The Virtual Solar Observatory

Design Proposal

2002 November

The VSO Study group

Stanford University	Richard. S. Bogart Karen Tian
National Solar Observatory	Frank Hill Stephen Wampler
Montana State University	Piet Martens Alisdair Davey
Solar Data Analysis Center	Joseph B. Gurman George Dimitoglou

Executive Summary

Space-based solar physics missions such as *Yohkoh*, *SOHO*, and TRACE have acquired or continue to amass large datasets, and the next five years will see further growth as Solar-B, STEREO, and SDO data become available. Many of these datasets are mirrored, that is, served from more than one location on the Internet, at least partly for reasons of bandwidth. Similarly, groundbased helioseismoligy networks such as GONG+ are acquiring data at higher rates than ever before, and the SOLIS synoptic program promises to add Pbytes (1 Petabyte = 10^3 Terabyte = 10^6 Gigabyte) more. (By comparison, starting in 2007, SDO should produce ~ 1 Tbyte of raw data a *day*.)

At the same time that the volume and multiplicity of sources is growing, the technology to serve huge data volumes online, with negligible labor costs after startup, has become commonplace. The move toward distributed data service, the ease of setting up such services, and the longstanding recognition that the best place for data service is where the scientific insight into the data also reside, have led many in the solar physics community to believe that centralized data centers can be replaced with a system of unified, networked access to distributed resources, that would better serve the needs of research solar physicists, while requiring fewer resources that could otherwise be directed toward research.

The NASA Sun-Earth Connections 2001 Senior Review of operating missions and data centers funded the Solar Data Analysis Center (SDAC) to construct and deploy a prototype Virtual Solar Observatory (VSO) to test the distributed access concept. After a six-month design study carried out by a group composed of solar physicists and data access experts at two universities, the National Solar Observatory, and the SDAC, and involving the community as broadly as possible, we are ready to recommend a VSO prototype architecture with the following, principal features:

- the ability to search all participating data sources with a common interface,

- the use of existing data query facilities at the current data sources,

- the use of industry-standard protocols such as eXtensible Markup Language (XML) in its internals,

- access through a Web browser interface to a remote VSO server, an Application Programming Interface (API), or *via* a VSO instance on a local computer, and

- extensibility to multiple additional features, including the registration of searches to enable other solar physicists to, for instance, attempt to reproduce the results of published literature.

In accordance with the funding profile recommended by the 2001 SEC Senior Review, we propose, after a community comment period extending into 2003 January, to proceed with the development and deployment of a VSO prototype involving data served at four institutions (Stanford, Montana State, NSO, and the SDAC) in Fiscal Year 2003, and to improve and extend the VSO in FY 2004. That extension will take the form of assistance to smaller data providers in the form of programming advice, assistance in procuring network-attached storage, and aid in constructing inexpensive databases, in order to allow inclusion of those sites in the VSO.

I. Introduction: Why Do We Need a VSO?

The number of online sources of large amounts of solar data is increasing steadily, and the volumes of most of the new sources due to become available over the next five years will be significantly larger than the size of those currently available. By 2007, we will have to confront Solar Dynamics Observatory experiment archives that will grow by over a Tbyte a day when decompressed. Not only is it becoming harder to determine the Web locations of all the servers and understand the different query building tools, but there is no longer a compelling reason to concentrate all the data for even a single, multiple-instrument mission in a single physical location.

At the same time, the technologies for serving such quantities of data online are becoming inexpensive and simple to deploy (network attached storage, open source SQL databases, &c.), and no unique information technology (IT) expertise is required to implement such devices or methods. The TRACE and, more recently, RHESSI experience have shown that implementing rapid, online access to the entire mission scientific data set, virtually from the day of launch, is well within the reach of PI team members. Thus, both the scale of the data access soon to face the solar physics community and the technology argue in favor of a virtual access point to physically distributed data services. The ability of such a “Virtual Solar Observatory” (VSO) to simplify the diverse query schemes of different service sites must be seen as another strong argument for implementing a VSO.

The 2001 Senior Review of NASA Sun-Earth Connections (SEC) Operating Missions and Data Centers approved additional funding for the Solar Data Analysis Center (SDAC), in essence, to put itself out of business by building and implementing a VSO.

