PDR – code and a radiative transfer code). He/she sends parameters to the PDR code which computes temperature and abundances at each point of the cloud. These results are sent to the radiative transfer code which produces the synthetic spectrum to be compared to the observations.

Determination of Physical Conditions in Interstellar Clouds: An astronomer studies a particular interstellar cloud for which he gathered column densities of some observed atoms and molecules. He/she searches and finds a grid of pre-calculated results. Registries point towards: a grid of pre-calculated results from a PDR code, the software aimed to search information in the grid and the PDR code used to build the grid.

Theory TAP service: Astrophysicists doing simulations wish to publish their results in a VObs compatible manner; in this case they wish to make their database of simulation catalogues available for querying via the VObs query language, publishing it as a TAP service. What is required is a specification that includes such datasets and a query language that allows the specifics of theoretical datasets. The datasets will include mock catalogues of observational objects such as galaxies, stars, but also of "unobservable" objects such as dark-matter halo merger trees extracted from N-body simulations.

Tools for cosmological simulations (Simulated S-Z maps): This use-case describes one implementation of a set of modular tools for analysing and observing cosmological N-body simulations. These 'modules' can be combined sequentially into an AstroGrid style workflow to generate simulated Sunyaev-Zel'dovich maps from an N-body simulation, the input for the following module being the output of the preceding, or implemented individually returning the results in a standardised format. The final an output can be directly compared to 'observational' data. The key underlying requirements are the development of a standard format for simulated data, associated metadata to fully describe the contents of simulated archives and support for theory specific queries in VOQL.

Intermediate scale (Nbody/Stellar Evolution in Globular Clusters / MHD Simulations of Astrophysical Jets): The first problem addresses the evolution of globular clusters, where the dynamical and stellar evolution is followed for every star in the cluster. The second problem describes the evolution of extragalactic jets, both energetic FRII MHD jets and less energetic FRI MHD jets in clusters of galaxies that form buoyant bubbles in the intracluster medium.

Galactic Stellar Content Simulator: offer a service which supplies a realistic distribution of the Galactic stellar content in a given sky field. Could be used in Virtual Telescopes[20]; could be combined with other simulators (e.g. galaxy cluster simulators, Kuiper belt simulators, etc.). Once the service is located in a registry, and the description fits the user requirements, the user must be able to supply input parameters (limiting magnitudes, field positions, ...) and then get back observational parameters of a Galactic stellar distribution which can be combined with the output of other simulator(s) and input in e.g. an instrument simulator.

Compare HR diagram for observational and theoretical cluster data: A user wants to compare observational and theoretical HR diagrams for given cluster parameters (ex 'cluster age' or named cluster like "47 Tuc").

5.1.3 SimDB and SimDAP

The IVOA TIG is actively working on two related standards, the Simulation Database (SimDB) and the Simulation Data Access Protocol (SimDAP). These two specifications were derived form an original effort for a Simple Numerical Access Protocol (SNAP). SNAP found its origin in a meeting of theorists and VO engineers in Cambridge, UK[21] (Feb 2004). The goal of SimDB is to define an interface to a database storing metadata about cosmological simulations (in the sense of section 4.3.1). Users should be able to register simulations there and query for interesting simulations. The model for the metadata includes pointers to web

[20] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/VirtualTelescopeConfiguration
[21] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/CambridgeTheoryWorkshopFeb06

services that give access to the actual simulations. The protocol for these services is defined by SimDAP.

This whitepaper is not the place to discuss these standards in more detail Developments on them can be followed on the IVOA wiki pages on theory, but in particular also in the SVN repository[22] of the Volute project on Google Code.

5.1.4 TSAP and S3

At the most recent interoperability meeting in Baltimore USA (Oct 2008) a proposal was made for a simple protocol for accessing micro-physics simulations[23], the "Simple Self-described Service" protocol (S3). This proposal is under discussion in the TIG and can be seen as an extension to an older proposal for a *Theoretical Spectral Access Protocol* (TSAP).

TSAP was the first standardisation proposal to handle theoretical data in the IVOA, in particular accessing theoretical spectra. This approach is now part of the Simple Spectra Access Protocol (SSAP) as a use case for theoretical spectra. Some services offer theoretical spectra using the SSAP standard and applications (VOSpec, VOSA...) give access to them. Figure 5 provides a snapshot of theoretical SSAP services from VOSpec. Unfortunately, the registration of purely theoretical SSAP services (and theoretical services in general) is not yet well established.

S3 is based in the ability of a theoretical data server to describe itself in a simple standardized way. This approach follows the one present in SSAP for theoretical spectra. S3 follows the same philosophy as TSAP and generalizes it to include other kind of theoretical data. At present there exist many pilot implementations: 7 made by the SVO, 2 by the ITVO and 3 by Virtual Observatory groups at Mexico and Brazil. These implementations provide a huge range of theoretical products, like evolutionary tracks, isochrones, photoionization model results, synthesis models results (other than spectra), and star formation histories of galaxies. Also it has been used to produce scientific refereed papers (e.g. [6]). The proposal is under discussion in the TIG.

5.1.5 Other IVOA standards and theory

Various standards developed by the IVOA working groups are either neutral to whether the resources are observed or simulated, or have some explicit support for theory, though generally in a limited sense. Here we give a short overview of the most relevant ones.

Registry

The registry allows publishers to describe their resources and users to discover these. There is a registry resource data model [7] that contains some elements that allow publishers to indicate that a resource is the result of simulations rather than observations. There is as of yet no support for describing simulations in any more detail. The development of such *registry extensions* is one of the tasks that

[22] http://code.google.com/p/volute/source/browse/#svn/trunk/projects/theory
[23] http://www.ivoa.net/Documents/latest/S3TheoreticalData.html

should be included in theory standards such as SimDB, SimDAP and the S3 proposal.

Data Access Layer

The DAL working group has developed a standard for accessing spectral data sets. In this effort explicit attention has been given to spectra derived form simulations. See again Figure 5 which shows that quite a few purely theory spectra archives have been implemented and registered.

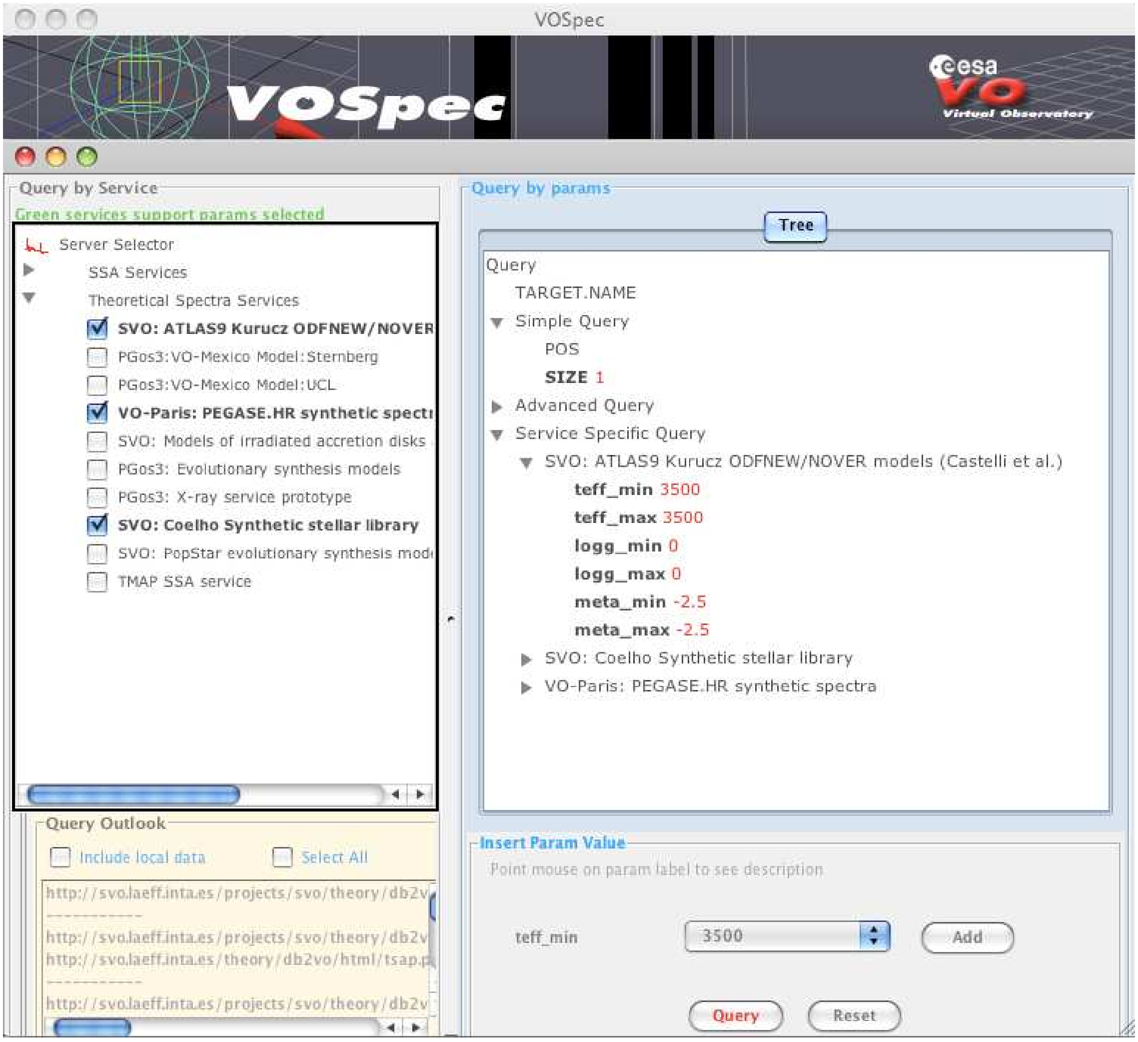

Figure 5 Selection of theory SSA services in VOSpec.

