2/26/2004

Theory in the VO

Gerard Lemson[1] and Jörg Colberg[2]

Abstract

In this document we discuss the introduction of theoretical results and data, for example that generated by large scale numerical simulations, into ongoing virtual observatory (VO) activities. The whitepaper has two main target audiences. The first is 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, we attempt 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, we describe how theoretical archives and related services imply interesting new requirements on these efforts.

Our approach in this whitepaper is to adopt the style of a FAQ, in the hope of provoking discussion based on the particular questions asked. In certain instances, only tentative answers are suggested, since it is expected that the reader will provide the relevant answers. In other cases we provide answers which are not necessarily IVOA approved, but are drawn from the authors ongoing participation in the IVOA process together with extensive experience in the analysis of large scale cosmological simulations as well as a background in a business IT environment heavily aimed at data modelling and (web) service design. In some cases, we give answers to questions that have yet to bee addressd in the IVOA, in particular in the area of what it means to publish archives and the role that data modeling efforts can or should play in relation to this.

We further support the creation of a specialized working group dealing with theory data in the IVOA. This group should ensure that other working groups are made aware of theory specific issues and that these are addressed. Ideally the group should propose a road map for the development of theory specific standards and reference implementations based on these. The current document, or later versions thereof, can form the basis of a charter for this working group.

---

[1] German Astrophysical Virtual Observatory (GAVO), Max-Planck-Institut für extraterestrische Physik, Garching, Germany. E-mail: gerard.lemson@mpe.mpg.de
[2] National Virtual Observatory (NVO), University of Pittsburgh, USA.
E-mail: astro@jmcolberg.com

Contents

1 Introduction

The Virtual Observatory (VO) is gaining momentum. At the recent ADASS conference in Strasbourg, it was estimated that the VO was mentioned in two out of three presentations (Bob Hanisch at IVOA Interoperability meeting).

Until now the various VO efforts have confined themselves to observational data products and services. More recently, considerable interest has been shown in including products of theoretical research. This push comes especially from groups involved in large-scale computer simulations who want to publish their results in a VO compatible form. In addition, for a long time observers have asked for theoretical data to compare their data with.

The request has come to various national VO's to support these efforts. The German Astrophysical Virtual Observatory (GAVO) has in its charter the establishment of simulation services using the Grid paradigm (Matthias Stenmetz at AIP in Potsdam). GAVO is furthermore closely cooperating with the group led by Simon White at the MPA in Garching to publish the results of their various simulations. The US (N)VO is planning a theory demo at the AAS meeting in January 2004 (Peter Teuben). A British proposal for funding a "Virtual Universe" project has been partially approved.

Many archives, both observational and theoretical, already make their data products available over the internet. In general a browsing interface is offered that allows users to discover potentially interesting products to download. One may wonder how this is different from publishing one's archives through the VO. And what does it mean to publish theory in the VO? Is it different from the "standard" observational VO and if so how? In this white paper we compose a list of questions such as these, about items we feel will be especially relevant when publishing theoretical data products and services in the context of a VO. We also provide answers, but the main purpose of this approach is to provoke discussion and to provide an explicit medium for feedback. To this end we will send the white paper to a number of people who have previously shown interest or are thought to be interested in a theoretical component to the VO. This group includes prospective users, suppliers of services and data products, and people already working on VO projects.

We want to avoid the impression that we see these efforts as distinct from the existing VO efforts, even though at certain places in the text we may refer to the theoretical VO (TVO). On the contrary, we propose to create a special theory working group within the IVOA. The idea behind this group is to ensure that theory specific requirements are communicated to other working groups of the IVOA, and conversely that existing standards will be incorporated in the various theory specific efforts. The current document could serve to create a charter for this working group that then could be put forward to the IVOA executive committee for approval.

As a final note it needs to be said that the choice of presentation, as a list of questions, leads to a certain redundancy in the text. For example, the fact that the simple cone search protocol is not useful for theoretical archives is repeated a few times in different contexts. When using this text to create other documents this redundancy may have to be removed.