II. The Strawman VSO Concept

We propose here an architecture and feature set for a *prototype* VSO. the prototype will not include all the features that might eventually become part of the VSO, nor even all the features that are concurrently under development elsewhere that should become parts of the VSO's functionality. It is possible to approach the design of such a system in at least two different ways. In one (top-down), all possible features and uses of a system are studied, and the best solution for as many as possible is proposed. This is the approach taken by the European Grid of Solar Observations (EGSO); see section V. Alternately, one can approach a system design from the bottom up, and ask what the essential element or elements of the design have to be in order to have a functioning and useful system.

The VSO study group decided, after examining different approaches to abstracting the procedures for solar data identification and access, to build the "smallest box" possible around that problem, rather than attempting to draw a box around all possible aspects of a VSO. Despite the difference in approaches, it currently appears likely that the EGSO design will converge on an approach with many components in common with our "small box."

It is also critical to note that the VSO will not be "complete" during its prototype stage, and even if adopted by the solar physics community, will only become truly useful if there are adequately funded, peer-reviewed opportunities to add functionality, *e.g.* in the form of user interfaces, remote processing to reduce data transfer, event and feature lists, methods of connecting the VSO with other SEC-related data systems, and so on.

Features of the prototype

We propose to proceed with the development of a prototype VSO in FY2003, and in FY2004 add both more data services and more functionality. At each step in the development of the VSO, we will insure the solar physics community a voice in the structure and direction of the VSO.

A protoype VSO will offer:

- virtualization of data search, data discovery, and query refinement,
- multiple interfaces (browsers, application programming interface),

- leveraging of existing data services (rather than creating *e.g.* new metadata standards),

- direct user access to data (without the VSO as an intermediary), and

- the ability to expand to several more data sources in the implementation and maintenance phases of the VSO (FY04 and beyond).

A solar physicist should be able to use the VSO *via* a browser interface to search for data applicable to a problem he or she wishes to study; refine the search criteria interactively through the VSO; and then retrieve the data directly from the data source (rather than generating twice as much network traffic by having the VSO act as an intermediary for the data as well as for the query). Alternately, one could use a an application programming interface (API) to search for (for example) the most recent image of a given type, regardless of observatory or instrument, to update one's own Web site, without a user interface.

The prototype VSO would *start* with data served by the NSO (including GONG, GONG+, and SOLIS), the SDAC (including a variety of space mission and ancillary data), Stanford University (including Wilcox Solar Observatory, *SOHO* MDI, and other helioseismic data sources), and Montana State University (including an extensive *Yohkoh* database). (The datasets at these data services are listed in Appendix C.)

The prototype VSO proposed here will *not* include:

- a central catalog,

- grid computing, or

- any features that limit or restrict access to data or software (authentication).

The study group found that the sites most likely to be involved in the initial stage of a prototype (*i.e.*, those at their own institutions) already offered sophisticated query engines; there was no need to design a new one.
Features that we hope could be added soon after the prototype is operational include:

- catalog caching (to speed queries),

- multiple instances of the VSO (to make the VSO truly virtual, it could run on local machines instead of on a small number of remote servers), and

- logging of searches (without identifying the party who carried out the original search).

We believe that multiple instantiation is a feature that many users would desire; the only “cost” associated with it should be the requirement to transmit logging information to one of the “primary” servers, so the funding agency can know how extensively the VSO is being employed, and so the searches could be logged.

Logged searches, identified only by date and time, could be cited in the acknowledgments of articles published using the data, and would thereby allow reproduction of results, a sound practice currently largely in disuse in solar physics. Logged searches would also enable interested researchers to try different analysis methods on the same data sets.

