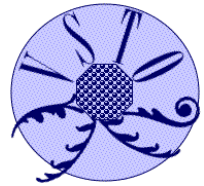
The Virtual Observatory *Exposed*



Peter Fox*

*HAO/ESSL/NCAR

Thanks to Deborah McGuinness[$#], Luca Cinquini[%], Patrick West[*], Jose Garcia[*], Tony Darnell[*], James Benedict[$], Don Middleton[%], Stan Solomon[*], eGY and others.

[$]McGuinness Associates

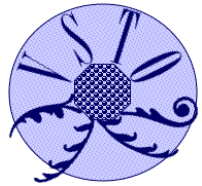
[#]Knowledge Systems and AI Lab, Stanford Univ.

[%]SCD/CISL/NCAR

Outline

- Terminology and general introduction
- Where is the need coming from?
- What should a VO do?
- Inside VOs (in Geosciences)
- Final remarks

Terminology



- Workshop: A Virtual Observatory (VO) is a suite of software applications on a set of computers that allows users to uniformly find, access, and use resources (data, software, document, and image products and services using these) from a collection of distributed product repositories and service providers. A VO is a service that unites services and/or multiple repositories.
- VxOs - x is one discipline, domain, community, country
- NB: VO also refers to Virtual Organization

*e*GY definition

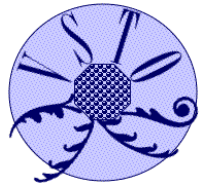
- The purpose of a Virtual Observatory is to increase efficiency, and enable new science by greatly enhancing access to data, services, and computing resources.
- A Virtual Observatory is a suite of software applications on a set of computers that allows users to uniformly find, access, and use resources (data, documents, software, processing capability, image products, and services) from distributed product repositories and service providers.
- A Virtual Observatory may have a single subject (for example, the Virtual Solar Observatory) or several grouped under a theme (the US National Virtual Observatory, http://www.us-vo.org/, which is for astronomy). A Virtual Observatory will typically take the form of an internet portal offering users features among the following.
  - Tools that make it easy to locate and retrieve data from catalogs, archives, and databases worldwide
  - Tools for data analysis, simulation, and visualization
  - Tools to compare observations with results obtained from models, simulations, and theory.
  - Interoperability: services that can be used regardless of the clients computing platform, operating system, and software capabilities
  - Access to data in near real-time, archived data, and historical data.
  - Additional information - documentation, user-guides, reports, publications, news, and so on.
- Virtual observatories are in varying states of development around the world - relatively well developed in some areas, while still a novelty in others. In the former case, eGY can be useful for publicizing and promoting greater use of the existing capabilities. In the latter case, eGY can be used to justify and stimulate the development of new capabilities. In all cases, eGY can be useful for informing the provider/user communities, for coordinating activities, and for promoting international standards.

Data: Diversity, Integration, Size, …

- Data policies are still highly variable or non-existent - how can data be managed to solve challenging scientific problem, societal problems without the continued need for a scientist to kno… details of complex data management systems
- Not just large (well organized, long-lived, wel… projects/programs want to make their …
- What does a large-scale, inte… like today?
  - *Most data still… use*
  - …directories, metadata,
  - …n by meta-mechanisms for …(but still need **performance**)
  - …ta assimilation
- …creasing realization: need management for all forms of 'data'

Need for VOs and size matters; personal data management is as big, or bigger problem as source data management; "my<VO>.org"

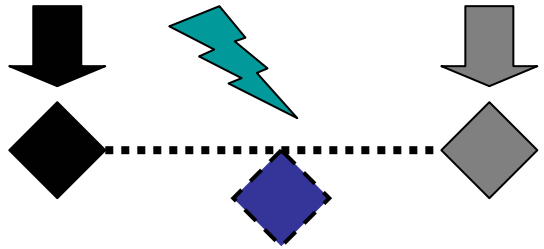
What should a VO do?



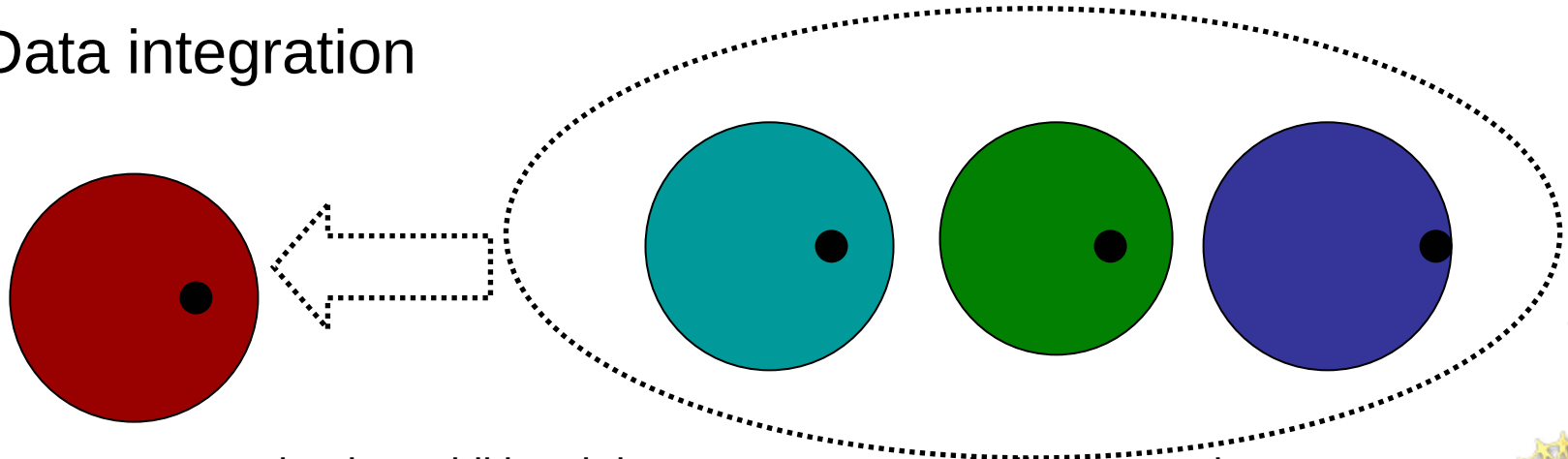
- Make “standard” scientific research much more efficient.
  - Even the principal investigator (PI) teams should want to use them.
  - Must improve on existing services (mission and PI sites, etc.). VOs will not replace these, but will use them in new ways.
- Enable new, global problems to be solved.
  - Rapidly gain integrated views from the solar origin to the terrestrial effects of an event.
  - Find data related to any particular observation.
  - (Ultimately) answer “higher-order” queries such as “Show me the data from cases where a large coronal mass ejection observed by the Solar-Orbiting Heliospheric Observatory was also observed *in situ*.” (science-speak) or “What happens when the Sun disrupts the Earth’s environment” (general public)

Virtual Observatories

- Conceptual examples:
- In-situ: Virtual measurements
  - Related measurements

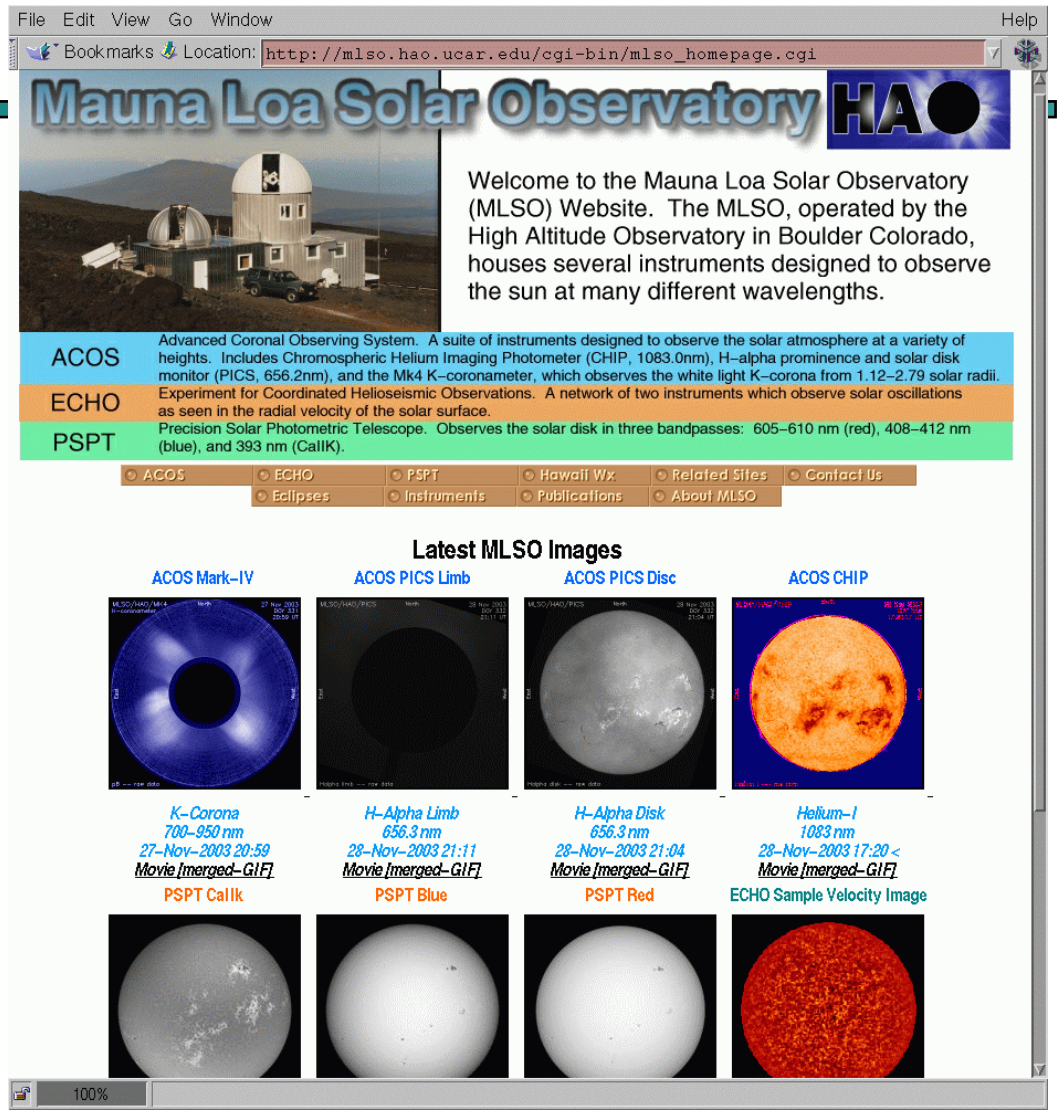


- Remote sensing: Virtual, integrative measurements
  - Data integration



- Both usage patterns lead to additional data management challenges at the source **and** for users; now managing virtual ‘datasets’