2 Why publish simulations in the VO?

Why might one want to publish one's data online at all? Possible answers are:

- To allow independent checks of conclusions based on theoretical results.
- To allow further analysis by third parties based on the published data.

The reasons for publishing *in the VO* are related to the idea that it would be nice if different archives provided a common interface to the world. Ideally, data are described in a uniform (meta-data) language, data products are delivered using a standard format, and identical services can be used to discover and query the archives and to retrieve data products of interest. What's more, a VO will make it easier to create services that query and combine multiple archives at the same time, something that is often called *federation* of archives. Additional reasons for publishing one's theory products in the VO are:

- To allow comparisons with similar results/methodologies or with the corresponding data by observers/theoreticians.
- To make theoretical results more easily accessible and understandable for observers.

In the future, one might hope that the following reasons provide an even stronger motivation

- Journals may allow/require links to actual data products and/or software used in published work.
- Referees may insist they be able to reproduce certain results through (T)VO services. (Admittedly, this is a somewhat optimistic view of the refereeing process.)

3 How are data published in the VO?

We understand the term *to publish* as follows:

> *To make data products in an archive available through services that are accessible via a VO supplied internet site.*

The *data products* are under the control of the archive. It is up to these archives to specify which data they want to make available. The *services* are supported by the archive and registered with a VO. That means that there is a VO facility that knows of the existence of the service and that can make these available through a web portal.

In principle, there are at least two requirements to data publication. It should allow *discovery* and *manipulation*. The former means that VO users can pose queries through a query service (that may include simple browsing) to discover data products that are potentially of interest for them. This requires that data products be described in some uniform manner (see below for a more detailed motivation). *Manipulation* requires that a user be able to get access to the actual data products themselves. This may imply retrieval by some means (downloading), it may mean a further drilling down in the actual data products, or some other service.

4 How can data products be described?

It is not sufficient if the VO discovery interface was merely a list of websites where archives exist and from where one can download data in some form. Instead, it requires a single view in which different archives can be compared with each other using a *single* interface. The archives must describe their data products *in some detail* to this interface. Such an interface then allows common querying.

To support this infrastructure in a properly defined manner requires that we create a uniform model of the *meta-data* describing the data products (see below for more details). Furthermore, we need a persistent meta-data repository where the meta-data are stored and that supports querying. This repository can be viewed as a more refined *Registry* as defined in the IVOA registry working group. In contrast to the Registry the meta-data repository is aimed at describing the *data* but not the services for retrieving the data. The Registry could/should link into this. It is anticipated that the VO will create reference implementations that allow archives to have their own repository that can be queried through the internet.

5 Why a common data model?

We give two answers. The first is based on the current situation in astronomy: Figure 1