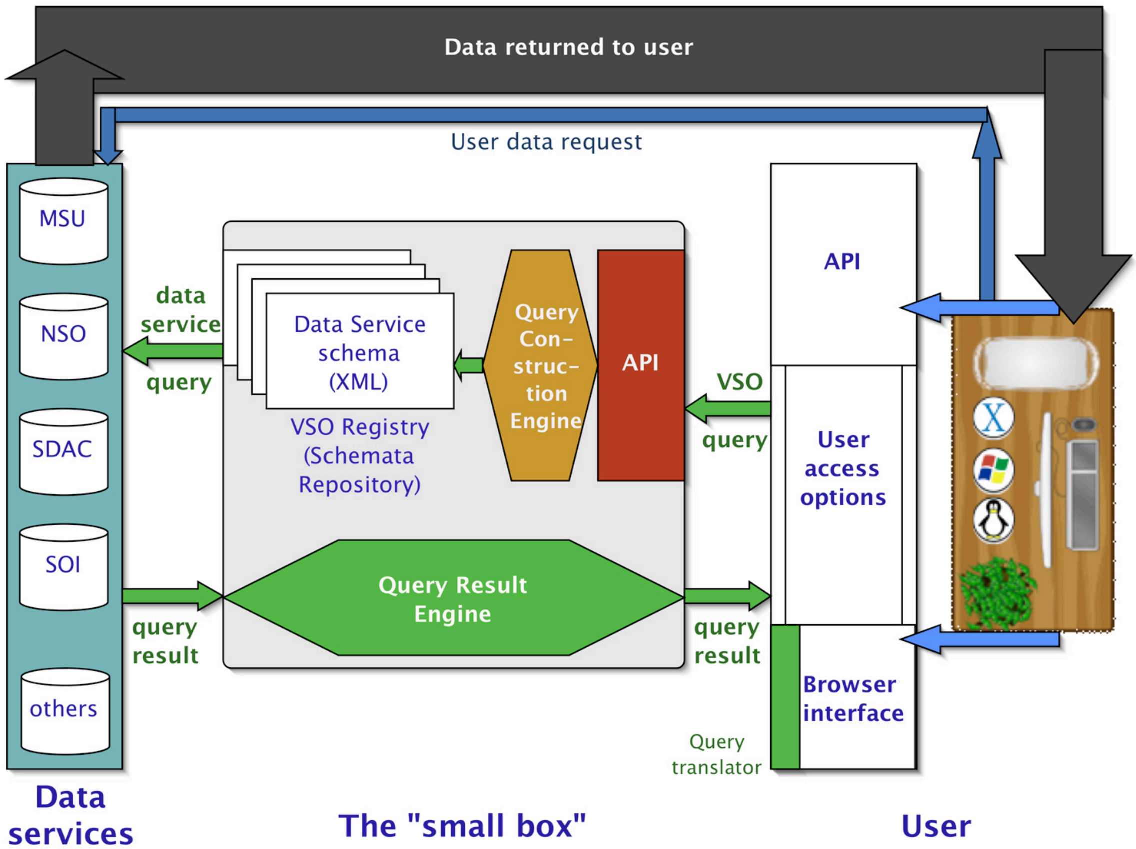

Figure 1. Conceptual diagram of the proposed *prototype* VSO. As in all virtual observatory concepts, a broker facility mediates between the data user (right) and the data services (left). In this design, however, only queries and query results pass through the broker; the actual data transfer occurs outside the VSO, thus significantly reducing network data traffic.

User-initiated actions (initial query of the VSO, eventual data request) are in blue; VSO communications are in green; and the final data transfer is in charcoal. The user initiates a query *via* either a browser interface, which communicates with the VSO API or directly through the VSO API. The query is then routed to a *query construction engine* which uses the XML schemata describing the data services to determine the service(s) to which to route queries in formats native to them. The query results from the data services are routed by the *query result engine* back to the API or browser; the user or the user's software then decides whether to request the data from the data service(s). (The user's desktop is drawn in such a way as to be platform-agnostic.) While the prototype will probably include only the four named data services, more could be added at any time thereafter.

VSO expansion

The VSO will be fully useful only if it can be expanded beyond the prototype stage, both in the number and variety of data services accessible *via* the VSO and in the functionality offered. The "maintenance" level of funding for the VSO proposed in the SDAC's "out-year" (FY2004 and FY2005) budget in the 2001 Senior Review included funds for roughly one programmer full-time equivalent (FTE). At least 50% of that effort would be available to offer advice (*e.g.* on the transfer of tapes-on-a-wall archives to network-accessible media), the deployment of inexpensive network-attached storage (NAS) hardware, the development of MySQL templates for constructing databases, and a limited amount of travel to new data source sites for short periods to help with such VSO deployment.

Extension of the capabilities of the VSO must, therefore, depend on research opportunities to provide, for instance:

- feature and event lists,
- remote processing capabilities to reduce transferred data volume (*e.g.* automated selection based on the presence or absence of features or events in the data), or even
- scheduling of observations when no data can be found in existing data services, if automatically queued observations are available.