Other, older DAL standards are not yet explicitly tailored to support simulation data, but efforts are underway to develop new versions of these where simulated data sets will get a similar place.

The main issue with these services is that the query protocols are still very much tailored to observational data sets. In fact the mandate is for theory services to return *no* results when any non-relevant observational parameters (position, time) are specified in queries. Querying is possible using custom parameters, but the support for retrieving these is as of yet very indirect.

The Table Access Protocol (TAP) standard is under development and will prescribe the publication of tabular datasets, most likely in relational databases. It is clearly possible to store results of simulations in databases (e.g. [8]), this standard will be relevant for theory as well.

In standards such as these, that were originally aimed at observational data sets, but might be relevant for corresponding synthetic results, it is important to be aware that the protocols are flexible enough so that they not preclude the simulated data sets to be discovered and queried. As an example we mention that it should be possible to pose queries that do not include observational-only parameters, such as position on the sky.

Data Models

The Space Time Coordinates (STC, [9]) data model has minimal support for dealing with simulated data. In particular, when describing the coordinate system for a given data set, one can use a value of RELOCATABLE to indicate that the data set is the result of a simulation. It is unclear from the STC specification exactly what type of simulations is envisioned to use this. It seems likely that simulated observations are intended, not the generic kinds of simulations described elsewhere in this whitepaper.

Similarly the characterisation data model [10] claims to support observations, but a comparison to the requirements from for example the SimDB effort shows that it is in fact not quite suited for the most generic types of simulations. Again the hidden assumptions behind these claims are that the simulations produce results similar to the observational data products that were mainly targeted.

Semantics

Data models for simulations are in general less explicit than the data models for images and spectra for example. To allow publishers nevertheless to add standardised and semantically meaningful information when registering a resources we need to provide them with a vocabulary that contains the terms they could use to describe the resource. The Semantics WG has developed various relevant standards that support this. First they specify how such vocabularies should be defined, namely using SKOS [11] . It has been shown in the SimDB effort that this standard is perfectly suitable for incorporation in data models.

Second there are some efforts in the Semantics working group aimed at defining meaningful vocabularies. Of particular interest for theory is a vocabulary for describing astrophysical objects and processes. This is under construction in the Semantics WG and will provide a formally agreed on list of words that can be used to describe the objects being simulated and their properties. More specialised vocabularies containing words for computational concepts may have to be defined by the Theory Interest Group themselves.

VO Query Language

The Astronomical Data Query Language (ADQL, [12]) defined by this WG is a language for querying databases, based on SQL92[24]. As stated above, simulation data can be stored in a relational database just as well as observed data so this standard is of relevance. ADQL adds features to SQL92 that are specific to querying objects on the sky and that can be seen as an extension to the Simple Cone Search protocol. These are of no direct relevance to theory, but one can envision later versions of the language to be extended in different directions, also theory. In fact SimDB assumes that the metadata database will be queried by ADQL through a TAP interface.

Applications

The standards from this WG pertain to client side applications. The Simple Application Messaging Protocol (SAMP) allows one to tie multiple applications together and interchange messages between them. This is useful if one application does not provide all support required and will be relevant for theory as much as for observations.

An example of this that was shown during the DCA workshop on *how to publish data in the VO*, June 2008, ESO Garching, Germany[25]. As illustrated in Figure 6, TOPCAT[26] was used to query the Millennium Database web site[27] (see below) and the results were sent to VisIVO[28] for display using the fore-runner of SAMP, the Euro-VO VO-TECH developed PLASTIC[29].

24 http://savage.net.au/SQL/sql-92.bnf.html
25 http://www.euro-vo.org/dcaworkshop2008/
26 http://www.star.bris.ac.uk/~mbt/topcat/
27 http://www.g-vo.org/Millennium
28 http://visivo.cineca.it/
29 http://plastic.sourceforge.net/

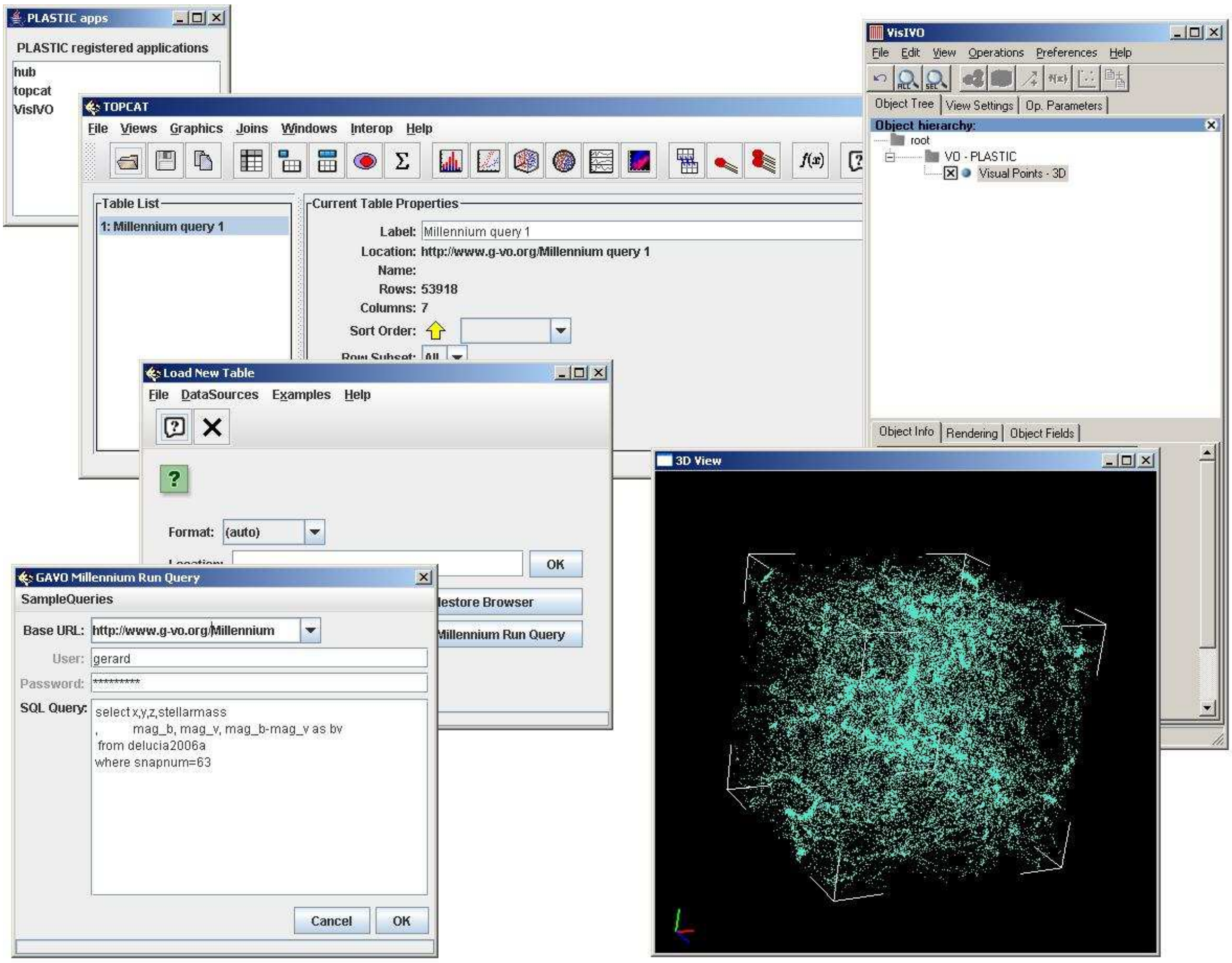

Figure 6 Illustration of application interoperability with SAMP applied to theoretical data.

Grid and Web Services

The standards from this group are very general and applicable to simulations as well as observations. But in particular this WG's efforts are of relevance when wanting to publish services that perform simulations, rather than giving access to their results. For example, the Universal Worker Service[30] standard for describing one's service allows services to be discovered and called, possibly through user interfaces generated from the description.