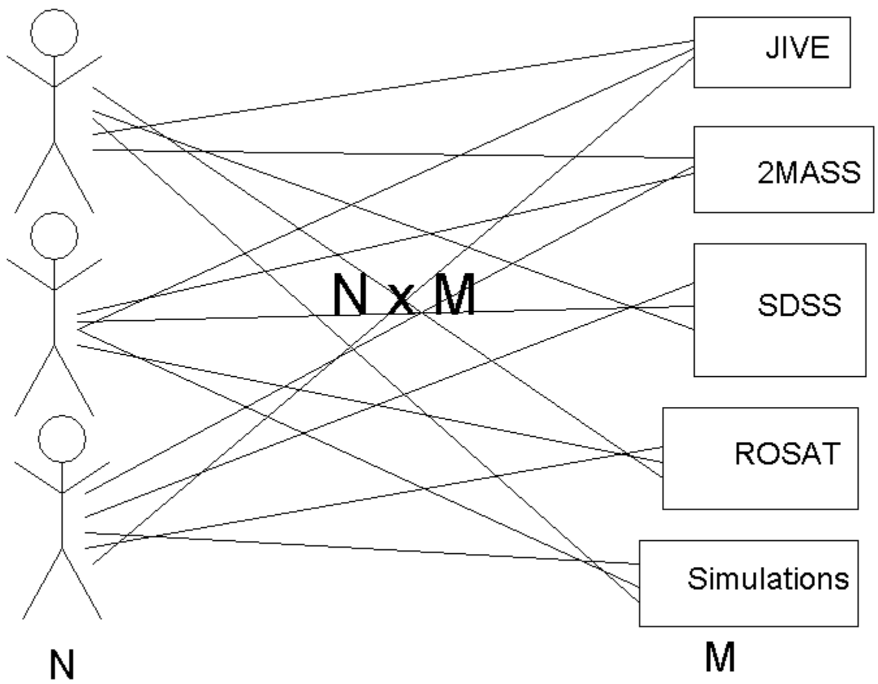

Figure 1: The problem

Many users are interested in many archives. In general, each archive has its own proprietary schema definition and data storage format. This implies that users have to learn each of these schemas and formats if they want to understand the data. The total amount of effort thus scales like the number of users N times the number of archives M.

The introduction of a common data model will improve this situation *drastically*, as the following figure shows. Data that are exposed to the outside world have to conform to a standard model. The idea is that through interfaces defined by the VO users will only have to understand the VO's data model but not the individual underlying models of the archives (where they are different from the VO ones). Archives have to be able to describe their data in terms of that same model, and they have to *map* their schema to it. In addition, they will have to be able to answer queries posed in terms of the model and return data products that are some predefined binding of the model to the data-exchange language. The total amount of work then scales favourably.

Another useful metaphor for the use of a common model is the following. We can say that the data in each archive is described in a separate language unique to that archive. Instead of every user being required to learn all the different languages, we invent an "Esperanto". The only requirement for the participants of the VO is now that they are able to speak and understand Esperanto. The common domain model plays the role of "Esperanto".

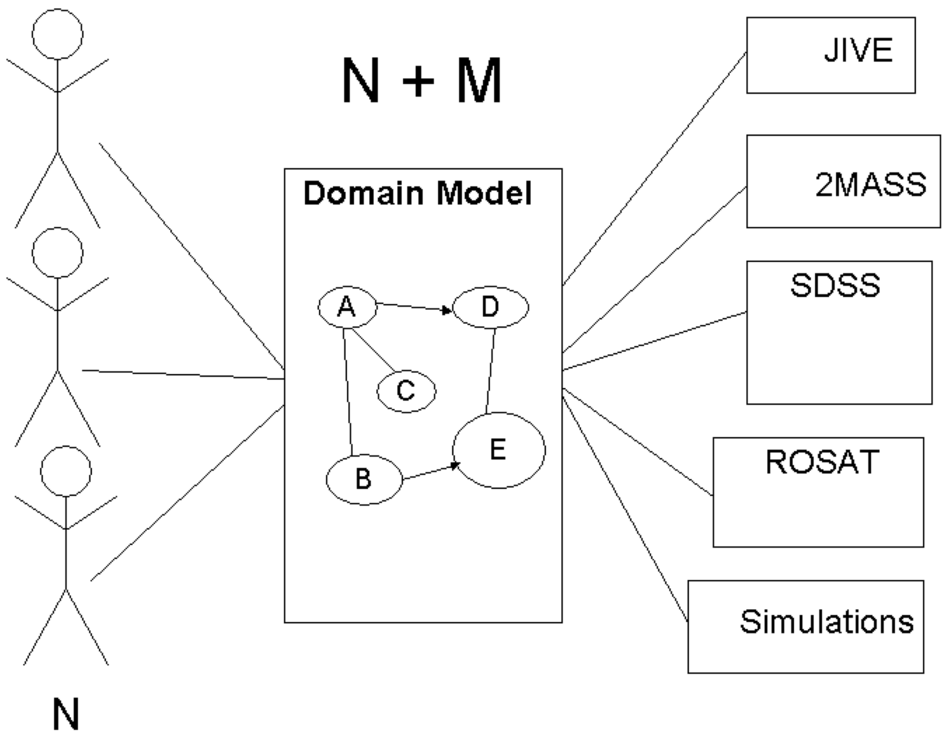

Figure 2: The solution

This common data model is equivalent to an ontology as used in discussions about the semantic web. Note that different solutions to the problem may be proposed. However, in the end *any solution will benefit from a common language*.