Observations of the solar atmosphere



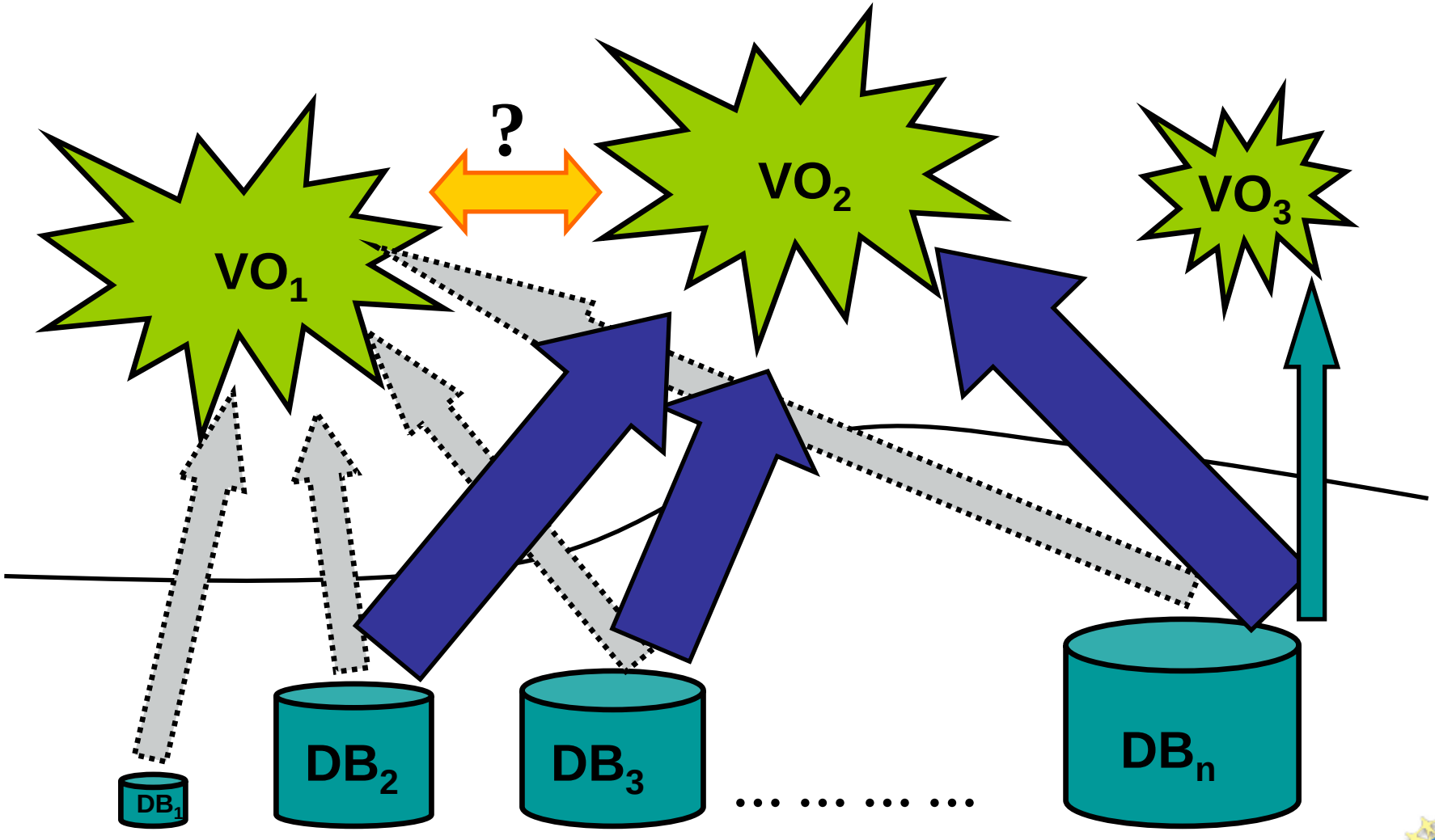
Near real-time data from Hawaii from a variety of solar instruments, as a valuable source for space weather, solar variability and basic solar physics