VOTable

Any tabular dataset can be stored in a VOTable, so also those resulting from simulations. One issue that has been discussed in the past is whether the metadata in the FIELD elements as embodied by the Unified Content Descriptors (UCDs[31]) are rich enough. See [13] for a first discussion of this issue. Ultimately this is a question for the Semantics group, and a procedure is in place for enriching the UCD list in a controlled way as required [14] .

30 http://www.ivoa.net/internal/IVOA/AsynchronousHome/
31 http://www.ivoa.net/Documents/latest/UCDlist.html

5.2 Theory in national VObs projects and elsewhere

The fact that within the IVOA no theory specific standards have been developed so far, has not stopped people from investigating how to publish theoretical results on line, and implementing corresponding services. Some of that work was done within the context of national VObs projects. Those in particular have in general tried to implement services which aim to follow the VObs philosophy and have been in general focused on the research fields covered by the national community of theorists, or some projects in challenging astrophysical computation. Here we describe those and other activities.

The Astrophysical Virtual Observatory (AVO) was an FP5 European project which defined some use cases with theory elements. In the Science Reference Mission (SRM) document[32] edited in April 2005, the reference to theory is mainly focused on the comparison of observational datasets with models. Only 2 over 10 science cases addressed in the SRM did not include the use of the theory to interpret the observations. Two dealt with the use of catalogues of mock observations whereas the rest was about Spectral Energetical Distribution interpretation based on spectral or isochrone modelling.

AstroGrid[33] has compiled a list of VObs use cases[34] some of which have a clear theoretical component. Most of these use cases deal with the comparison of observational datasets with theoretical models. AstroGrid has developed the Common Execution Architecture (CEA): a way of defining and registering the interface of an application such that it can be executed on a server, or on a grid, under control of a standard web-service. This arrangement can execute registered simulation-codes on demand. The CEA is part of the EuroVO infrastructure and has been used outside AstroGrid (e.g. to run the Meudon PDR code on a cluster in Paris). CEA may later be developed as an IVOA standard.

The German Astrophysical Virtual Observatory (GAVO[35]) has from its inception paid special attention to theory. The first results of this were prototype web applications providing access to visualisation codes acting on hydrodynamical simulations of galaxy clusters[36] and to an online simulator of the CMB sky at the Planck and WMAP frequencies[37]. More recently the Millennium Database service[38] was published and has been very successful. This web application provides online access to the results of the Millennium Simulation stored in a relational database [8] . So far over 250 users have registered. Over 160 papers have been written based on the results of the Millennium simulation, roughly half of which have used the GAVO web service to access the results.

The France VO (FVO[39]) has an active workgroup devoted to theory, and is participating actively to TIG activities. Various participating projects have created interesting services which serve as prototypes for many of the ideas in the

[32] http://www.euro-vo.org/pub/fc/cases/srm.pdf
[33] http://wiki.astrogrid.org
[34] http://wiki.astrogrid.org/bin/view/VO/UseCaseList
[35] http://www.g-vo.org
[36] http://www.g-vo.org/hydrosims
[37] http://www.g-vo.org/planck
[38] http://www.mpa-garching.mpg.de/millennium
[39] http://www.france-vo.org/

theory-VO. Examples are the Besancon galaxy model[40], the GalMer[41] database of galaxy merger simulations, the Meudon model of Photo-Dissociation Regions[42] and the Horizon database of cosmological simulations[43]

The Italian VO[44] also has various projects related to theory organised as the Italian Theoretical Virtual Observatory (ITVO[45]). Inside the ITVO project, two WEB portals have been developed (ITVO@Trieste and ITVO@Catania), that provide access to theoretical data stored in three archives whose physical location is in various institutes. They offer also many services such as cut-out and preview of the data by using "VisIVO[46], that is a visualisation and analysis free software for astrophysical data. VisIVO can handle both observational and theoretical data, and now there exist also a Server and Web version. Inside the ITVO, BaSTI[47] has been developed, the "Bag of Stellar Tracks and Isochrones". BaSTI is a web portal which allows for the first time the search of stellar simulated data and population synthesis models via a relational database, and provides also a number of population synthesis tools. In addition, it will soon offer the possibility of plotting the output data by using STILTS[48]. ITVO is also active in the TIG both in the development of the SimDB and SimDAP standards and in the evaluation of the S3 proposal.

The US VO (NVO[49]) has been pursuing theory related projects from the very beginning. Peter Teuben has maintained a web page[50] with links to theory VObs-like projects and services and Rick Wagner is co-chairing the SimDAP effort in the IVOA as well as setting up a theory SkyNode in San Diego[51].

The Spanish VO (SVO[52]) and the ESA-VO project collaborated in the definition of the first ever protocol to access Theoretical Spectra in the VO context. Together they defined the initially called Theoretical Spectral Access Protocol (TSAP, see also 5.1.4) later merged with the Simple Spectral Access Protocol (SSAP) to allow access to theoretical spectral services. ESA-VO also implemented access to registered theoretical services in their VOSpec tool, allowing display and handling of both theoretical and observational spectra within one single application. Most recently SVO has produced the S3 proposal (see 5.1.4) for a standard for publishing micro-simulations. SVO has also been actively pursuing science cases using VObs techniques. In particular, the development of VOSA[53], an application able to use observational and theoretical data from the VO to generate scientific results (see [6]).

---

[40] http://www.obs-besancon.fr/modele/
[41] http://galmer.obspm.fr/
[42] http://www.france-ov.org/twiki/pub/GROUPEStravail/ObsVTheorie06avr5/FLP_workshop06_twiki.pdf
[43] http://www.projet-horizon.fr
[44] http://vobs.astro.it/
[45] http://wwwas.oats.inaf.it/IA2/ and http://itvo.oact.inaf.it
[46] http://visivo.cineca.it/
[47] http://wwwas.oats.inaf.it/IA2/BaSTI/ or http://www.oa-teramo.inaf.it/BASTI
[48] http://www.star.bris.ac.uk/~mbt/stilts/
[49] http://www.us-vo.org/
[50] http://bima.astro.umd.edu/nemo/tvo/
[51] http://lca.ucsd.edu/data/sca/
[52] http://svo.laeff.inta.es/
[53] http://svo.laeff.inta.es/theory/vosa

In Appendix B we present a table with examples of projects that have a strong theory-VObs flavour. Some of these have been described above as efforts of national VObs projects. Some are implemented on-line, others are science projects with strong VObs-like features, but are so far only represented by publications and could be seen as use cases. This list is by no means comprehensive, but will be used as reference list for the conclusions and recommendations section

5.3 EuroVO-DCA

The EuroVO-DCA project has devoted a work package on theory and the present whitepaper is one of its deliverables. Here we summarise the other theory related activities of the project.

5.3.1 Partner projects

Some of the partners in EuroVO-DCA have devoted some of their resources to theory. Many of the projects listed in the previous section were sponsored by EuroVO-DCA. Other support included paying for participating in the IVOA interoperability meeting in Beijing (May 2007). But also individual visits and some impromptu workshops were supported. An example of the latter is the SNAP workshop[54] that WP4 organised in cooperation with the IVOA Theory Interest Group.

5.3.2 Theory Census

During the 2008 the EURO-VO DCA project carried out a census of the astronomical data centres throughout Europe. Parts of this census were two questionnaires concerned with data centres hosting theoretical data and services.

The details of the full survey can be found in the WP2 deliverable Census of European data centres (deliverable D5, [15]). Here we present a summary of those details of the census that might be relevant for the present white paper.

As a result, 16 institutes identified 24 archives and 10 services.

There is a range of types of simulations. They can be classified as large-scale simulations and micro simulations (c.f. sec. 4.3). The former are primarily related to cosmological simulations of dark matter halos and galaxy mergers. The latter ones model and simulate the micro-physics of mainly stellar spectra and stellar populations. Other simulations are related to molecular line libraries, astrophysical jets, stellar accretion disks, quantum mechanical calculations of Stark broadening, and web links to other theory web sites.

The most interesting difference with respect to observational archives/services is the type of data provided. Simulations in general do provide more physical data than observational data. Their structures are different from those of classical images or spectra. Whereas e.g. spectral line shapes are data types relevant to both observations and simulations, there are many other, non-classical data types like hierarchical merger trees from galaxy evolution, isochrones for stellar evolutionary models, quantum-mechanical parameters, particle properties (coordinates, velocities, masses, etc.), html links to other theory web sites, and

[54] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/GarchingSNAPWorkshop200704

physical parameters of jet simulations. Theoretical services are mainly related to the observational interface to the simulations.

The interest in adopting VObs standards for metadata and data access is high. The level of VObs compliance is varying, but almost all archives offer at least VOTable access and simple access protocols (SSA, SIA). Two archives have implemented the current versions of the theoretical database model and data access (SimDB and SimDAP, formerly termed SNAP). Though, none of the archives have a formal requirement for VObs compliance. And as mentioned in section 5.1.4, about 14 implementations exist of the S3 standard proposal.