6 Do one need to prescribe a common *data storage* format?

It is often argued that it would be a good thing if all simulation archives stored their data in a common data format. This proposal goes beyond the description of the data products in a common language, as described in the previous section. One proposal for such a format is HDF5 (http://hdf.ncsa.uiuc.edu/HDF5/) that is being investigated for this purpose by several simulation groups (Simon White, private communication). The advantages of such a common data format are that users of the VO do not have to worry about reading data containers from different archives, but can write their analysis tools against a common structure.
Note that it may not be required that the archives *store* their data in a prescribed format. It may be that they be required to provide software components (*adapters*) that translate from the storage to the common data format at runtime. In either case this common data format has to be defined. Experience from the FITS process shows that this is probably a long process and that acceptable solutions will have to be found in the meantime.
Ultimately, an answer to this question must come from a poll of participating archives, whether they are willing to conform to a standard format, be it for storage or for translating adapters.

7 What is the theory/observational interface?

One of the main goals of the TVO is to enable *federation* of theoretical with observational archives. That is, we want to enable the comparison of observational with equivalent theoretical data products in a uniform manner (see for example the British Virtual Universe proposal at http://star-www.dur.ac.uk/~csf/virtU/virtU-final.pdf). An example is the comparison of synthetic galaxy catalogues, such as those arising from semi-analytical algorithms applied to dark matter simulations, with observational catalogues from the SDSS. Another example is color-magnitude diagrams for globular clusters, both observed and simulated. Thus, emphasis will be on put the publication of these products and the construction of services (algorithms, visualisation tools, etc.) that enable this kind of comparison. Obviously, this ties in directly with the needs of observers. At the same time, it automates what theoreticians have to do when they want to estimate the validity of their models.

8 What kinds of theory data products can be published?

Standard observational data products that have so far been dealt with explicitly in the VO are images, spectra, and source catalogues. It is planned that also time-ordered event lists and radio visibility data will be supported. Likewise we have to identify some standard theoretical data products. This may help decide on which standard services to define for the TVO (see 10.1). Here we give a list of some products, starting with those that are of particular use to the theory/observational interface.

- Data products that have been proposed for theory/observation interface:
  - synthetic observations of X-Ray clusters (Springel, Tormen) vs. XMM/Chandra observations (Böhringer, Schuecker et al.)
  - color-magnitude diagrams of globular clusters observed (David Zurek et al.) vs. simulated (Zurek et al., Hut et al.)
  - galaxy catalogues from semi-analytical work (Kauffmann et al., Frenk et al) vs. observations (for example SDSS)
  - galaxy merger simulations (Steinmetz) vs. observations (...)
  - Planck CMB simulations with non-trivial topologies (Banday)
  - Ly-α forest simulated (Nusser et al) vs. observed ones (Nusser/Sheth)
- Non-observational products:
  - particle lists
  - halo catalogues
  - halo merger histories

9 How can one model/describe simulations?

The description of results of simulations in a model will include describing the simulations that produced them. In a project like the VO, where the goal is to give as wide a view of astronomical efforts as possible, it is dangerous to limit applications to a few examples of products that should be published. Here, we aim to discover a representative set of *classes* of theoretical models/simulations that people are interested in. It is the task of the IVOA data-modeling group to describe these in a way that extracts the common elements, but also allows for the differences.

There may be multiple ways to classify simulations, many of which are potentially of interest. Here, we present four classifications, namely by simulation subject, by simulated physical processes, by the software algorithms used in these simulations, and by the produced types of data products. A complete description of a simulation may require more than one of these classifications to be represented.