120 users

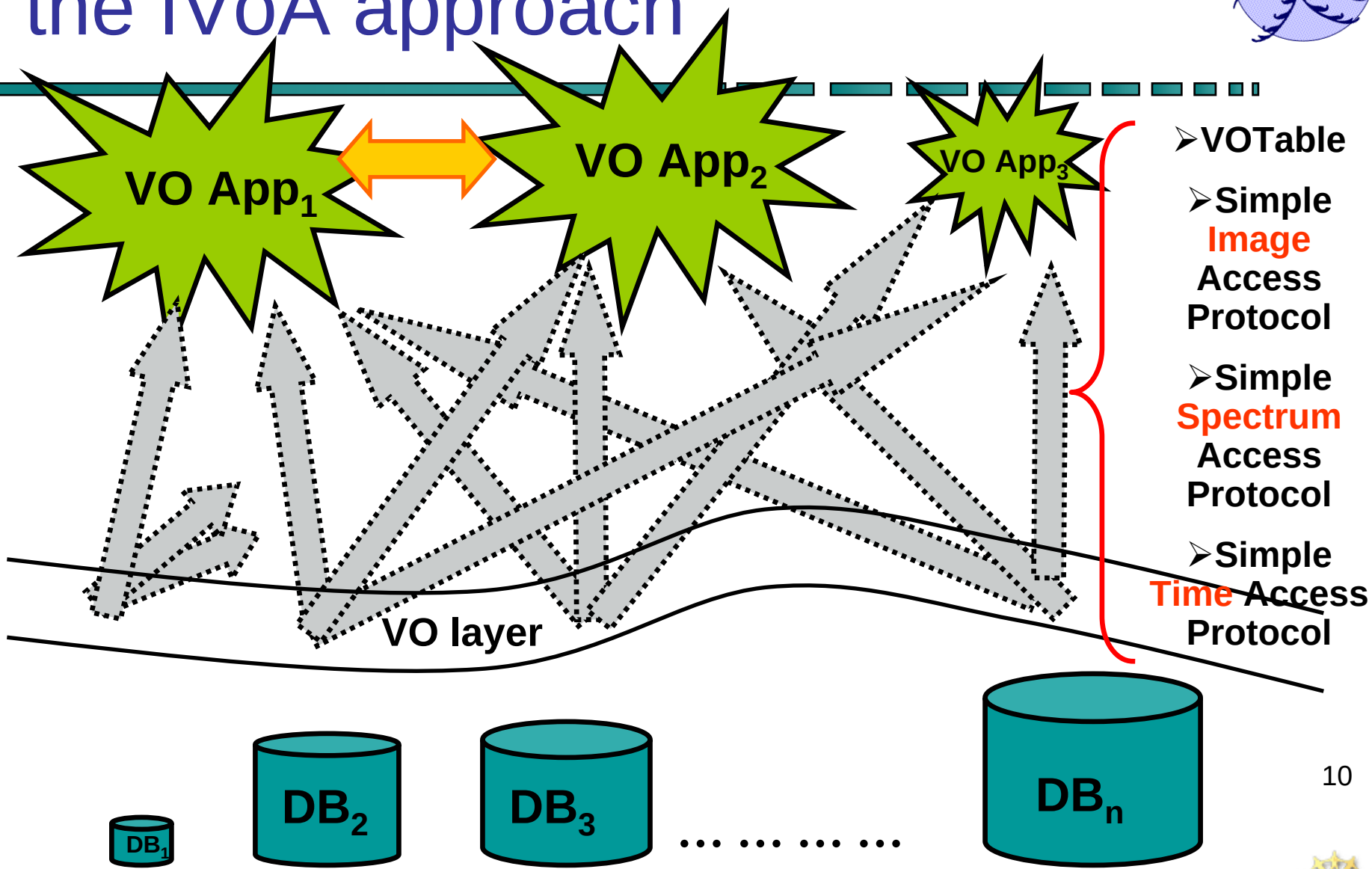
300,000 datasets

10TB +

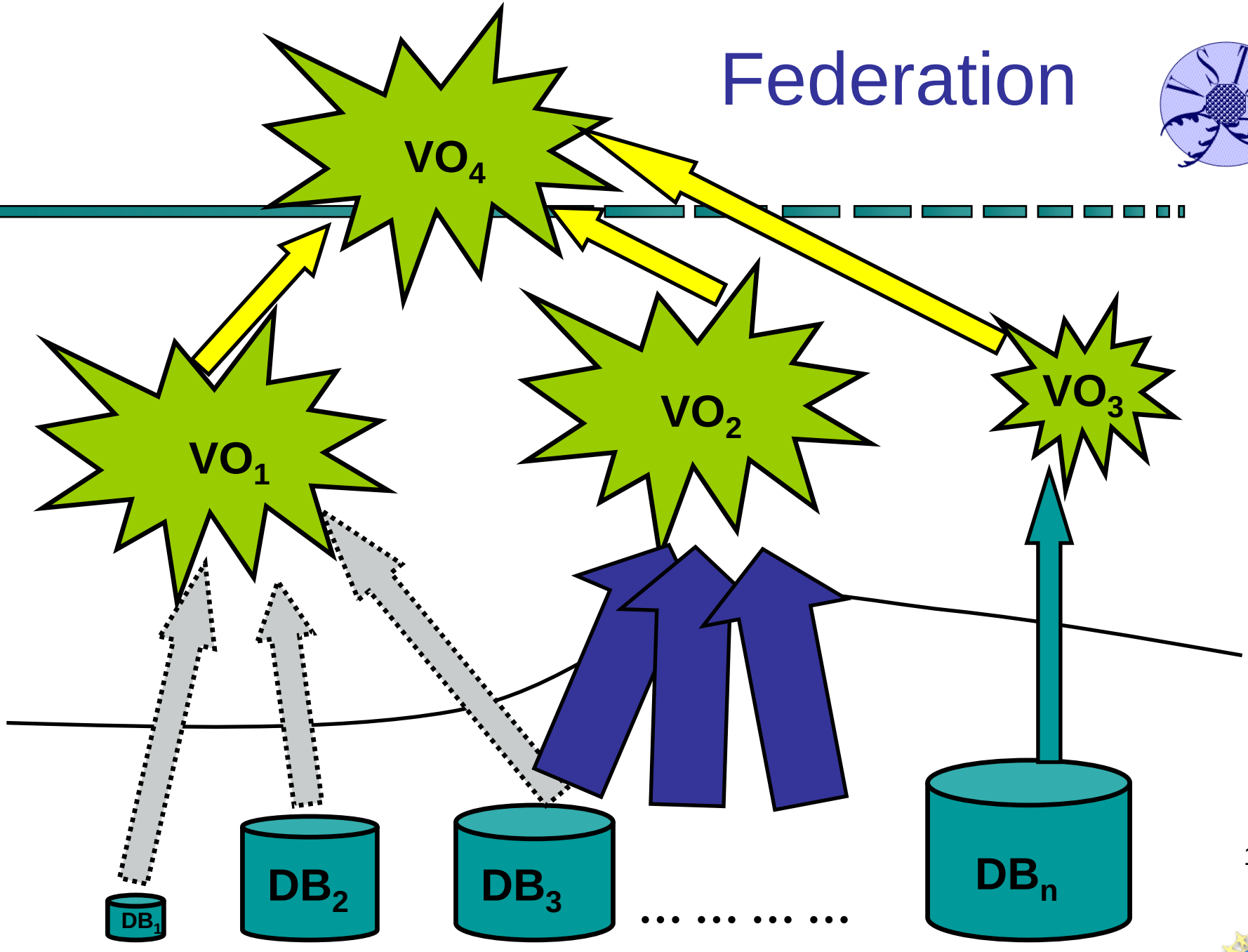
Importance of (interface) stds - early days of VxOs



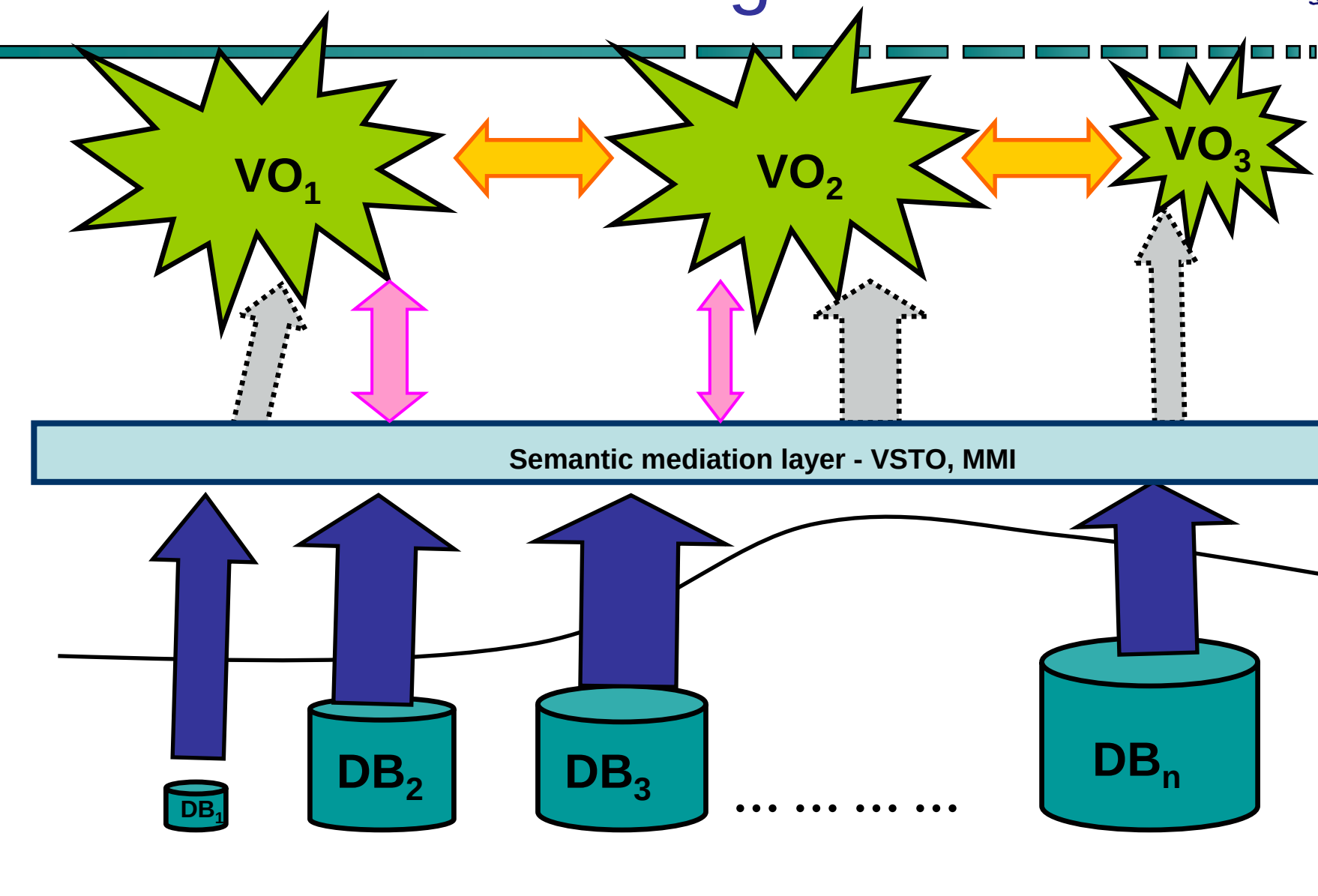
Importance of (interface) stds - the IVoA approach