5.3.3 WP4 Workshop

Theme

As part of the process of understanding the requirements for the integration of theoretical astrophysics (i.e. simulations) into the VObs, it is important to bring together users and developers. The EuroVO-DCA workshop "Theory in the Virtual Observatory", held in Garching in April 2008, was aimed towards this goal. Here is an extract from the announcement describing the motivation for the workshop:

"The Virtual Observatory is an international astronomical community-based initiative. It aims to allow global electronic access to the available astronomical data archives of space and ground-based observatories, as well as simulation databases. It also aims to enable data analysis techniques through a coordinating entity that will provide common standards, wide-network bandwidth, and state-of-the-art analysis tools. VObs efforts have mainly concentrated on observational data archives and services, but recently results from theoretical research have started attracting more attention. The main goal of this workshop was to outline needs and challenges that theoretical astrophysics will be facing in the coming years, and to identify how the unique capabilities intrinsic to the Virtual Observatory concept can meet them."

To achieve its goals, the workshop targeted two main audiences:

- The first is the community of theoreticians who wish to take the initial steps necessary to publish their results online, making them accessible to the world-wide community, or who want to make available on-the-fly services to extract or analyze theoretical data.

- The second audience comprises users of such published data sets and services. This includes other theoreticians, but a particular goal is to support observers who seek theoretical data and/or on-line services to make possible detailed comparisons with their observational results.

The Workshop Topics were 3+1D simulations, micro-simulations, theory-theory interoperability, theory-observational interface and computational infrastructure. All sessions were followed by discussions.

- An introductory session was aimed to setting the stage of the virtual observatory in general, its results so far and the special emphasis on theory.

- The second session dealt with what are generally called *cosmological simulations*, those that aim to directly model the evolution of a part of 3D space. One feature of these is that in general the results are very large and very different from more common observational standards and there is in general no pre-existing data standardisation. Presentations were given about the largest N-body simulations to date, both pure dark matter and hydro-dynamical, and thoughts, some implemented, how to publish these to the community.

- The third session was titled micro simulations. Here we heard discussions on simulation approaches in the field of stellar evolution codes, population synthesis, PDR codes and attempts to publish these. Here a clear result was that different sub-fields want to use each other's results and standardisation efforts have been considered, though not yet implemented.

- One of the promising features of the Virtual Observatory is that different archives and services might be combined together by publishing them in an interoperable manner. For observational archives this is quite obviously of use, as information on the same objects may be available in different observations form the same part of the sky. For theory this is less obvious. We had two sessions to investigate interoperability, one on theory-theory, one on theory-observations. Theory-theory interoperability was represented amongst others by presentations on code comparison and code combination projects. See Figure 7 for an example how this might work in the field of stellar evolution. Theory-observational interoperability deals with the various means by which observers can use the results of theory, be it via virtual telescopes or advanced fitting methods.

- The final session dealt with computational infrastructure. Here we heard talks about applications of the Grid, relational databases, VObs aware tools to the publication and analysis of theory data for a larger audience.

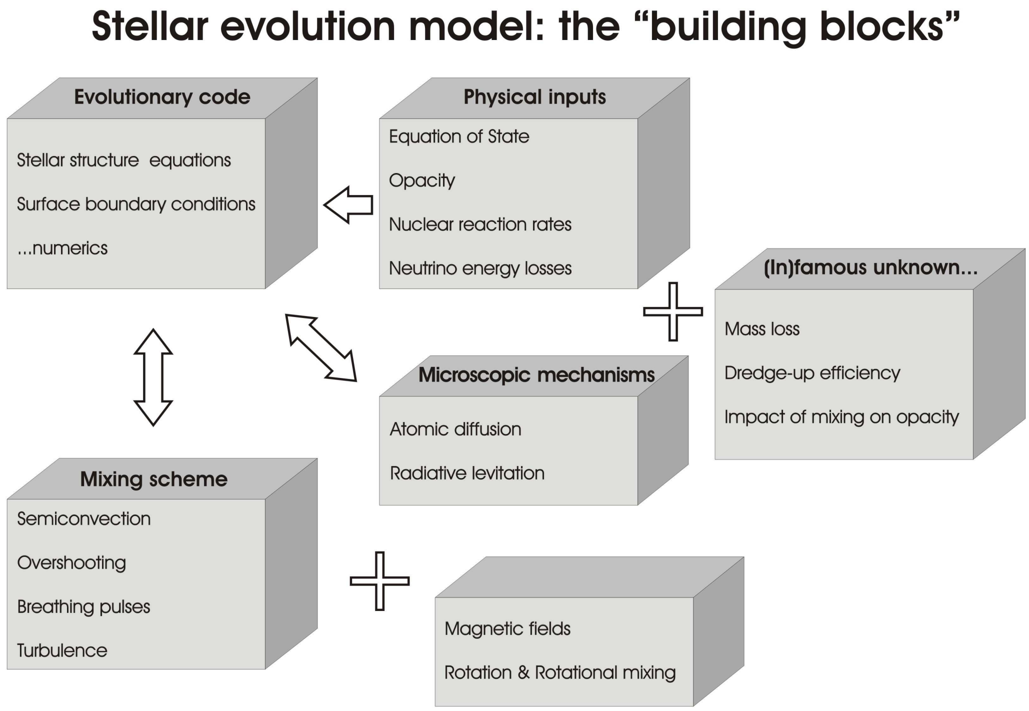

Figure 7 Example of theory-theory interoperability from the field of stellar evolution taken from the presentation by Santi Cassisi during the EuroVO-DCA workshop on "Theory in the Virtual Observatory".

Participants

A total of 49 participants attended the workshop (23 invited speakers, 5 contributed talks). They came from several European and EU-associated countries, as well as, from abroad (USA, Japan, and Mexico). The broad attendance bridges the European and global community.

Results

Each of the sessions was followed by a discussion. Interest in VObs techniques and interoperability is very high. As recent work shows, though the theory metadata models and data access standards are not fully mature yet, it is already possible to publish simulation data and services using the current state.

Here we give a brief summary of the final discussion:

- In order to foster the evolution of the theory standards, it is important to get theoretical researchers involved into the IVOA standardization process. It was noted that the involvement should not focus on the technical level but on the scientific/work requirements.
- For (potential) users to get a better view of the current situation it was mentioned that a cookbook/tutorial style documentation is urgently needed.
- Micro-simulations have turned out to deserve a renewed focus (after the support for theory spectra in SSAP) of the VObs efforts in terms of data models and interoperability. Especially the interoperability is of great interest, since individual micro simulations can build on each other.

- Are there two classes of theoretical products/services? On the one hand, data services related to huge data sets (cut-out, mock observation, rough analysis, quick look, etc.) should be installed close to the data. Most of the large datasets are produced by almost standalone codes (N-body, hydro, etc.) for 3+1D simulations. On the other hand, "quick" codes could be implemented as "on-the-fly" services and might be inserted in workflows (typical of micro-simulations).
- There might be synergy between large surveys (e.g., SDSS) and large simulations in terms of internal consistency, uniformity, and size that prevents practical downloading of the entire dataset. Perhaps some effort should be made to adapt existing tools and applications that are used for survey data sets to theory datasets instead of "reinventing" tools just for theory.
- There are many concerns about credits, description of the physical assumptions, and about (the need for) preservation of the archives. These require intense interaction between experts in the fields and the VObs partners.

Proceedings

The detailed content of the workshop can explored at the EuroVO-DCA wiki (http://cds.u-strasbg.fr/twikiDCA/bin/view/EuroVODCA/WCAWP4Workshop). There will also be proceedings to publish in early 2009 in Memorie della Società Astronomica Italiana – Supplementi, which will give a valuable resource to assess the current state in the theory & VO communities.

6 Conclusions and Recommendations

The description of the state of the TVO in the previous sections shows that this is an active area of development, both regarding the developments of the framework itself, and the willingness of scientists to participate in the efforts.

It was shown that publishing one's simulation results online in a VObs-like or even VObs–compatible manner is possible, and moreover can be a very fruitful endeavour. When the data are of interest users will appear and the impact of the published works will increase accordingly.

There are many more opportunities beyond large scale simulations than have so far been explored, both formally and through *ad hoc* service implementations. In particular the area of micro-physics simulations appears to offer a large and diverse range of possible services that can be of great benefit to other theoreticians and observers alike.

It is thus clear that works remains to be done, on standardisation in the IVOA, on the implementation of standards by data centres and on the active participation of scientists, both as producers and users of VObs services. In the following sections we will draw more detailed conclusions and make recommendations based on a working definition of what constitutes the VObs framework.

These recommendations should be seen also in the context of the Astronet Infrastructure Roadmap[55]. The VObs effort is well appreciated as an important aspect of a strategic plan for European astronomy. Both sets of recommendations complement each other ideally.