- classified by *subject* of simulations
  - CMB
  - Large-scale structure
    - analysis: gravitational lensing, Lyman alpha cloud spectra, pencil beams, semi-analytical galaxy formation, gravitational clustering, clusters
  - galaxy clusters
  - galaxy formation
  - galaxy mergers
  - globular cluster
  - molecular clouds
  - stellar evolution tracks
  - supernovae
  - accretion disks
  - gravitational waves from merging black holes

    - planetary systems
    - spectra
    - jets
- classified by type of evolution equations:
    - gravity
    - (magneto-)hydrodynamics
    - effective physics (semi-analytical, stellar evolution)
    - “transfer equations” (for example CMBFAST)
    -
- classified by implementation choice
    - particle based
        - PP
        - tree based
    - grid based
        - fixed
        - adaptive
    - mixed
        - PPPM
        - AP3M
        - TreePM
        - AMR
- Classification by kind of data product:
    - particle list
    - grid

10 What kinds of services should be offered?

Apart from providing access to the data products themselves, it will be possible, or sometimes even required to publish services through the TVO. We distinguish three general classes of services. The boundaries between them are not well defined.

10.1 Query services

Query and browsing services are aimed at the discovery of specific data products in an archive based upon specific query parameters. Experience from the VO efforts so far indicates that it will probably be useful to define some simple, standard query services that are easily implemented by simulation archives. Examples from the observational VO are the Simple Cone Search, the Simple Image Access Protocol, and the Simple Spectrum Access Protocol.

As discussed above, these particular query services are hardly ever relevant for theoretical results. Therefore, we may want to define some alternative query services that are relevant. An example might be to return all particles from a cosmological N-body simulation within a given sized volume randomly positioned in space at a given redshift.

Consideration of some standard services should include consideration of the output format. Currently, the standard format for data interchange in the VO is VOTable. We must investigate whether that format is useful/applicable for most kinds of theoretical data structures as well and, if not, what requirements are posed on an alternative.

10.2 Analysis services

An analysis service is defined as a software component that performs a manipulation of data to extract new information. This is also often called *data mining*. The aspect that discriminates analysis services from query services is that the latter deal with relations between different data elements that have been pre-designed into the structure of the data. Data mining services enable discovery of correlations that are *not* explicitly modeled.

The most important standard example from the observational VO is a cross-match service, which identifies common objects in different source catalogues. For reasons described in 13.2, such a service is hardly ever important for theory archives, simply because there is no sense of object identity across different archives.

Examples are:

- Virtual (or synthetic) telescope. Service that "observes" simulation results to produce "images" that can be directly compared to observations. Examples: XMM/Chandra (Tormen), Planck (MPA group). Also (simpler) optical subsets from semi-analytical galaxy catalogues.
- *Comparators* for comparing the results of these synthetic telescopes to the actual observations.
- Statistics calculators such as n-point functions, morphology indicators etc – there is ongoing work at the University of Pittsburgh on this that will first publish an n-point code that will work on a data set the user can define.
- Halo finders.
- Visualisation services

10.3 Simulators

Some groups are investigating the possibility of offering services that allow users to run simulations through the VO. We here extend the definition of "simulation" to include every algorithm that creates *new* data, possibly from data products that have been published in the VO already. The distinction between this and the analysis/data mining services of the previous section is somewhat blurry, though.

Some proposals are:

- N-body codes for galaxy mergers (Steinmetz)
- Semi-analytical galaxy formation algorithms on halo-merger trees (Kauffmann)
- N-body codes linked to stellar evolution codes for globular cluster simulations (Teuben, NVO demo at AAS 2004).

11 How can we make existing services visible in VO?

There are many theory groups that have written applications/services for their data products simply for their own scientific work. These groups may be interested in providing these services to the general community. How will we be able to make these so-called *legacy* services available through the VO?

The VO will likely have to assist these archives by writing appropriate “wrapper” applications. At the University of Pittsburgh, experience has been gained in this respect by wrapping “legacy” n-point correlation function calculators with webservices. This is just an example for implementing a pre-existing code and plugging it on top of some data. In the end, it does not make a difference what data you run code on. Once exposed to the web the whole thing becomes a service right away. Similarly at GAVO a project is in progress to create template adapters for intermediating between a Java based webserver and legacy services implemented in FORTRAN.