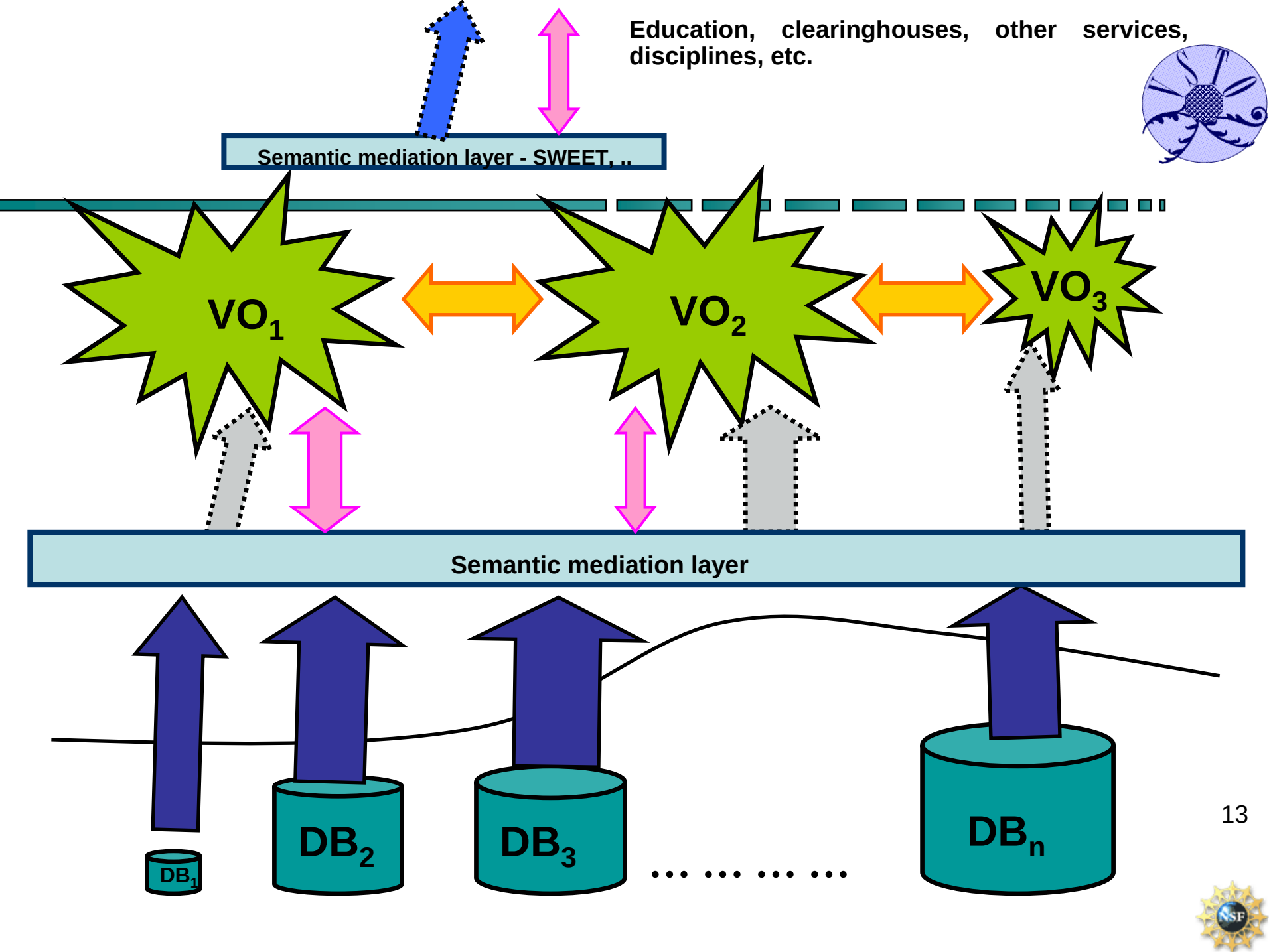


Federation



Importance of (interface) stds - Semantic VOs - e.g. VSTO



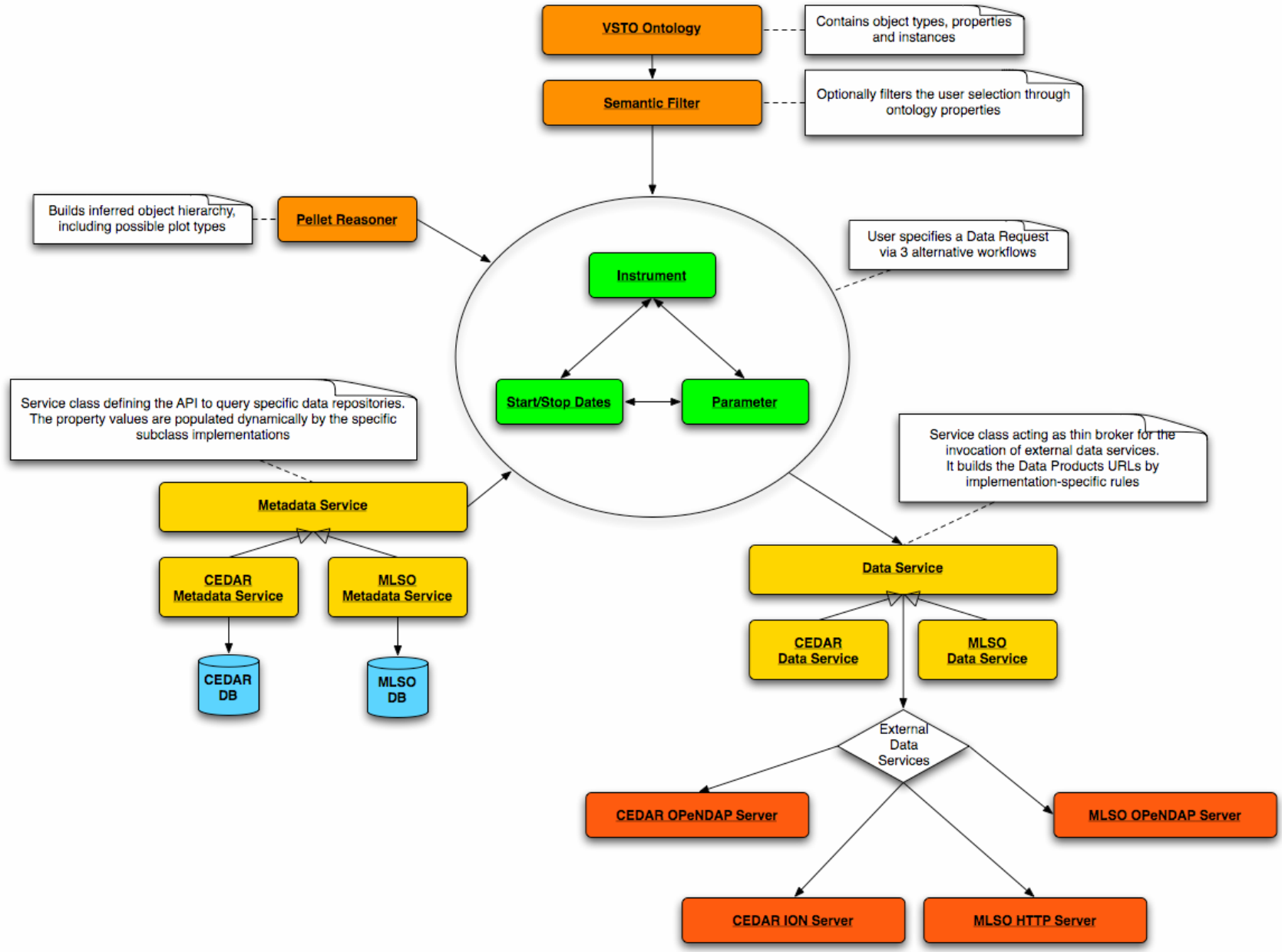

Education, clearinghouses, other services, disciplines, etc.
Semantic mediation layer - SWEET, ..
VO1
VO2
VO3
Semantic mediation layer
DB1
DB2
DB3
… … … …
DBn

Issues for Virtual Observatories

- Providing for multiple VOs: consider federating/aggregating rather than one-on-one
- Scaling to large numbers of data providers
- Crossing disciplines
- Security, access to resources, p
- Branding and attribution [illegible] from and who gets the credi[illegible] this an authoritativ[illegible]
- Pro[illegible]gating key information as it [illegible] of services, copies of processing
- L[illegible]ty, preservation, stewardship, rescue
- Funding for participation - how to leverage existing efforts
- Interoperability at a variety of levels (~3)

Semantic Web: ontologies, reasoning, etc. are one approach to address many of these issues

VSTO - semantics and ontologies in an operational environment: vsto.hao.ucar.edu, www.vsto.org



NCAR

Virtual Solar Terrestrial Observatory

Data Workflow #1a

Data Request Summary

1. Instrument:

2. Start Date:
Stop Date:

3. Parameters:

Input Step 1 of 3: Choose Instrument

Please select an instrument

You may filter the instruments selection by *one* of the following criteria:

Filter by Physical Domain: --No Filter-- Filter by Instrument Type: --No Filter--

☐ Show Instrument Code Reload

[?] **Instrument:**

- ○ OpticalInstrument > Interferometer > FabryPerot > Arecibo P.R. Fabry-Perot [?]
- ○ OpticalInstrument > Interferometer > FabryPerot > Millstone Hill Fabry-Perot [?]
- ○ OpticalInstrument > Interferometer > FabryPerot > Peach Mountain Fabry-Perot [?]
- ○ OpticalInstrument > Photometer > Chromospheric Helium Imaging Photometer [?]
- ○ OpticalInstrument > Photometer > MK3-K Coronameter [?]
- ○ OpticalInstrument > Photometer > MK4-K Coronameter [?]
- ○ OpticalInstrument > Photometer > H-alpha prominence and solar disk monitor [?]
- ○ Radar > IncoherentScatterRadar > Irkutsk Russia I.S. Radar [?]

Cancel Next >

VSTO Workflow 1c

http://www.vsto.org/data/useCase1c.htm

Virtual Solar-Terrestrial... | VSTO Home | VSTO Workflow 1c | VSTO Workflow 1a | VSTO Workflow 1b

NCAR

Virtual Solar Terrestrial Observatory

Home | Data | Communities | About Us | Logout

Start by Instrument | Start by Dates | Start by Parameter

Data Workflow #1c

Data Request Summary

1. Parameter: NeutralTemperature
2. Start Date: 2000/05/01
   Stop Date: 2000/05/11
3. Instrument: Millstone Hill Fabry-Perot

Avalaible Output

Data Files: STREAM [?] DAS [?] INFO [?] TAB [?] OPeNDAP [?] IDL [?] FLAT [?]

Data Plots: Time Series [?]