6.1 The *VObs framework*

To describe how to fit theory into the framework of the VObs we will need to define what constitutes this framework. What decides whether a certain data collection is "in the VObs", or whether a certain service is "VObs-compatible", what should one do to "participate in the VObs"?

Our definition of the framework consists of a list of requirements. These requirements are aimed at a community consisting of parties (people and organisations) playing particular roles in the process. We identify the following roles:

- **Scientist (Sci)**: producer/publisher/user of VObs resources. Includes also institutes hosting archives and services. Where relevant we will use these more detailed roles:
  - **Prod**: the scientist as producer of scientific results. The existence of parties playing this role is basically always assumed. We will do the same here and not make explicit statements about this role.

[55] http://www.astronet-eu.org/IMG/pdf/Astronet-Book.pdf

- **Pub**: the scientist publishing results online.
- **User**: the scientist using online/VObs resources for his/her research.
- **Inst**: the scientific institute, organisation providing resources to do actual work.

- **IVOA**: the International Virtual Observatory Alliance, standards body for the VO. One particular sub-role here is
  - **TIG**: the theory interest group of the IVOA.
- **National VObs (nVObs)**: a funded VObs project, possibly a partner of the IVOA. Seen as the glue between the IVOA and some national *hinterland* of scientists.

To participate in the VObs effort requires one to play at least one of these roles and to conform to one or more of the requirements applicable to this role. The list we use is non-exhaustive and open to argument about the MUSTs and SHOULDs (see [16]). We see it as a convenient set of features to which TVO efforts by participating parties (roles) can be matched and from which recommendations can be derived. The following list gives the requirements and indicates which roles are likely most relevant.

1) Participation (**Sci**).

   i) Publication: Scientific results MUST be made available online[56]. (**Pub**)

   ii) Usage: Online resources SHOULD be used by scientists. (**User**)

2) Standardisation

   i) Definition: Interoperability standards MUST be defined. (**IVOA**)

   ii) Participation: The scientific community MUST be involved in the standardisation process. (**Sci**)

3) Implementation

   i) Mapping: metadata SHOULD be assigned to published data sets and services according to IVOA data model standards. (**Pub**)

   ii) Protocols: services SHOULD implement IVOA standard protocols. (**Pub**)

   iii) Data formats: Services SHOULD provide results in an IVOA-standard data format. (**Pub**)

   iv) Applications: VObs aware client tools SHOULD be written. (**nVObs**)

[56] We will refer to such online, published results as resources. See **Erreur ! Source du renvoi introuvable.**For the interpretation of the term "resource" in the VObs context.

v) Registration: Online resources MUST be registered in an IVOA compatible resource registry. (**Pub**)

4) Resource allocation

i) Standardisation efforts MUST be supported. (**Inst, nVObs**)

ii) Implementation efforts MUST be supported. (**Inst, nVObs**)

iii) Outreach SHOULD be supported. (**Inst, nVObs**)

The requirements under 1) simply claim that the VObs is created for use in the scientific community and therefore scientists will have to participate in it. Without scientists' participation the VObs will fail. The two sub-items of 1) characterise boundary conditions for the VObs to work.

1 i) says that scientific results must be made available, otherwise the VObs has no content. This requirement does not yet imply any standardisation, only that the results can be accessed and is therefore relatively easily obeyed. A simple web page with links to files for download is sufficient, as is a simple cgi-bin script for more active support. Even FTP qualifies. Though conformity to this requirement is necessary (hence the MUST), it is not sufficient to claim compatibility with the VObs framework.

1 ii) says that the published resources should be used to further scientific work. Usage will become a MUST if (as we expect) the VObs is relevant to scientists needs. For that is the claim of the VObs, that it *facilitates*, or even *enables* scientific research. It can only do so if it is used to do new science, not if it only serves as a repository of old results without usage.

2) embodies the generally accepted claim that for the VObs to fulfil its promises standards have to be defined. We claim here that it is important that scientists are also involved in this process. This may be limited to providing scientific use cases, but also testing and possibly more technical feedback should be provided and incorporated into the process.

3) simply claims that the standards should be implemented. We make a distinction between registration, which MUST be part of the framework and the implementation of the other standards, which "merely" SHOULD be done. We believe that as long as standards for a certain type of service or data do not exists, services can still be published online, but are part of the VObs (if and) only if they are registered.

4) finally expresses the fact that the VObs does not come for free. Explicit support structures such as national VObs projects are needed both to provide the human resources for standardisation, implementation and curation as well as hardware and software infrastructures. But it also includes the allocation of resources in a scientific funding proposal for VObs related efforts.

The following sections will summarise the status of the TVO in terms of these requirements and will define corresponding recommendations where we feel further attention is required. The recommendations are aimed at different audiences. Where applicable we will indicate these here by the roles defined above.

We add a special role to the Euro-VO AIDA project. Being explicitly funded with "networking" duties, a project like this can have an important role in the organisation and funding of international meetings. This was seen already in the impromptu SNAP workshop and meetings co-organised by EuroVO-DCA. EuroVO-AIDA is also assigned to provide service activities for data centres, users and outreach, and Joint Research Activities, in particular for the development and maintenance of interoperability standards.

- **AIDA**: in its role as facility centre, data centre alliance and technology centre.

6.2 Participation

6.2.1 Publication

As section 5 and Appendix B show, there are many examples of theory resources that are available online. These span a large range of scales and cover many areas of theoretical astrophysics. There is a large range in complexity of the type of services, ranging from downloads simple web pages to complex web applications combining database search, standards implementation and dynamic visualisation.

Some examples in Table 1 are represented only by a publication reference, the data products are not published online. And clearly there must be many more potential data sets and services out there that deserve to be published. Hence we recommend

R **Prod+Pub**: It is worthwhile to spend some effort on publishing one's results. There are many scientists out there who may be interested in your data for their work, often for very different reasons than the data were produced for originally.

R **Prod**: Once results are published, make sure to *publicise*. This goes beyond the IVOA registration to be discussed below. Ensure that publications contain links to the services, possibly even make a special announcement of it, for example in astro-ph (see [8] , [17]), and mention them in contributions at conferences.

R **TIG**: Public relation should be performed for the VObs amongst theory community. Urge them to publish their results online, in whatever form.

R **nVObs+IVOA**: It would be good if IVOA would explicitly include simulations whenever the VObs is explained. Currently many quotes limit themselves to "space and ground-based observatories[57]"

6.2.2 Usage

From the example of the Millennium Database[58] we see that interesting and publication worthy science can be done using online theory services. One problem for potential users is that they must be able to find the services.

[57] http://www.euro-vo.org/pub/general/intro.html

R **TIG**: Public relations should be made for the VObs amongst theory and observational community. Explain about existing services and provide use cases.

R **nVObs**: Include theory services in VObs hands-on and other workshops. Even if they are not (yet) standardised.

6.3 Standardisation

Standardisation lies at the core of what most would call the "framework of the VObs". And it is here that most work is required for theory. To develop standards that can do justice to the large variety of potential and actual services is a difficult task.

6.3.1 Definition

Standards should be defined for theory data sets, but which standards? Where can standardisation offer the potential for new science? This is the first and main task for the TIG, to find areas where standardisation may help the scientists.

R **TIG**: Investigate which areas might benefit from standardisation.

One potential area was extensively discussed during the EuroVO-DCA theory workshop. In the field of stellar evolution, stellar atmospheres, stellar population synthesis various specialisations touch upon each other and already from within these communities the ideas of standardisation of data formats for example had been discussed.

R **TIG**: engage the communities of stellar evolution, stellar atmospheres, stellar population synthesis and extract requirements for standardisation.

To do so may require attending scientific meetings, or organising small workshops with attendance from both VObs engineers and such scientists. Such workshops can be defined by national projects or at European level, possibly together with the TIG.

R **nVObs+Sci+AIDA+TIG**: organise workshops where parties from different communities can meet for standardisation discussions.

The actual standards definition will be under the control of the IVOA, with theory standards being overseen by the theory interest group.

R **IVOA+TIG**: subsequently define, in collaboration with these communities, the standards.

Under construction are the SimDB and SimDAP standards.

R **TIG**: complete SimDB+SimDAP specification

[58] http://www.mpa-garching.mpg.de/millennium

- **R** **TIG+Sci**: interact with scientists to ensure SimDB metadata is compatible with their idea of metadata and test whether those can be mapped to the SimDB data model.

- **R** **nVObs**: SimDB has characteristics of a registry and requires more sophisticated support than simple data access services. Consequently we recommend that national VObs-s treat these as they do registries and provide centralised implementations where scientists can register their simulations.

- **R** **TIG+nVObs**: it would be useful to produce an out-of-the box implementation of the SimDB standard and provide this to the community. As extra use case of such a service can be mentioned a local simulation repository for institutes where many simulations are being managed.

The one accepted IVOA standard for which explicit attention was given to theoretical data products, namely SSAP, has been used extensively for theory spectra. This is good news, as it shows that the approach works.