12 How do we ensure IVOA compliance?

The TVO should be a part of the overall international VO movement as embodied by the IVOA. As described in great detail in the next section, there are a number of places where current IVOA standards and specifications are not directly applicable to theoretical archives and services. On the other hand, it is clearly desirable to create a working theory/observational interface so that data products from theoretical archives can be treated in a similar manner as observational ones.

We therefore believe that it is important that the TVO be an integral part of the whole IVOA process, but that care is taken that its particular requirements are taken into account.

To this end we propose that a separate IVOA *theory* working group be created within which discussions can be held in a more public manner and which will ensure that the particular requirements are taken into account. We hope that the current document and any feedback it will create can be used to create a charter for this working group.

13 Why does theory need special attention in the IVOA?

This is an important point and we provide answers in more detail.

13.1 Position based query protocols irrelevant for TVO

The IVOA has defined a number of simple query protocols that are easily implemented and provide a simple way to publish many observational data products in the VO. These protocols - such as the Simple Cone Search and the Simple Image Access Protocol - are based on absolute positions on the sky. Results of simulations usually are not tied to a specific absolute position on the sky, which implies that these standards are irrelevant for most theoretical data products[3].

In light of the positive experience of the approach in the IVOA it will likely be very useful to define some simple query protocols for simulations as well and this effort should be one of the first for the TVO to undertake. Clearly though these protocols will need to query for different properties than spatial location.

[3] There is a class of simulations for which absolute positions are relevant, namely those whose initial conditions were designed such that the simulation reproduces observed structure, such as Mathis et al (2002, MNRAS 333, pp. 739-762). Examples like this one are exception rather than the norm, though.

13.2 Federation based on similarity, not identity

Different independent observations of a particular region of the sky can be compared and federated based on an assumed *identity* of objects they contain. Various geometrical cross-match algorithms have been defined that relate objects based on closeness in space, some of them taking into account positional uncertainties. Unless one studies different simulations started from the same initial conditions, such a cross-match and the consequent federation based on identity is irrelevant for theory data products.

This does not mean that different theory products cannot be compared with each other or that observational results cannot be compared directly with theoretical data. It simply requires different kinds of matching "join" specifications.

Finding a "match" in a theoretical archive for a source observed in some catalogue will not be based on closeness in position, but on closeness in physical parameter space. Parameters on which theoretical archives will be searched will consist of *physical properties*, such as masses or sizes of simulated objects, or temperatures or luminosities.

13.3 New observables

In observations results are almost invariably based on detected photons. In theory this generally is not the case. Entities and properties that might not be directly observable are commonly used for theoretical models and are thus included in the results. Usually, mock observations are produced from these. It can be anticipated that creating mock data from theoretical data will be one of the important services for theoretical archives. But this means that a whole new set of properties, namely unobservable theoretical entities, must be modeled.

As another consequence, the IVOA working group on UCDs (http://www.ivoa.net/forum/ucd/) will likely need to be involved to deal with these new theoretical concepts/entities. These concepts have no place in the observational databases that were used to define the UCD dictionary. But they will be required to properly describe data in VOTable **if** that data format is going to be used for transmission of both theoretical and observed data.

13.4 Exact vs observational properties

Theoretical data products contain, by construction, exact knowledge of *all* quantities of interest. To compare such results to observational data, one may need to modify these quantities in a way that mimics observations, thus getting different results.

For example, a color-magnitude diagram of simulated globular clusters in general is different from one derived from a virtual observation, in which stars may be blended etc. In many cases, virtual observations are obtained by artificially observing the theoretical model data through a virtual telescope with exactly the same properties/limitations as the real one. The distinction between ideal data and observed data must be made explicit.

13.5 Models

Physical models play an important role for theoretical work and they have to be included in the data models. Different products often do not have the same physics applied to them. Observations are governed by similar constraints, which eases the comparison. Simulation results can be compared to *understand* effects of different models. Observations *constrain* the possible models. This implies that simulation products can be queried by the kind of physical model that underlies them, something that seems less relevant for observations.