Such expansion of functionality must be funded separately from VSO development and maintenance, since it must involve the user community and is properly a research effort. If the community believes that such an effort is worthwhile, it should make its needs known to the funding agencies to that new or existing programs (*e.g.* NASA LWS TR&T, NSF data infrastructure); preliminary indications are that the funding agencies are receptive to focusing some of their research opportunities on such activities.

For more information

See the VSO Website: `http://virtualsolar.org/` .

III. Technical Approach

Motivation

The data providers and sources are geographically distributed across the country. Almost all of the providers make their data available on the World-Wide Web. The access facilities vary from sophisticated, Web-enabled databases to simple file systems available via FTP. The corresponding metadata descriptions (catalogs) maintain a similar range of sophistication ranging from on-line searchable databases to text (*i.e.* ASCII, HTML) files.

A requirement for a successful VSO is the provision of a transparent view of and access to all these datasets.

Combining all of the datasets, or their catalogs, in a central repository may appear to be a valid approach, but it is unrealistic. Such an architecture simply would not scale to accommodate the ever-growing volume and complexity of data (particularly from space-based missions and ground-based helioseismology networks), even if synchronizing and updating the metadata proved practicable. At the very least, it would require an additional layer of data expertise for each data service, something a small research community can ill afford. Since each data provider already has a network-based access method, we propose to leverage those existing (and presumably evolving) investments in query capability to construct a VSO that presents the user with a uniform interface for a broad selection of data services.

We introduce the notion of VSO as a processing engine rather than a "database of databases." The design presented here is lightweight, distributed, and maintains only descriptions of the data holdings and access methods for each data provider.

The Prototype

Overview

We view the function of the VSO as a metadata intermediary between end-users and data providers.

The VSO will consist of three major components:

- Query Construction Engine
- VSO Registry
- Query Results Engine

We also discuss briefly the front (user interface) and back ends; throughout, refer to Figure 1.

Front end: How will I be able to use the VSO?

Users will be able to access the VSO either by using a pre-defined, HTML-based Web form or by developing a program that communicates with the VSO directly.

The VSO will feature a standard Application Program Interface (API) to permit any program to build and submit queries to the Query Construction engine. As a result, the VSO provides the flexibility to users to develop their own custom User Interfaces (UI's).

Query Construction Engine

The Query Construction Engine is a program that receives input from a user interface and constructs a valid query using the information stored within the Schemata Repository.

The function of the Query Construction Engine is to translate the generic search descriptions (from user input) to data service-specific query parameters. This involves locating the candidate data service(s) and refining the search. As each data provider has described its data holding to the VSO in the registry, (*e.g.*, observable and time range information), the VSO will select one or more candidate data services whose descriptions match the generic search criteria. Once the candidate data services are selected, data service-specific questions can be formulated to aid the user in refining the search.

VSO Registry

The VSO Registry (see Figure 1) maintains a collection of schemata and their instances. Each schema describes the organization of the data at a provider's site while the instance of the schema provides the descriptions of the contents and access methods (data volumes, location) of the data provider. The XML Schemata are not copies of the catalog contents of data providers but provide templates for an XML description of their datasets.

Back end: Query Execution

Query execution and data export are the two types of services that data services are expected to provide to the VSO. We don't expect participating data services to do anything significantly different from what they are already offering, *i.e.*, allowing searches for a dataset and exporting it if requested. To participate in the VSO, each participating data service needs only to provide two kinds of interfaces: one to accept query parameters and carry out queries, and another to handle data export requests.

Query Result Engine

The function of query result is to merge the query results from multiple data services. The engine will, in so far as possible, hide the idiosyncrasies of each data service's exporting methods, and provide an integrated view to the user. Information such as the size of the dataset are likely to be of interest to the user, and will be presented at this point.

Technology

Why XML?

XML (eXtensible Markup Language) provides a mechanism to identify structures in a document. It is text-based and platform-independent. In addition to assigning meanings to contents via user-defined XML tags (reminiscent of the "keyword=value" approach), XML can identify the structural relationship among tags (e.g., the hierarchical structure of a document), therefore creating a structured document. It is this feature of "structured content" that make XML attractive to us. Because of its popularity, we can leverage many existing technologies to process XML documents.