Change Input

Click on the Back button to change your data selection, or Cancel to end the workflow

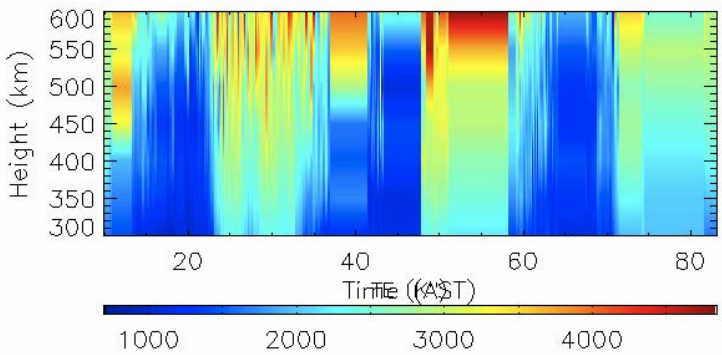
< Back | Cancel

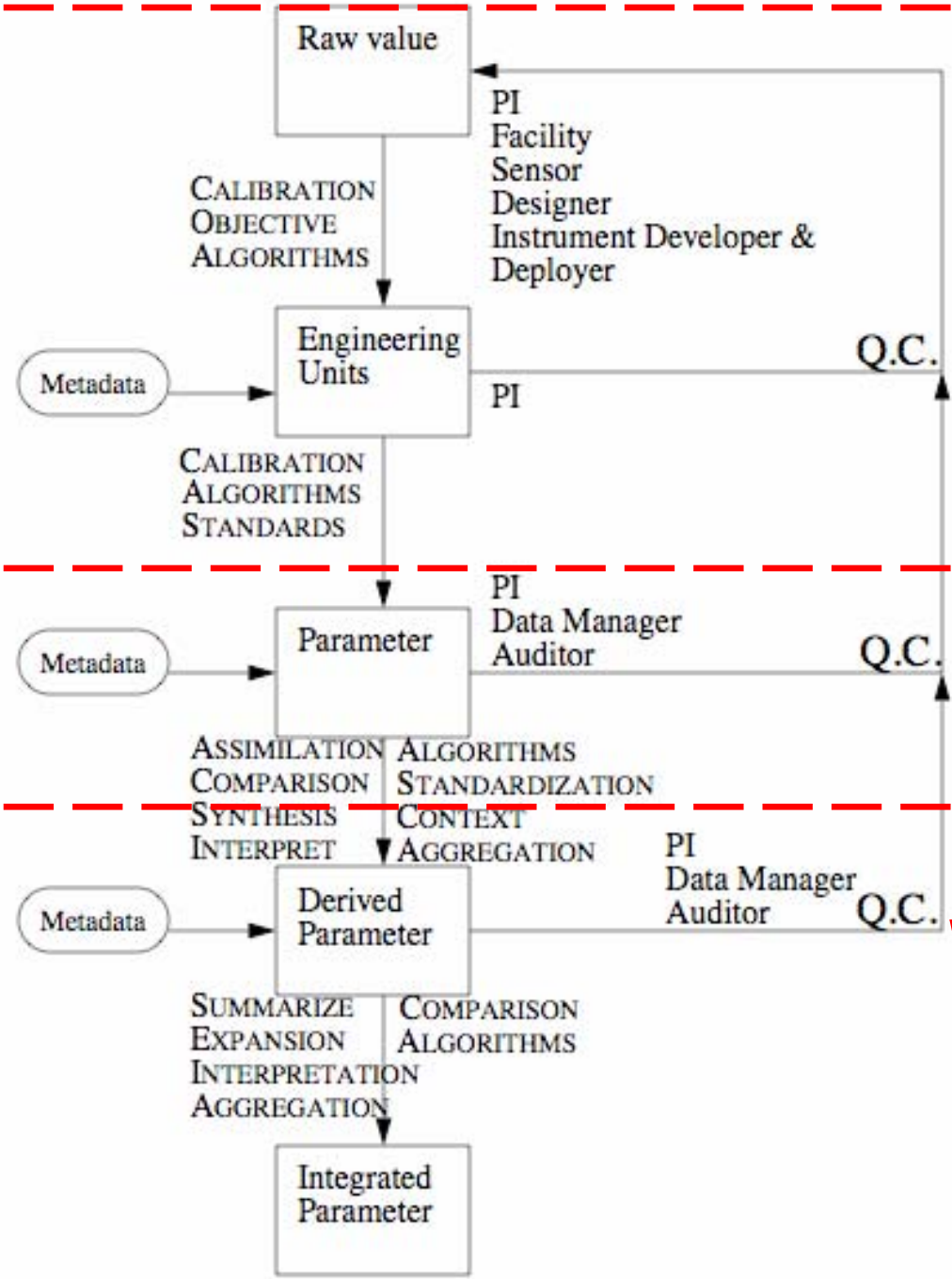
http://cedarweb.hao.ucar.edu:8081/cgi-bin/...H=2&DAY=9&NDAYS=4&PLOT_SUBMIT=PLOT_SUBMIT

http://cedarweb.hao.ucar.edu:8081/cgi-bin/ion-p?ION__E1=PLO

Instrument: 53 — Irkutsk Russia I.S. Radar
Operating Modes:
    53/9801 — Nu Te Ti Vi
Parameters:
    560 — te — Electron temperature
Starting: February 09, 1999
Ending: February 13, 1999

These plots are produced for visual browsing of the data and should not be used in publications without citing the data provider and CEDARWEB.



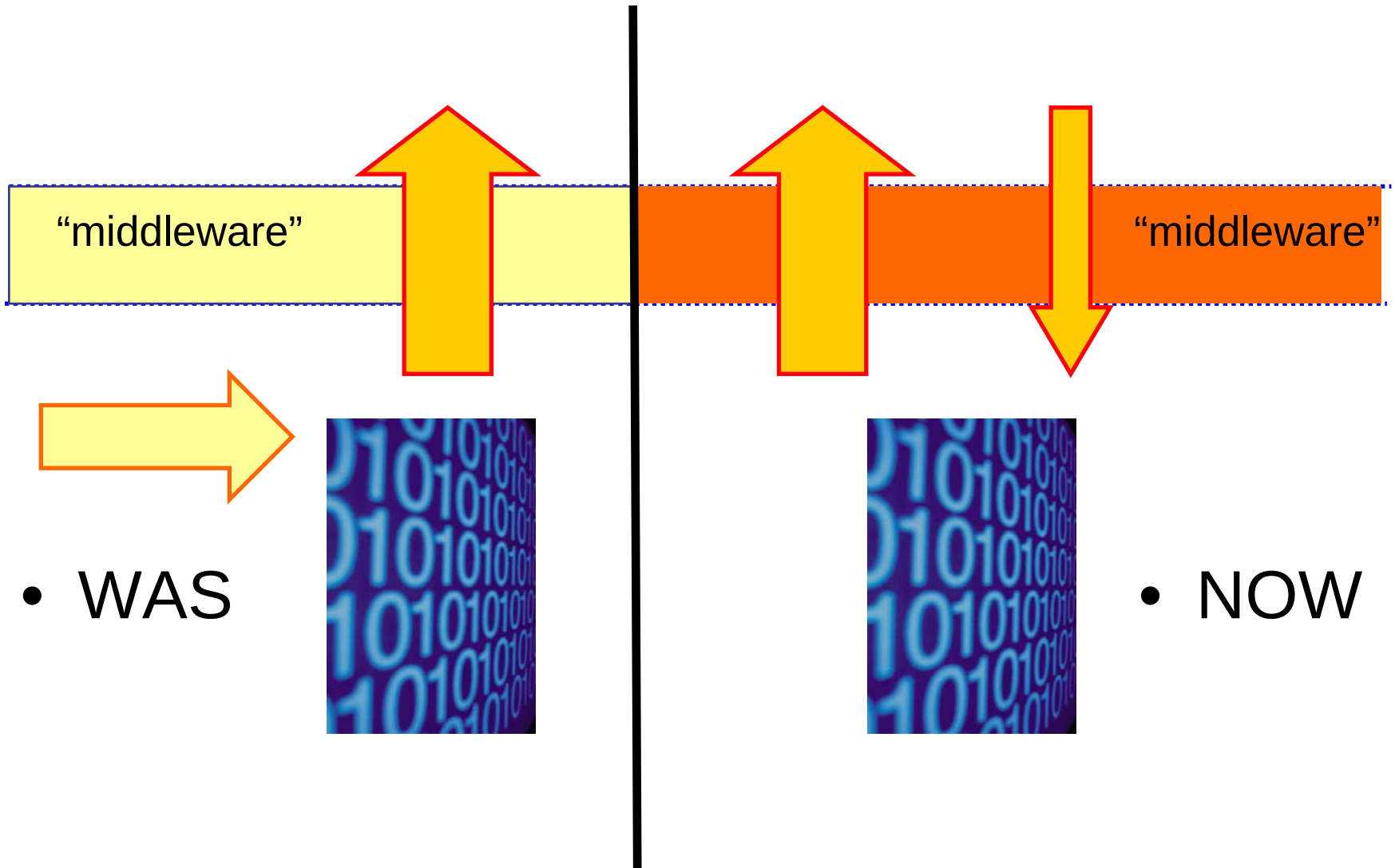


MUST BE HERE

- Mostly here

Sometimes here

Modern VOs and Data Frameworks NOT just for outflow!!