- **R** **IVOA**: ensure that theoretical data products of the appropriate type are properly supported in the future versions of "observational" DAL protocols such as SIAP and SCS.

A weak point of applying SSAP to theory spectra is that the protocol and its associated data model are not very expressive with regards to the provenance and parameters of the spectra. Attempts have been made to see whether the SimDB data model is suitable for this type of simulations. It seems to be generic enough to do so, though some changes must be made to it.

- **R** **IVOA+TIG**: investigate the evolution of SimDB (and SimDAP) for supporting non-cosmological types of simulations. This should be done in a later version from the currently proposed one.

The SimDB data model is rigorous, but because of this rather complex. Recently an IVOA Note was written on a "simple, self-describing service protocol" (S3, [18]), which claims to cover a subset of the metadata of the SimDB data model, and to be particularly useful for micro-simulations, something for which SimDB+SimDAP were not originally designed. From the S3 *Note* it seems in principle to have application beyond theory, for example it seems it might serve as an extended simple cone search, allowing queries on other parameters but position.

- **R** **TIG+IVOA**: Coordinate the assessment of the S3 note in the IVOA. Pay particular attention to its applicability for publication of and interoperability between results of micro-simulations.

Most tools that deal with spectra and allow users to query SSA services are not very flexible regarding theory specific query parameters. VOSpec is a positive exception in that it allows one to discover and manipulate the specific input parameters for each of the service.

R **nVObs**: update interfaces of client tools to support querying on custom parameters.

In the SimDB specification an important role is played by semantic vocabularies. These are assumed to provide lists of common concepts that can be used as value to metadata attributes. This has to be worked out further. Requests have come from the community as well (e.g. Gus Evrard, private communication) about the desire to have common models for, for example, dark matter halos, galaxy clusters that go beyond mere words, but also include their possible properties and relations to each other. This requirement goes beyond simple vocabularies, in the direction of full ontologies.

R **IVOA+TIG**: provide semantic vocabularies, ontologies and/or data models that describe astrophysical concepts.

Note that vocabularies like this are also the appropriate place to compare observations and simulations that have nothing in common but the target object under investigation. A simulation of a galaxy merger is clearly of interest to an observer who investigates galaxy mergers, and to search for these in a repository of simulations it is good if common terms are used in both worlds, whether theoretical or observational.

6.3.2 Participation

It is important that standards are being developed that support science cases. It is also important that during the process of standards development the target community is being asked for feedback on the standards. At the current time this is especially important for the SimDB and SimDAP standards which are close to a first draft.

R **TIG**: get feedback on the SimDB and SimDAP standards from the scientific community.

During the DCA WP4 workshop in Garching, a clear science case emerged for interoperability between different branches of the micro-physics community. For example, stellar population synthesis requires results from models on stellar evolution and stellar atmospheres

R **TIG**: the micro-physics community has clear and interesting use cases for interoperability. Examine this further and propose or develop a standard to deal with this.

During subsequent discussions the suggestion was made that such use cases should be discussed during scientific meetings of the communities as well, as those are perfect opportunities where the scientists get together. Most of these will not visit IVOA interoperability meetings.

R **Sci+TIG**: discuss standardisation use cases during community specific science meetings.

6.4 Implementation

6.4.1 Mapping

Providing metadata in an appropriate standard is an often underestimated part of the implementation of standards. It requires knowledge of both sides of the effort, the actual science domain and the (meta-) data model to which it is mapped. We see it as the task of national VObs projects to assist in this task, but potentially also of the TIG. One of the first tasks of the TIG was to find national representatives for theory. These have not been used explicitly up till now, but maybe that could change once theory becomes more embedded in the IVOA process. One of their tasks could be to be the natural contact for technical issues like this, if only to further delegate to appropriate experts. One could also think of organising specific "support" Workshop(s) in the EuroVO-AIDA context.

R **nVObs+TIG+AIDA**: provide feedback and assistance on questions regarding the mapping of simulation metadata to the appropriate models.

6.4.2 Protocols

There are not yet any *accepted* theory-specific standard protocols defined by the IVOA, but on the short term some are expected, in particular SimDB and SimDAP, while formal standardisation work on micro-simulations can be expected soon, motivated by the S3 Note. Furthermore the Simple Spectral Access Protocol already pays explicit attention to theory spectra and some tools, in particular VOSpec are able to plot these.

R **Sci**: when one wishes to publish synthetic spectra to the VObs, use the SSA protocol

R **Sci**: when one wishes to publish cosmological simulations to the VObs, and once the standards have been completed, register the simulations in a SimDB and use the SimDAP protocol for services providing access to these simulations.

R **Sci**: when one wishes to publish (micro-) simulations that are not covered by SSA or SimDB/SimDAP, one should follow the IVOA standardisation developments in this field. Motivated by S3 standardisation efforts are under way, and we urge scientists to participate in these efforts, for example by implementing prototypes.

R **Sci**: when wishing to publish your data in any standard, ensure that the metadata for a simulation are generated as part of the simulation and post-processing pipeline.

R **nVObs + AIDA**: commit to providing persistent SimDB implementations for use by scientists who wish to register their simulation results and related services. This might be a task for the Euro-VO Facility Centre.

R **Sci**: when implementing services that can be used in a pipeline, take where appropriate the GWS standards into account: Universal Worker Service, VOSpace, CEA.

R **Sci**: when implementing non-public service requiring authentication, investigate the corresponding GWS standard.

R **Sci**: investigate the use of relational databases for publishing large simulation results and publish these using the ADQL+TAP specifications.

6.4.3 Data formats

Most of the online theory services investigated in this whitepaper are able to produce data products in the VOTable and/or FITS formats. On the other hand, most of the data products from especially the larger simulations come in proprietary formats that are not easily transformed to these formats, because of their size, or have more complex data structures than these relatively flat formats allow. SimDAP in particular will include a specification for the data formats of the potentially very large results of the service calls.

Note that if one's native data formats are not compatible with an IVOA standard, it is not required to translate one's own data. Depending on the standard it may only need to be done when (subsets) of the data are delivered by the service.

R **IVOA + TIG**: get requirements for and provide definitions of data structures that are more complex than flat tables (VOTable), or regular N-dimensional grids (FITS). Examples: results of adaptive mesh refinement simulations, tree structures, graphs.

R **IVOA**: consider a result delivery option where files in native data formats can be delivered, if they are accompanied by translation modules in some language. I.e. do not insist that the delivered data is already in a standard format, but leave it up to the end user to decide how to deal with the result.

6.4.4 Registration

When searching a registry for simulations one can use the Content/type attribute of the resource model. The NVO registry at STScI/JHU offers an advanced search capability[59] where the constraint `Type like 'Simulation'` constrains that attribute. The query produces 14 resources, most of which are SSAP services giving access to model spectra. Many of the services described earlier are however absent from this list. One reason for this is simply that the services have not been registered. Another reason is that registered theory services are not correctly or at least completely described. For example the TMAP[60] theory spectral

[59] http://nvo.stsci.edu/voregistry/QueryRegistry.aspx?advanced=true
[60] http://vo.ari.uni-heidelberg.de/ssatr-0.01/TmapArchiveInfo.html

access service has Type='Archive', and does not include 'Simulation' in that attribute.

R IVOA: Ensure that theory standards have their counterpart as extensions in the registry data model.

R TIG: Ensure that scientists are aware of IVOA registries and register their services there.

6.5 Resource allocation

Nothing comes for free, resources will have to be allocated if theoretical results should be incorporated in the VObs. Here we make some estimates on this.

6.5.1 Standardisation

People have to participate in IVOA efforts to define standards. It includes participating actively in the writing of documents and discussions online, or in telecons. It also includes participating in the bi-yearly IVOA interoperability meetings and possibly in special purpose workshops and meetings (for example the SNAP workshop in Garching, April 2007).

But these efforts need to be funded. This funding currently mainly comes from national VObs projects, or institutes that actively participate in VObs efforts, plus some funds from non-sustainable EC-funded projects. Theoretical institutes so far have only played a relatively minor role in the standardisation efforts. But as we claimed above, we feel it is important that astronomers, both theorists and observers, should participate in the development of standards for theory. Hence

R **Sci**: anticipate participation in IVOA standardisation efforts and request resources to do so.

R **nVObs**: actively search out and possibly fund theorists to participate in IVOA standardisation efforts.

6.5.2 Implementation

The implementation and publication of theory services, whether standardised or not, requires further resources, not only personnel, but also in hardware and possibly software. And it does not stop with the implementation. Maintenance, curation, possible help-desks need to be taken into account as well. Providing the resources is in general the task of the service provider.

The service provider may not be the institute that produced the data; it can be a centralised data centre. But even in that case resources have to be allocated at the producer's site to interact with the service provider concerning data structure, metadata, mapping to standards, data formats, documentation, etc. This is a task that is often overlooked

R **Sci**: allocate resources for designing, implementing and maintaining a VObs-like service.