We will employ the capabilities of XML for both describing the data services and for a “Web services” approach to integrating the data service description, query, and response functions of the VSO.

How we will use XML

i. Data Service Description

To facilitate the integration of distributed data services (we distinguish between “data service” and “data provider,” since each provider might in fact host more than one data service, with a unique access method and organization), each data service must first be able to describe its data holdings to the VSO. We choose XML as a “universal language” with which each data service “communicates” with the VSO. The VSO registry collects these descriptions, and serves as a resource discovery mechanism for the Query Construction Engine.

Since there's no predefined meanings for XML tags, there isn't any preconceived semantics. We ourselves have to define the semantics of an XML document. The burden of the design therefore falls on the abstract meta-data model that characterizes the data services of participating data providers. This meta-data model is implemented as an XML Schema which in turn specifies the structure and datatype of XML documents. The challenge is to create a meta-data model that is extensible, so that it can support dynamically changing requirements, as data providers and services are added and evolve.

For example, in our dataset model, two basic elements are "observable" and "time range". The possible values of observables can be enforced using the following enumeration structure in an XML schema:

```
<xs:simpleType name="ObservableName">
  <xs:restriction base="xs:string">
    <xs:enumeration value="Dopplergram"/>
    <xs:enumeration value="Magnetogram"/>
    ...
  </xs:restriction>
</xs:simpleType>
```

The specification of time range is more interesting because there is more than one way to do it, *e.g.* defining the start and end of the time range is one way, and the center and extent another. For non-continuous observations, sampling rates are also needed. This is a typical example of inheritance in the object-oriented paradigm, and can be implemented using the abstract/substitutionGroup definition in XML schema. Spatial location is another element with comparable complexity and common inheritance.

ii. Service Integration

WS (Web Service) augments the service methods of the Web, namely POST, GET, and PUT. It builds upon the platform of XML and HTTP, and includes elements such as remote invocation (SOAP), service characteristics (WSDL), and directory service (UDDI).

SOAP (Simple Object Access Protocol) defines an RPC (Remote Procedure Call) mechanism. It employs HTTP as its transport and encodes the client-server interactions (*i.e.*, request and response messages) in XML documents.

WSDL (Web Services Description Language, as its name suggests, describes what a Web service can do, where it resides, and how to invoke it. According to this information, one should be able to construct an API to access the Web service.

We are interested in the Web Service model because it provides an alternative to the standard POST and GET methods.

New ideas

The architecture of the VSO offers some new computing ideas.

First, it is not a monolithic server with a database that stores metadata. Instead, it is an agile, lightweight application program that accepts input (user queries) and provides output (data location and access). We store XML descriptions of data providers as a data structure (it can reside in memory) instead of static files or records.

Second, we treat each user input and query not only as a transaction but as a real-time computation. By doing so, we are able view VSO's inner-parts as instantiated objects and perform operations or manipulate them by applying methods.

Finally, the deployment of VSO may be fully distributed with the ability to have multiple, fully customizable VSO's running on user desktops.

What a user would need to access the VSO

For the prototype version we will develop at least one browser interface, an HTML-based form that would allow a user to enter search criteria. The form will communicate via the VSO API, construct a query, send it to the appropriate data providers, and finally display the results. The user could refine queries iteratively to produce a final result, which would allow the direct download of data (or the queuing of larger requests) without the final request or fulfillment passing through the VSO itself.

This type of access will probably be the most typical kind of access. However VSO is not bound to only one User Interface (UI). Users are able to create their very own interface (web-based, command based) as long as they communicate properly with the VSO API. The VSO API will be generic enough and accessible via a least common denominator protocol (TCP/IP) allowing users to develop interfaces with any programming language that supports network connectivity (*e.g.* C, C++, Java, Perl, Python, the IDL SolarSoft tree, and IRAF). This would allow users to integrate an interface to the VSO into local data analysis or Website population software.

IV. Programmatics

Funding for the first six months of the VSO effort became available in mid-fiscal year (FY) 2002. Unsolicited proposals to perform the study together were received from three groups (see below), and this approach was approved by the VSO steering committee (see appendix A). The three study grantee groups and the SDAC became, informally, the VSO Study Group (See Table 1.) Work began in early May, 2002 and continued with fortnightly telecons, one group meeting (in October, 2002), and various offline contacts among study group members.