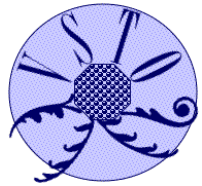
- WAS
- NOW

Final remarks

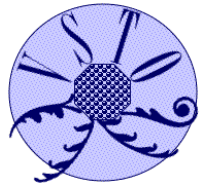
- Many geoscience VOs are in production
  - see *e*GY/VO poster (near this room)
- VO conference - April 2007 in Denver, CO
- *e*-monograph to document state of VOs
- Ongoing activities for VOs through 2008 under the auspices of *e*GY

- Contact pfox@ucar.edu

Garage

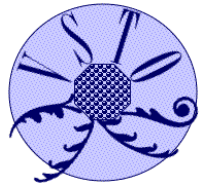


Lessons learned



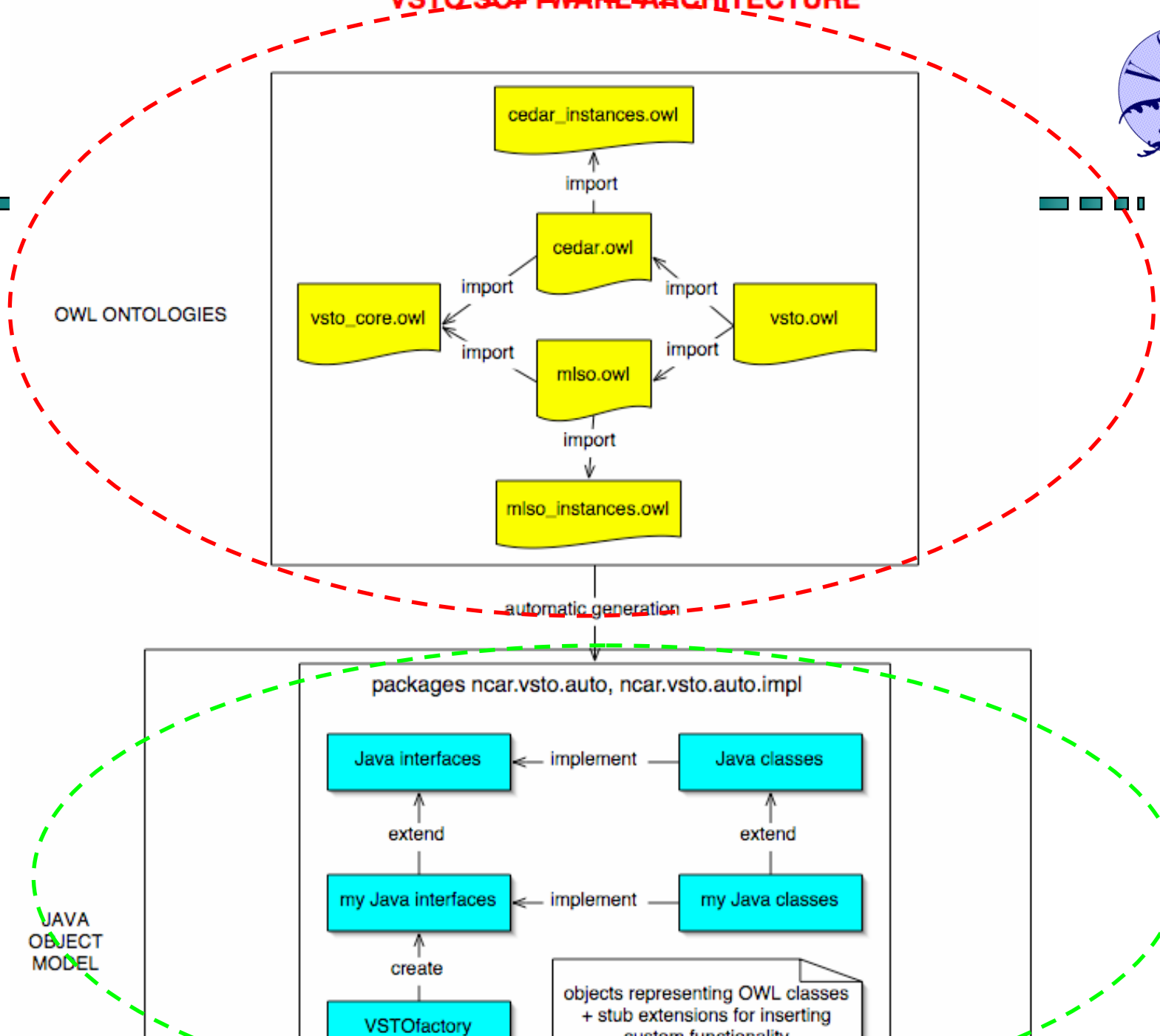
- Users, users, users
- Use cases, use case, use cases
- Same framework for all aspects of data and information flow
- Rapid development of intelligent light-weight framework and rely on services to do heavy-lifting
- Job does not end when the user gets the data (still working on this)

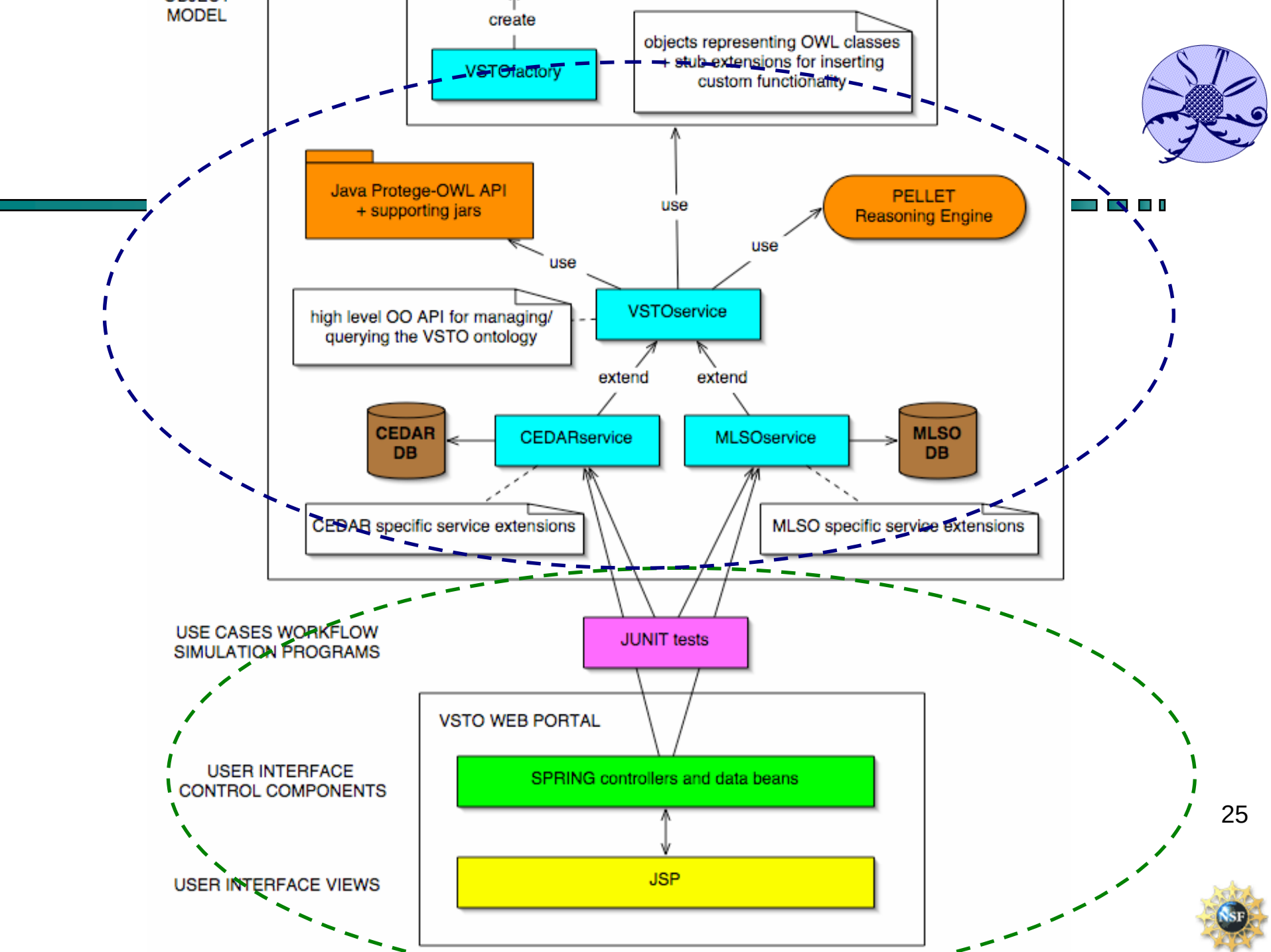
Lessons learned/ best practices



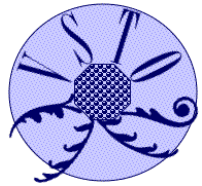
- A little semantics goes a LONG way, and a little more goes even further
- Interoperability: the few things we have to agree upon so that we need NOT agree on anything else (EC, 2005)
- Data management
- Communities
  - Providers and users are **peers**
  - Vetting of ontology - diverse community required
- People
- Software
  - We built and ‘trashed’ three prototypes in very short timeframes
  - Framework is independent of classes and individuals in ontology