6.5.3 Outreach

The VObs does not come for free. As discussed before, input is required from the scientist for the standardization process and the development of VObs applications. Furthermore, since this is a community effort, the allocation of resources (human and technical) is necessary for the VObs to become reality and a framework for the astronomers/theorists everyday life. On the other hand, the scientists have to become involved in the process. As an outcome of EuroVO science workshops[61,62] and astronomical conferences[63], though the community is aware of the VObs, but participation in and detailed knowledge about the (current state of the) VObs is still limited. It has also to be noted that in this context, the EuroVO-DCA census (sec. 5.3.2 and [15]) is biased towards data centres and service providers (both observational and theoretical) that are already aware of or involved in VObs activities. In order to foster the flourishing of the VObs effort, active promotion is indispensable.

R **IVOA & nVObs**: pro-actively approach individual scientist/groups promoting the VObs. Be present at non-VObs scientific meetings. Use for instance channels like (inter)national mailing lists.

R **Sci**: Realize that already now you use components of the VObs in your research (like, e.g., Vizier, Aladin, etc.). VObs usage is seamless, and thus the VObs framework is not seen from users, even when they use it!

6.6 Final words

We believe we have shown that it is well possible to introduce theory in the VObs framework, and that effort is on-going. But issues remain to be resolved. The architecture of the VObs needs no adjustment to accommodate theory. A resource registry for discovering interesting resources is as relevant for theoretical data products and services as it is for observational ones. Data models are required (and can be constructed) for describing theoretical resources. The approach to data access protocols based on a discovery and a retrieval phase is suitable for theory as well. The IVOA query language is relevant for filtering the sometimes very large data sets. Services can in principle be deployed, protected and chained together in workflows according to the standards of the Grid and Web services working group.

It is therefore mainly in the details of this process that work remains to be done. Particularly the standardisation for data description, data discovery and data access protocols require more work above and beyond what is currently available in the IVOA. The reasons for this are simply that the existing standards so far have dealt with observational data products and theory data products can be very different and very diverse. Work along these lines is progressing though at least for what we call cosmological simulations, but it seems likely that other types of simulations can be served with only minor changes to the models and data access service specifications.

---

[61] Multi-wavelengths astronomy and Virtual Observatory, 1-3 Dec. 2008, ESAC, Spain, http://esavo.esa.int/MultiwavelengthVOWorkshopDec2008/

[62] Astronomical Spectroscopy and the Virtual Observatory, 21-23 March 2007, ESAC, Spain, http://esavo.esa.int/SpectroscopyAndVOWorkshopMarch2007/

[63] JENAM 8-12 Sept. 2008, Vienna, Austria, https://www.cosmic-matter.org/indico/conferenceDisplay.py?confId=6

The question that remains is whether it will be sufficient for the astrophysical community to start participating. This may in the end be the greatest challenge, convincing theorist to make the extra effort of conforming to the standards and publishing their results in the VObs framework. Willingness for participation expressed by the respondents to the Census of European astronomical data centres is a very good starting point. Here it is important to develop successful prototypes and to assess relatively simple manners to take those extra steps. It is in this field that projects such as the Euro-VO are of great importance. Finally, our goal is that the VObs will not be singled-out anymore, since it would have become the way of doing science.

Appendix A: Theory Expert Group

- Santi CASSISI
  Santi Cassisi is Associate Astronomer at the INAF- Astronomical Observatory of Teramo, Italy. He is working in the field of Stellar Evolution and Stellar Population from the theoretical point of view. He is developing suitable numerical codes for Stellar Evolutionary computations and Population Synthesis analysis.

- Miguel CERVIÑO
  Miguel Cerviño has a staff position at the CSIC. He works in the field of stellar population synthesis. He has coordinate different efforts related with the elaboration of databases of population synthesis, atmosphere models and isochrones for its posterior use in analysis tool in coordination of the theoretical data models providers. He participates actively in the Spanish Virtual Observatory, and also collaborates with different VO initiatives in Mexico, Brazil and Italy related with synthesis models and analysis tools for galaxy evolution.

- Gerard LEMSON
  Gerard Lemson works for the German Astrophysical Virtual Observatory (GAVO[64]) at the Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg and the Max-Planck-Institut für extraterrestrische Physik in Garching, Germany. In that project he has been especially active in trying to apply VObs techniques to the publication of theory data sets and services. He is chair of EuroVO-DCA WP4 and was from 2004-May 2008 chair of the IVOA Theory Interest Group.

- Pedro OSUNA
  Pedro Osuna is the ESA-VO project Technical Coordinator. With a background on Theoretical Physics, he coordinates as a System Engineer all ESA Space Based Missions archives and VO technical work at ESAC. ESA-VO project is particularly interested in theoretical spectral models that can be overlaid with observational data. In this context, ESA-VO develops the VOSpec tool for VO spectral handling and analysis.

- Joop SCHAYE
  Joop Schaye is an associate professor at Leiden Observatory in the Netherlands. He works mostly on the formation of galaxies and the intergalactic medium, both at high redshift and in the local universe. He has extensive experience with large-scale hydrodynamical simulations and numerical radiative transfer.

- Nicholas WALTON
  Nicholas Walton is at the Institute of Astronomy, University of Cambridge.

[64] The German Astrophysical Virtual Observatory is supported by a grant from the German Federal Ministry of Education and Research (BMBF) under contract 05 AC6VHA.

He is involved with a number of large scale surveys where he is working to better understand Galactic structure. This involves access to simulations enabling comparison with observations. He is secretary of the International Virtual Observatory Alliance and also the Open Grid Forum's Astronomical Applications research group.

- Hervé WOZNIAK
  Hervé Wozniak is Astronomer at the Observatory of Strasbourg, France. His main field of research is twofold: the chemodynamical evolution of galaxies with simulation techniques, and the modelling of galactic and stellar dynamics. He is member of the HORIZON cosmological project. He is also chair of the IVOA Theory Interest Group (2008-2010), and the F-VO theory working group (2004 -).

Appendix B: Theory-VObs-like projects and services

Table 1 Theory-VObs-like projects and services.

Name	Short Description	Links
Cosmological simulations		
GalMer	Search and visualisation services on archive of galaxy merger simulations.	http://galmer.obspm.fr/
Millennium Database	SQL interface for querying a database of post-processing products of the Millennium Simulation [19] .	http://www.mpa-garching.mpg.de/millennium/ http://www.g-vo.org/Millennium
SimCat	SimDB prototype	http://cds.u-strasbg.fr/twikiDCA/pub/EuroVODCA/WCAWP4Workshop/RickWagnerEuroVODCA.pdf
Cosmic Data ArXiv	Website providing access to simulations and simulation codes.	http://t8web.lanl.gov/people/heitmann/test3.html
Trieste	Web portal for query a multi-level DB; from raw data to post-processing data directly comparable with observations (Gardget2, Enzo data)	http://wwwas.oats.inaf.it/IA2/ITVO/
Catania	Web portal for query and relational DB and produce cut-out data, preview (FLY, Enzo data)	http://itvo.oact.inaf.it
Micro-physics simulations		
BaSTI	A relational DB providing stellar model predictions and population synthesis data.	http://wwwas.oats.inaf.it/IA2/BaSTI
Meudon PDR code	SimDb applied to PDR simulations	http://www.ivoa.net/internal/IVOA/InterOpOct2008Theory/IVOA08_Baltimore_Franck.pdf
Mexican Million of Models Database	Photoionized Nebulae Database	http://cds.u-strasbg.fr/twikiDCA/pub/EuroVODCA/WCAWP4Workshop/morisset.pdf

GALAXEV	*"...a library of evolutionary stellar population synthesis models computed using the new isochrone synthesis code of Bruzual & Charlot (2003)."*	http://www.cida.ve/~bruzual/bc2003
SVO theoretical model server	A server of stellar spectra, isochrones and evolutionary tracks from different authors.	http://svo.laeff.inta.es/theory/docs/index.php?pname=Main
TMAP	*"...a tool to calculate stellar atmospheres in spherical or plane-parallel geometry in hydrostatic and radiative equilibrium allowing departures from local thermodynamic equilibrium (LTE) for the population of atomic levels."*	http://astro.uni-tuebingen.de/~rauch/TMAP/TMAP.html
VOSpec	Client tool providing access to Theoretical Spectra and allowing observational and theoretical services to coexist, providing also simple tools like normalisation of the theoretical spectra to discover qualitative features of the possible fits.	http://esavo.esa.int/vospec/
CMD 2.1	Web service for accessing stellar and populations synthesis data from Padova.	http://stev.oapd.inaf.it/~lgirardi/cgi-bin/cmd
IAU Comm. 35 - Resources	Many more resources for stellar and galaxy evolution studies.	http://iau-c35.stsci.edu/Resources/index.html
Theory-theory interoperability		
MODEST	*"... a loosely knit initiative of various groups working in stellar dynamics, stellar evolution, and stellar hydrodynamics. Our aim is to provide a software framework for large-scale simulations of dense stellar systems, within which existing codes for dynamics, stellar evolution, and hydrodynamics can be easily coupled."*	http://www.manybody.org/modest/
The Aspen-Amsterdam void finder comparison project	*"...the first systematic comparison study of 13 different void finders constructed using particles, haloes, and semi-analytical model galaxies..."*	http://adsabs.harvard.edu/abs/2008MNRAS.387..933C
Cosmological radiative transfer codes comparison project	*"..aims to check and validate the participating codes by performing a range of standardized test problems."*	http://www.cita.utoronto.ca/~iliev/rtwiki/doku.php