Technical restrictions or methodologies also play an important role in theoretical work. Different algorithms used for the same problem might lead to different ranges of validity of the result. For example, cosmological N-body simulations can be run with a large variety of codes all of which have somewhat different limitations as far as resolution, smallest resolved scales, largest resolved scales (which, curiously enough, is very often neglected), etc. are concerned. These restrictions also have to be included in the data model to allow for comparisons between different theoretical results and between theoretical results and observations.

13.6 New kinds of services pose new requirements on existing models

The technical infrastructure used for examining results of simulations is the same as that used to create them. It is therefore more easily imaginable to offer theoretical services that create new data than in the case of observational data, at least until robotic telescopes come online. One may anticipate that such services pose new and different requirements on the VO infrastructure than those posed by publishing observations. For example, in the data modeling efforts for observations, telescope configurations need be described only from the point of view of describing and querying the resulting data. Online v*irtual telescopes* will create mock images by observing simulation results based on appropriate mock telescope configuration parameters. This configuration will hopefully be described by a model that is common to the corresponding real telescope, but will likely pose new requirements for that model.

13.7 Pre-creating artificial observations from simulations

In the construction of the theory/observational interface we foresee an interesting asymmetry between observations and theory. It seems that comparisons between the two will often involve the extraction of observation-like data products from theoretical ones, whereas the opposite is less likely to be required, or even possilbe, at least on the short term. For example, cosmological dark-matter distributions have to be transformed into artificial weak-lensing maps, simulated galaxy clusters have to be observed with mock X-Ray telescopes for comparison to observed ones. A big issue here is how and where to do this. Some of these mock observations are very time consuming, and to do them on the fly for every query into the observational properties of simulated systems seems unfeasible. It seems therefore that one may want to store sets of *very commonly used derived data* together with the “raw” original data products.

As noted above, it is not anticipated that observations will need to be accompanied by some theory analogue derived from them, although a best fit model of a certain kind might be an interesting exception.

14 What should the IVOA theory working group do?

An IVOA theory working groups should:

- 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.
- Vice versa, incorporate standard approaches defined in those group when designing and implementing services on theoretical archives.
- For example, specify how to publish simulation archives by creating a theory data model within the IVOA data modeling working group.
- Define standard services relevant for theoretical archives
- Define relevant milestones and assign specific tasks to interested parties to implement these.

15 Who are the participants in the TVO?

The list below should be interpreted as a list of *roles* that different participants to the TVO may perform and to which the TVO must therefore cater. Consequently we do not wish to imply that there is no overlap between the different entries. A particular astronomer may play any of these roles at any time.

- *Data providers* who wish to publish their simulations to achieve greater exposure (and consequently more references).
- Theorists that do not have access to large computational facilities and would for their work like to gain access to the results of simulations carried out by other groups that do have those resources available.
- Theorists that want to compare their own simulations with other people's work.
- *Expert service providers* who want to publish their analysis and/or simulation services for greater exposure (and consequently more references).
- Users who want to use services written by experts, so they do not have to write these themselves.
- Users who want to apply their own analysis algorithms to results of large scale simulations performed by others.
- Observers wishing to interpret observations using simulation results.
- Observers wishing to fine-tune future proposals using simulations with virtual telescopes. Various observing proposal websites already offer a simple kind of simulator for predicting exposure time requirements, S/N etc (for example SCISIM for XMM, http://xmm.vilspa.esa.es/scisim/). These use in general simplified models of the systems to be observed. A refinement of such tools based on more detailed simulations could be offered by archives and VOs.
- Laymen such as teachers or amateur astronomers. It is an integral part of many VO proposals that this group of people should be served as well. NVO for example has a special position for public outreach.