In addition to their design work, members of the VSO Study Group presented their then-current ideas of what a VSO might look like at several community and committee meetings (see Appendix B).

SDAC	Joe Gurman George Dimitoglou
Stanford University	Rick Bogart Karen Tian
National Solar Observatory	Frank Hill Steve Wampler
Montana State University	Piet Martens Alisdair Davey

Table 1. The VSO Study Group

The work of the Study Group is now complete, with the production and dissemination of this report. Study Group members will participate in the VSO Birds of a Feather (BoF) session at the Fall AGU meeting; they have already discussed VSO goals and designs to as much of the solar physics and related communities as possible (see Appendix B).

Management

The SDAC Facility Scientist (Joe Gurman) is the *de facto* project manager/project scientist for the VSO. Charles Holmes at NASA Headquarters is the *de facto* program manager, in his role as Mission Operations and Data Analysis (MO&DA) manager for solar and heliospheric physics within the Office of Space Science.

Future effort

If the community approves the proposed direction of the VSO effort, we would proceed to:

1. develop and implement the prototype VSO beginning in FY2003 (since NASA contracting is a somewhat sluggish process, we expect the effort to begin no earlier than the mid-second quarter of calendar year 2003);

2. add features to VSO in FY2004, including both in-house (*e.g.* multiple instantiation, catalog caching, and/or search logging) and "research opportunity" features; and

3. maintain and expand the VSO in FY2004 and beyond, depending on funding.

Budget

The current budget of the VSO is shown in Table Budget-1.

US Government Fiscal Year (FY)	**US$K**
2002	160
2003	450
2004	453
2005	160
Total (approved by 2002 Senior Review)	1223

Table Budget-1.

Community input

It is clearly in the interest of the solar physics community, as well as members of related discipline communities in the Sun-Earth Connections area, to express their opinions about the design and direction of the VSO. They may do so, at any time, to the members of the VSO study group (see above) and to the members of the VSO steering committee, whose current membership is detailed in Appendix A. In the absence of wider community involvement, the VSO steering committee must act in the community's interest to oversee the direction of the VSO.

V. Relation to other efforts

EGSO

The VSO is not the only current effort attempting to construct a virtual solar observatory. The European Grid of Solar Observations (EGSO), an effort funded by the European Commission, is taking a more highly structured approach, with formal Work Packages and (currently at least) a more ambitious set of design goals. Two of the VSO study group member institutions (NSO and the SDAC) are unfunded US partners in the EGSO effort.

The design of the EGSO is still under discussion, but we can foresee EGSO-VSO design issues falling into either of two scenarios:

1. similar designs

2. disparate designs

In the similar design case, interfacing the VSO and EGSO should be straightforward, and involve little cost or effort to the VSO. In the disparate design case (one possibility under study, for instance, would feature a central catalog of data), the US EGSO partners would supply their part of the work required (from resources other than VSO funds), and the international solar physics community would have the benefit of two competing architectures between which to choose. Since we believe that competition is at least as good in the marketplace of ideas as it is in the commercial one, we welcome the latter scenario, since we lay no claim to precognition. We can live with the former scenario, too — as long as the solar physics community remains involved in critiquing the VSO design and implementation.

See: `http://www.egso.org/` for more information.

LWS Data Enviroment

There is, of course, a broader effort under way to provide a data environment for NASA's Living With a Star program (and hopefully the International LWS program, iLWS, as well). Ideally, the VSO should "plug and play" with such a system. If the VSO proves useful to the solar physics community, we would foresee the integration of the VSO into a an LWS data system in a way that maintains its usefulness to the base community, but broadens its functionality in such a way that non-solar physicists would find it of use as well. We believe that such efforts must follow the successful development and development of the VSO.

National Virtual Observatory

Much better known and funded than the VSO, the NVO (and its superset, the International Virtual Observatory Association (IVOA)), seeks to offer virtualized access to a large range of astronomical databases, data mining, and other features of interest to "nighttime" astronomers. There have been frequent questions and feelers as to the relationship of the VSO to these efforts; the response of the study group has so far been that while we are interested in any technology developed by the NVO efforts, the problem we are addressing is sufficiently different that we would rather have at least a working prototype before entering into any formal discussions with the "dark side." We are simply too small (~ 60 times smaller budget) and too different an effort to risk being subsumed into something as different as the NVO effort. The difference in funding sources (NSF for the NVO, NASA for the VSO) is another distinction that leads us to believe that for the short term, the VSO and NVO should remain separate entities, while we share technology and insights.