CoRoT/ESTA	Grids of models for use in CoRoT mission.	http://www.astro.up.pt/corot/
Stellar Code Calibration (preliminary)	*"A project to reduce numerical and technical uncertainties in stellar models."*	http://www.mpa-garching.mpg.de/stars/SCC/StarCode.html
theory-observational interface		
Virtual Cluster Exploratory	*"…a data access and analysis environment for simulated clusters of galaxies that mirrors the functionality of HEASARC archives and tools."*	http://vce.physics.lsa.umich.edu/
SkyMaker	*"…a program that simulates astronomical images."*	http://terapix.iap.fr/rubrique.php?id_rubrique=221
INAF IA2/ITVO	*"… an ongoing project with the goal of creating a distributed archive of data coming from numerical simulations."*	http://wwwas.oats.inaf.ist/IA2/
CMB sky simulator	*"…provides synthetic sky maps of the Cosmic Microwave Background."*	http://www.g-vo.org/planck/
XMAS	X-ray Map Simulator	
MOPED	Fitting model galaxy spectra to observations.	http://adsabs.harvard.edu/abs/2000MNRAS.317..965H
SEAGal/Starlight	Server for the study of stellar populations in galaxies, providing the results of the analysis of over 500k galaxies from the SDSS with (experimental) VO access.	http://www.starlight.ufsc.br/
MoMaF2	Mock surveys through cosmological simulation.	http://horizon-vo.univ-lyon1.fr/GalICS/Job.do Alternatively see : http://www.g-vo.org/mpasims/MoMaf2
SKADS Simulated Skies	*"…a set of simulations of the radio sky performed at the University of Oxford, suitable for planning science with the Square Kilometer Array (SKA) radio telescope."*	http://s-cubed.physics.ox.ac.uk/
Besancon galaxy model	Model of stellar population synthesis of the Galaxy with virtual observations.	http://model.obs-besancon.fr/

Appendix C: Acronyms

A -

ADASS: Astronomical Data Analysis Software and Systems

ADQL: Astronomical Data Query Language

AMR: Adaptive Mesh Refinement

AstroGrid: The UK VObs project

B –

BaSTI: Bag of Stellar Tracks and Isochrones

C -

CNRS: Centre National de la Recherche Scientifique

CEA: Common Execution Architecture

CMB: Cosmic Microwave Background

CSIC: Consejo Superior de Investigaciones Científicas

D -

DCA: Data Centre Alliance

DAL: IVOA Data Access Layer Working Group

DM: Data Model; IVOA Data Modelling Working Group

E -

ESA: European Space Agency

ESO: European Southern Observatory

EU: European Union

Euro-VO: The European Virtual Observatory project

EuroVO-DCA: Euro-VO Data Centre Alliance (FP6 Coordination Action)

EuroVO-AIDA: Euro-VO Astronomical Analysis for Data Access (FP7 Integrated Infrastructure Initiative)

F –

FP#: Framework Programme

FRI/FRII: Fanaroff-Riley type I and II

FVO or F-VO: France Virtual Observatory

G -

GalMer: Galaxy Mergers

GAVO: German Astrophysical Virtual Observatory

GWS: IVOA Grid and Web Services working group

I -

INAF: Istituto Nazionale di Astrofisica

INSU: Institut National des Sciences de l'Univers

ITVO: Italian Theoretical Virtual Observatory

IVO: International Virtual Observatory

IVOA: International Virtual Observatory Alliance

INTA: Instituto Nacional de Técnica Aeroespacial

L -

LU: University of Leicester

M -

MHD: Magneto-Hydrodynamics

MPG: Max-Planck Gesellschaft

N -

NOVA: Nederlandse Onderzoekschool voor Astronomie

P –

PDR: Photodissociation Region

PLASTIC: Platform for AStronomy Tools InterConnection

S -

S3: Simple Self-describing Services

SAMP: Simple Application Messaging Protocol

SCS: Simple Cone Search

SIAP: Simple Image Access Protocol

SimDB: Simulation Database

SimDAP: Simulation Data Access Protocol

SNAP: Simple Numerical Access Protocol

SPH: Smooth Particle Hydrodynamics

SSAP: Simple Spectral Access Protocol

STILTS: Starlink Table Infrastructure Library Tool Set

SVN: Subversion (version management software)

SVO: Spanish VObs

T -

TAP: Table Access Protocol

TEG: Euro-VO Theory Expert Group

TIG: IVOA Theory Interest Group

TVO: Shorthand for "Theory VO", to be interpreted as" the application of the VObs effort to theory", i.e. the subject of this whitepaper. *Not* to be interpreted as a separate effort from the general VObs!

U –

UCD: Unified Content Descriptor

V -

VO or VObs: Virtual Observatory

VObs.it: Italian Virtual Observatory

VOQL: VO Query Language

W -

WG: Working Group (in context of IVOA)

WMAP: Wilkinson Microwave Anisotropy Probe

WP: Work Package (in context of DCA)

X –

X-MAS: X-Ray Map Simulator

Appendix D: References and links

[1] Lemson G. and Colberg, J. 2004: *Theory in the VO*.
http://www.ivoa.net/pub/papers/TheoryInTheVO.pdf

[2] Ochenbein, F. *et al* 2004: *VOTable Format Definition, Version 1.1*
http://www.ivoa.net/Documents/cover/VOT-20040811.html

[3] Ochsenbein, F. *et al* (1999): *Using XML for Accessing Resources in Astronomy*
http://vizier.u-strasbg.fr/doc/astrores.htx

[4] FITS Working Group (2008): *Definition of the Flexible Image Transport System (FITS), FITS Standard Version 3.0*
http://fits.gsfc.nasa.gov/standard30/fits_standard30.pdf

[5] Hanish, R. and Linde, A. E. 2003 *IVOA Document Standards*
http://www.ivoa.net/Documents/cover/DocStd-20031024.html

[6] Bayo, A. *et al* (2008): *VOSA: virtual observatory SED analyzer. An application to the Collinder 69 open cluster*
A&A 492, 277.

[7] Plante, R. *et al*: *VOResource: an XML Encoding Schema for Resource Metadata*
http://www.ivoa.net/Documents/cover/VOResource-20080222.html

[8] Lemson G. and the Virgo Consortium 2006: *Halo and Galaxy Formation Histories from the Millennium Simulation: Public release of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the LambdaCDM cosmogony*
http://xxx.lanl.gov/abs/astro-ph/0608019

[9] Rots, A. H.: *Space-Time Coordinate Metadata for the Virtual Observatory, Version 1.33*
http://www.ivoa.net/Documents/cover/STC-20071030.html

[10] Louys, M. *et al*: *Data Model for Astronomical DataSet Characterisation, Version 1.13*
http://www.ivoa.net/Documents/cover/CharacterisationDM-20080325.html

[11] Gray, A. J. G. *et al* : *Vocabularies in the Virtual Observatory, Version 1.16*
http://www.ivoa.net/Documents/cover/Vocabularies-20081104.html

[12] Ortiz, I. *et al* 2008: IVOA Astronomical Data Query Language, Version 2.00
http://www.ivoa.net/Documents/cover/ADQL-20081030.html

[13] Shaw, L. & Walton, N. (2004): *UCDs for Simulations*
http://www.ivoa.net/internal/IVOA/IvoaTheory/UCDproposal.pdf

[14] Derriere, S.*et al* (2005): *An IVOA Standard for Unified Content Descriptors*
http://www.ivoa.net/Documents/latest/UCD.html

[15] EURO-VO Data Centre Alliance, *D5 - Census of European Data Centres*

[16] Bradner, S. 1997: *Key words for use in RFCs to Indicate Requirement Levels*
http://www.ietf.org/rfc/rfc2119.txt

[17] Di Matteo *et al* (2007): *Chemodynamical evolution of interacting galaxies: the GALMER view*
http://xxx.lanl.gov/PS_cache/arxiv/pdf/0709/0709.2553v1.pdf

[18] Rodrigo, C. *et* al (2008): *S3: Proposal for a simple protocol to handle theoretical data (microsimulations)*
http://www.ivoa.net/Documents/Notes/S3TheoreticalData/S3TheoreticalData-20081015.pdf

[19] Springel, V. *et al* (2005): *Simulating the joint evolution of quasars, galaxies and their large-scale distribution*
Nature 435 (2005) 629-636