VSTO SOFTWARE ARCHITECTURE

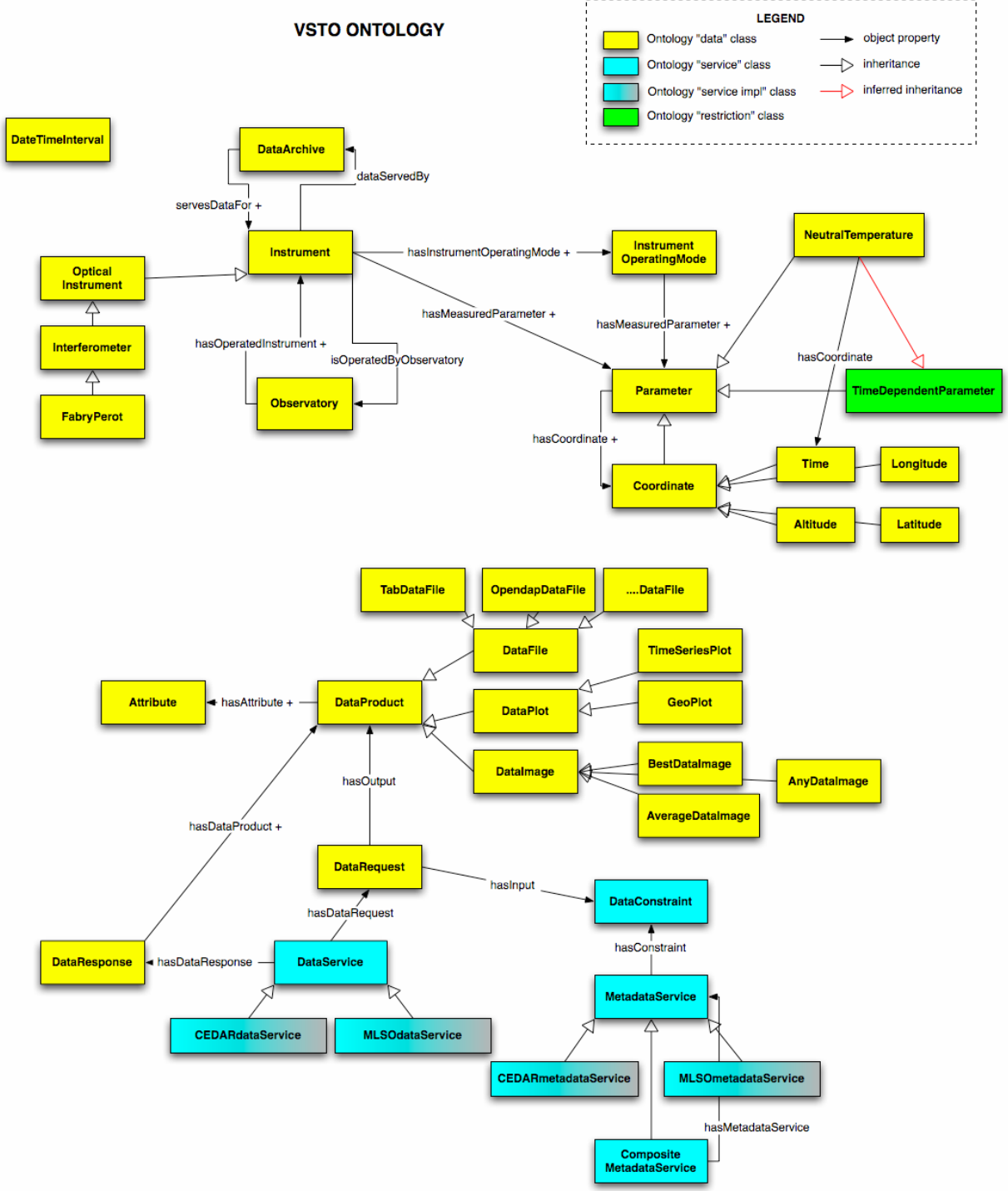


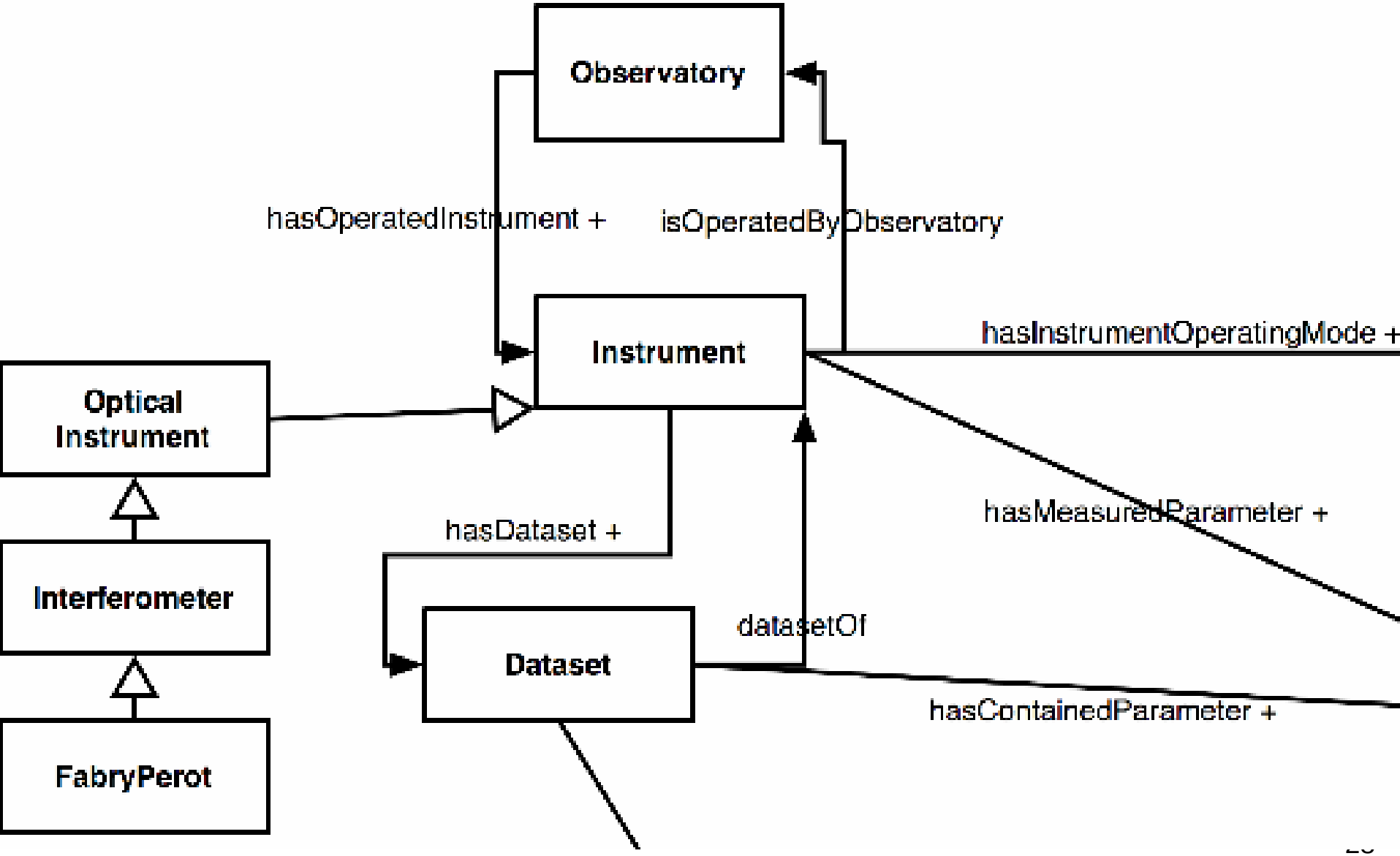

MODEL
create
VSTOfactory
objects representing OWL classes
+ stub extensions for inserting
custom functionality
Java Protege-OWL API
+ supporting jars
use
use
use
PELLET
Reasoning Engine
high level OO API for managing/
querying the VSTO ontology
VSTOservice
extend
extend
CEDAR
DB
CEDARservice
MLSOservice
MLSO
DB
CEDAR specific service extensions
MLSO specific service extensions
USE CASES WORKFLOW
SIMULATION PROGRAMS
JUNIT tests
VSTO WEB PORTAL
USER INTERFACE
CONTROL COMPONENTS
SPRING controllers and data beans
USER INTERFACE VIEWS
JSP

What's new in the VSTO?

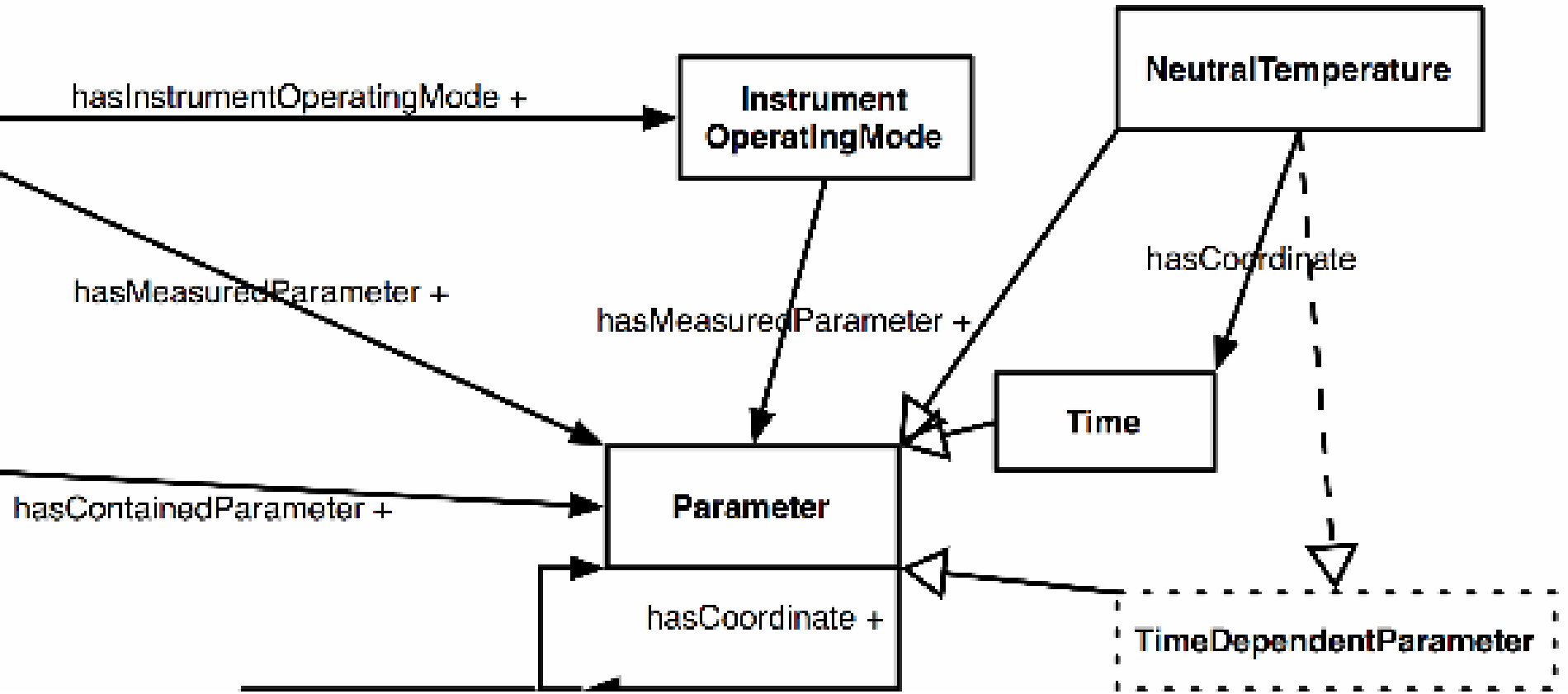


- Datasets alone are not sufficient to build a virtual observatory: VSTO **integrates** tools, models, and data
- VSTO addresses the **interface** problem, effectively and scalably
- VSTO addresses the **interdisciplinary** metadata and **ontology** problem - bridging terminology and use of data across disciplines
- VSTO leverages the development of schema that adequately describe the
  - **syntax** (name of a variable, its type, dimensions, etc. or the procedure name and argument list, etc.),
  - **semantics** (what the variable physically is, its units, etc.) and
  - **pragmatics** (or what the procedure does and returns, etc.) of the datasets and tools.
- VSTO provides a basis for a **framework** for building and distributing advanced data assimilation tools