16 What are the responsibilities of the various participants?

- Data creators:
  - create the data and store in archive

    - describe process of data creation in standard modelling terms
    - describe data products according to IVOA standards
    - implement automated publication and registration mechanism
- Data providers:
    - enable web access to archives
    - choose data products to be published
    - register data products with IVOA
    - support discovery/query services on data products
    - support federation
- Service providers:
    - implement data discovery/query/analysis/creation services
    - enable web access to results of these services

17 What concrete tasks can we define already?

Here we present a list of tasks that we think will have to be executed.

- archive publication and querying
    - Define a conceptual data model for simulations compatible with IVOA dm.
    - Create a reference implementation for a meta-data repository
    - Design and implement automated registration services for meta-data repository
    - Implement query services on repository
    - Define standard queries/protocols for theory analagous to SCS/SIAP for observations:
        - "volume of space at given cosmological time", possibly subsampled by certain factor
        - properties (constituents) of identified objects as function of time
        - synthetic spectra for specific galaxies/stars/...
        -
- analysis services
    - Create webservice interfaces for existing services such as
        - halo finders
        - statistics calculators (Colberg)
    - Create tools for creating mock observational products from simulations ("virtual telescopes")
        - Virtual Chandra for X-Ray clusters (Gardini et al)
        - Mock SDSS from semi-analytical galaxy catalogues
        - Weak lensing on LSS
        - Strong lensing, SZ, .. on (X-Ray) clusters
        - Globular cluster observations
        - Mock spectra
        - Mock CMB maps (CMBFAST online, SZ, integrated Sachs-Wolfe)
        - Mock Ly-α spectra
    - Create tools for comparing observations with mock-observations theoretical products, for example
        - X-Ray clusters

        - Galaxy catalogues
        - Galaxy mergers
        - Globular cluster CMDs
    - Expose visualisation services
    -
- simulators
    - initial conditions generator
    - galaxy merger simulator (including genetic algorithms for initial conditions?) (GAVO)
    - all these have to be wrapped by a web interface
    - semi-analytical galaxy formation wrapped by web interface

18 What reference implementations can we define ?

- Semi-analytical galaxy formation
- X-Ray clusters simulation vs observation
- Galaxy merger simulation reproducing observation
- Ly-α forest simulations
- Advantages of storing theory data in a standard database: example, store simulated galaxy catalogues in SDSS, in same format.
-

19 What efforts are already in progress or planned?

Here is a list of VO–like efforts that people are involved in already or that are planned for the near future. **Please add your own efforts to this list.**

- Theory NVO demo (Teuben et al)
- Simulation data modeling (Lemson)
- Ly-α forest simulations (Nusser)
- Power-law cosmological simulations and merger-tree/halo structure models (Colberg, Sheth)
- CMD archiving (Zurek et al)
- MPA/Virgo simulations publishing (Springel, White, GAVO)
- MPA Planck simulation publishing (MPA, GAVO)
- MPE+MPA+Research Network proposal on X-Ray clusters, simulated and observed (Böhringer et al)
- British "Virtual Universe" proposal (Frenk, Lahav, Walton)

20 References

Here is a list of useful references and links:

1. IVOA http://www.ivoa.net and links from there to member VO projects.
2. MPA simulation download site: http://www.mpa-garching.mpg.de/galform/virgo/index.shtml
3. MODEST site http://www.manybody.org/modest/
4. GAVO proposal (http://www.g-vo.org/...)
5. Virtual Universe PPARC proposal http://star-www.dur.ac.uk/~csf/virtU/virtU-final.pdf

6. NVO publications (http://www.us-vo.org/publications.html)
7. Peter Teuben's "minutes" of meeting at ADASS below
8. Gardinai et al astro-ph/0310844 (an example of a virtual or synthetic telescope).
9. HDF5 data specification, http://hdf.ncsa.uiuc.edu/HDF5/

Acknowledgements

This document finds it origins in a discussion the authors had with Dave De young, Peter Teuben and Tony Banday at ADASS XIII in Strassbourg. We have received useful comments on earlier versions of this document by Tony Banday, Geraint Lewis, Peter Schuecker and Wolfgang Voges.