See: `http://us-vo.org/` for more information on the NVO.

Appendices

A. The VSO Steering Committee

Member	Affiliation
Robert Bentley (chair)	Mullard Space Science Laboratory University College London
Samuel Freeland	Lockheed Martin Corporation Advanced Technology Center
J. Todd Hoeksema	NASA Headquarters, Code S and Stanford University
Stephen R. Walton	California State University, Northridge Department of Physics and Astronomy
Dominic Zarro	L-3 Communications Analytics Corp.

B. Community exposure

VSO-related presentations have been given at each of the following:

- American Astronomical Society meeting (Washington DC), 2002 January (invited talk)
- American Astronomical Society Solar Physics Division meeting (Albuquerue NM), 2002 June (poster; BoF session)
- European Grid of Solar Observations (EGSO) meeting, 2002 October (presentation)
- NASA Sun-Earth Connections Data and Computation Working Group (SECDCWG, Washington DC), 2002 October (presentation)
- NASA Living With a Star (LWS) workshop (Scaggsville MD), 2002 November (poster)
- American Geophysical Union Fall meeting (San Francisco CA), 2002 December (BoF)

C. A sample data service schema

The example below is a preliminary and highly imperfect description of the holdings of the National Solar Observatory's Digital Library. It is not meant to represent precisely the eventual NSO data service schema for the VSO, but instead, what a generic schema of the type for the proposed VSO prototype might look like.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<nso:schema xmlns:nso="http://vso.nso.edu/XMLSchema"
targetNamespace="http://vso.nso.edu"
xmlns="http://vso.nso.edu"
elementFormDefault="qualified">

<nso:annotation>
  <nso:appInfo>NSO-VSO Data Description</nso:appInfo>
  <nso:documentation xml:lang="en">
  This Schema defines the National Solar Observatory data holdings for the
Virtual Solar Observatory.
  </nso:documentation>
</nso:annotation>

<nso:element name="facility">
  <nso:complexType>
    <nso:choice>
      <nso:element name="observatory" type="nso:token"/>
      <nso:element name="telescope" type="nso:token"/>
      <nso:element name="instrument" type="nso:token"/>
    </nso:choice>
  </nso:complexType>
</nso:element>

<nso:element name="data_type">
  <nso:simpleType>
    <nso:restriction base="nso:string">
      <nso:enumeration value="LOS_Magnetogram"/>
      <nso:enumeration value="Vector_Magnetogram"/>
      <nso:enumeration value="Spectrum"/>
      <nso:enumeration value="Intensity_Image"/>
      <nso:enumeration value="Doppler_Image"/>
      <nso:enumeration value="Time_Series"/>
      <nso:enumeration value="Parameter_Table"/>
    </nso:restriction>
  </nso:simpleType>
</nso:element>

<nso:element name="ut_obs_start">
```

```
  <nso:complexType>
    <nso:all>
      <nso:element name="ut_date_start" type="nso:date"/>
      <nso:element name="ut_time_start" type="nso:time"/>
    </nso:all>
  </nso:complexType>
</nso:element>

<nso:element name="ut_obs_end">
  <nso:complexType>
    <nso:all>
      <nso:element name="ut_date_end" type="nso:date"/>
      <nso:element name="ut_time_end" type="nso:time"/>
    </nso:all>
  </nso:complexType>
</nso:element>

<nso:element name="time_cadence">
  <nso:complexType>
    <nso:all>
      <nso:element name="time_step" type="nso:decimal"/>
      <nso:element name="time_step_units" type="nso:token"/>
    </nso:all>
  </nso:complexType>
</nso:element>

<nso:element name="wavelength">
  <nso:complexType>
    <nso:all>
      <nso:element name="wl_start" type="nso:decimal"/>
      <nso:element name="wl_end" type="nso:decimal"/>
      <nso:element name="wl_step" type="nso:decimal"/>
      <nso:element name="wl_units" type="nso:token"/>
    </nso:all>
  </nso:complexType>
</nso:element>

<nso:element name="spatial_type">
  <nso:simpleType>
    <nso:restriction base="nso:string">
      <nso:enumeration value="full_disk"/>
      <nso:enumeration value="corona"/>
      <nso:enumeration value="local_area"/>
    </nso:restriction>
  </nso:simpleType>
</nso:element>

<nso:element name="heliographic_coordinates">
  <nso:complexType>
    <nso:all>
```

```
      <nso:element name="longitude_start" type="nso:decimal"/>
      <nso:element name="longitude_end" type="nso:decimal"/>
      <nso:element name="longitude_step" type="nso:decimal"/>
      <nso:element name="longitude_units" type="nso:token"/>
      <nso:element name="latitude_start" type="nso:decimal"/>
      <nso:element name="latitude_end" type="nso:decimal"/>
      <nso:element name="latitude_step" type="nso:decimal"/>
      <nso:element name="latitude_units" type="nso:token"/>
    </nso:all>
  </nso:complexType>
</nso:element>

<nso:element name="cartesian_disk_coordinates">
  <nso:complexType>
    <nso:all>
      <nso:element name="x_start" type="nso:decimal"/>
      <nso:element name="x_end" type="nso:decimal"/>
      <nso:element name="x_step" type="nso:decimal"/>
      <nso:element name="x_units" type="nso:token"/>
      <nso:element name="y_start" type="nso:decimal"/>
      <nso:element name="y_end" type="nso:decimal"/>
      <nso:element name="y_step" type="nso:decimal"/>
      <nso:element name="y_units" type="nso:token"/>
    </nso:all>
  </nso:complexType>
</nso:element>

<nso:element name="polar_disk_coordinates">
  <nso:complexType>
    <nso:all>
      <nso:element name="radius_vector" type="nso:decimal"/>
      <nso:element name="position_angle" type="nso:decimal"/>
    </nso:all>
  </nso:complexType>
</nso:element>

<nso:element name="spherical_harmonic">
  <nso:complexType>
    <nso:choice>
      <nso:element name="degree_l_start" type="nso:nonNegativeInteger"/>
      <nso:element name="azimuthal_degree_m_start" type="nso:integer"/>
      <nso:element name="radial_order_n_start" type="nso:integer"/>
      <nso:element name="degree_l_end" type="nso:nonNegativeInteger"/>
      <nso:element name="azimuthal_degree_m_end" type="nso:integer"/>
      <nso:element name="radial_order_n_end" type="nso:integer"/>
      <nso:element name="degree_l_step" type="nso:positiveInteger"/>
      <nso:element name="azimuthal_degree_m_step"
type="nso:positiveInteger"/>
      <nso:element name="radial_order_n_step" type="nso:positiveInteger"/>
    </nso:choice>
```

```
  </nso:complexType>
</nso:element>


</nso:schema>
```

D. Data services for the initial implementation of the VSO prototype

Data service	**Data served**
National Solar Observatory (NSO)	spectral atlases magnetograms (photospheric and chromospheric) Dopplergrams synoptic maps (magnetic field and other parameters) coronal emission-line scans Synoptic Optical Long-term Investigation of the Sun (SOLIS; from 2004) Global Oscillations Network Group (GONG) and GONG+
Solar Data Analysis Center (SDAC)	*SOHO* (except MDI high-rate) TRACE *Yohkoh* *CGRO*BATSE solar flare data GOES soft X-ray photometry *SMM* *OSO-7* raster images *Solar-B* (from 2005) STEREO (from 2005)
Stanford University solar group	*SOHO* MDI (including high-rate) Wilcox Solar Observatory magnetograms GONG subsets TON helioseismology database Mt. Wilson 60-ft. tower data LOWL helioseismology database
Montana State University solar group	*Yohkoh* database

E. XML References

1. http://www.w3.org/XML/
2. http://www.w3.org/TR/xmlschema-1/
3. http://www.w3.org/TR/NOTE-xml-schema-req
4. http://www.xml.com/pub/a/98/10/guide0.html
5. http://www.brics.dk/~amoeller/XML/overview.html
6. http://www.xfront.com/BestPracticesHomepage.html
7. http://www.w3.org/TR/wsdl
8. http://www.uddi.org/about.html