VSTO ONTOLOGY
LEGEND
Ontology "data" class
Ontology "service" class
Ontology "service impl" class
Ontology "restriction" class
object property
inheritance
inferred inheritance
DateTimeInterval
DataArchive
dataServedBy
servesDataFor +
Instrument
hasInstrumentOperatingMode +
Instrument OperatingMode
NeutralTemperature
Optical Instrument
Interferometer
FabryPerot
hasMeasuredParameter +
hasMeasuredParameter +
hasOperatedInstrument +
isOperatedByObservatory
Observatory
Parameter
hasCoordinate
TimeDependentParameter
hasCoordinate +
Time
Longitude
Coordinate
Altitude
Latitude
TabDataFile
OpendapDataFile
....DataFile
DataFile
TimeSeriesPlot
Attribute
hasAttribute +
DataProduct
DataPlot
GeoPlot
DataImage
BestDataImage
AnyDataImage
AverageDataImage
hasOutput
hasDataProduct +
DataRequest
hasInput
DataConstraint
hasDataRequest
hasConstraint
DataResponse
hasDataResponse
DataService
MetadataService
CEDARdataService
MLSOdataService
CEDARmetadataService
MLSOmetadataService
hasMetadataService
Composite MetadataService



➢**Exploring the ontology**


hasInstrumentOperatingMode +
Instrument OperatingMode
NeutralTemperature
hasMeasuredParameter +
hasMeasuredParameter +
hasCoordinate
Time
hasContainedParameter +
Parameter
hasCoordinate +
TimeDependentParameter

VSTO Home

http://www.vsto.org/home/home.htm

Google

Virtual Solar... | VSTO Home | Workflow 1b | Workflow 1c | Communities | Documentat... | Workflow 1a

NCAR

Virtual Solar Terrestrial Observatory

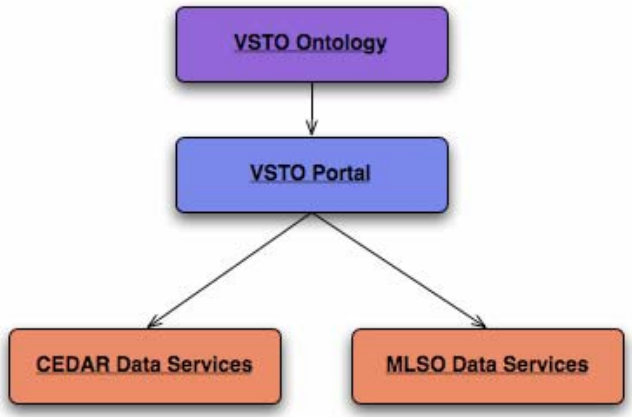
Home | Data | Communities | About Us | Login

Welcome to the Virtual Solar Terrestrial Observatory

The Virtual Solar Terrestrial Observatory (VSTO) is a unified semantic environment serving data from diverse data archives in the fields of solar, solar-terrestrial, and space physics (SSTSP), currently:

- Upper atmosphere data from the **CEDAR** (Coupling, Energetics and Dynamics of Atmospheric Regions) archive
- Solar corona data from the **MLSO** (Mauna Loa Solar Observatory) archive

VSTO Data Access



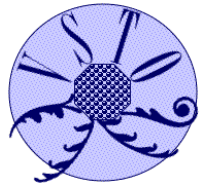
VSTO is a collaboration of the ESSL/HAO (High Altitude Observatory) and CISL/SCD (Scientific Computing Division) divisions at NCAR with McGuiness & Associates, funded by the National Science Foundation.

User: guest | VSTO Home | VSTO Project Web Site | Contact Us

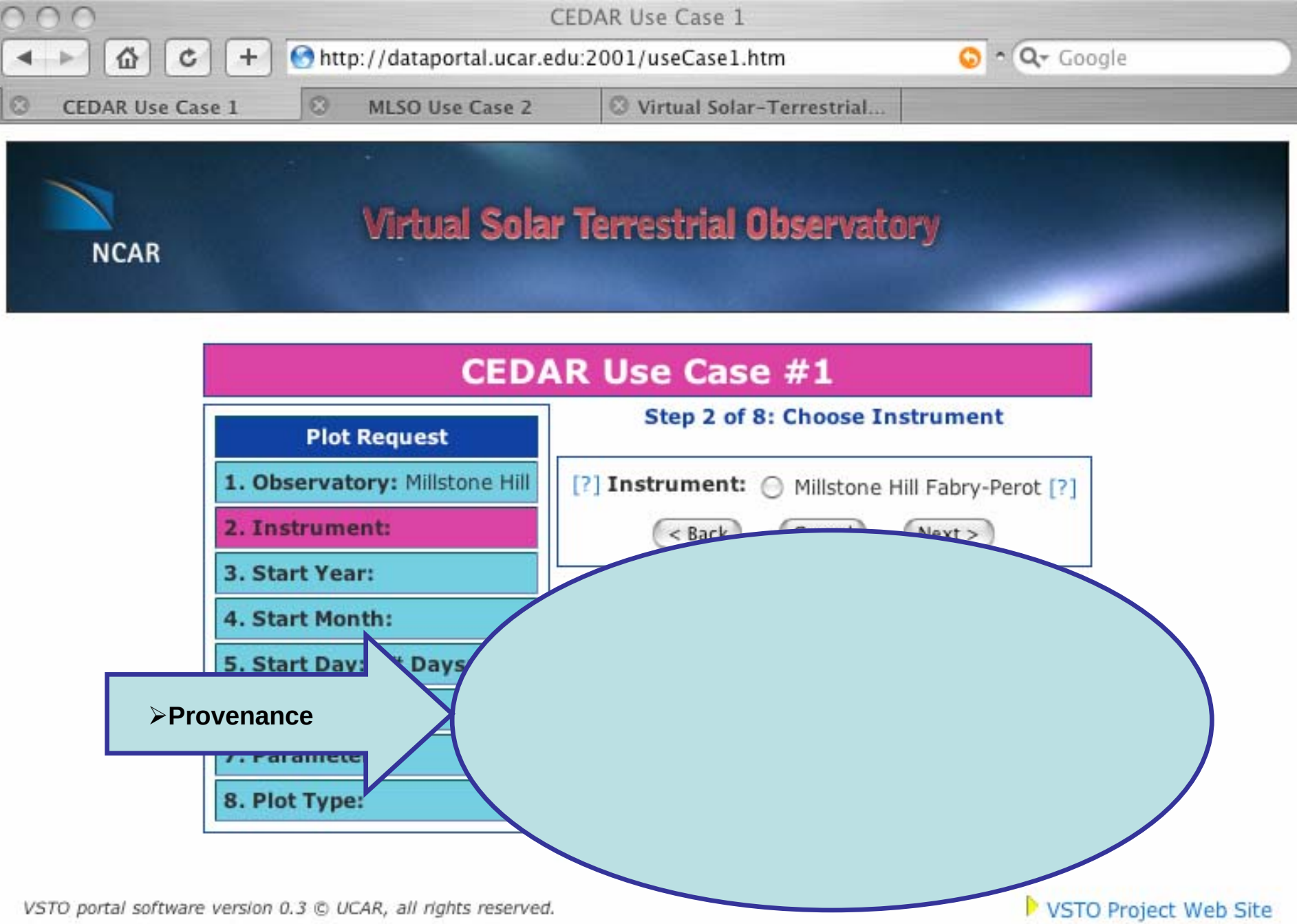
VSTO Portal Software version 1.0 © UCAR, all rights reserved.

Virtual Solar Terrestrail Observatory, funded by the National Science Fundation

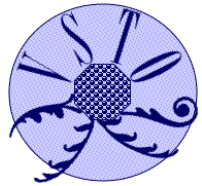
Languages and tools



- Semantic Web Languages
  - OWL Web Ontology Language (W3C)
  - RDG
  - OWL-S Messaging/services (Submitted W3C note)
  - SWSL/SWSF
  - WSMO/WSMF
  - ODM/ODD Ontology Definition Metamodel (OMG)
  - Editors: Protégé, SWOOP, Medius, Cerebra Construct, SWeDE
- Reasoners: Pellet, Racer, Medius KBS
- Other Tools for Semantic Web
  - Search: SWOOGLE swoogle.umbc.edu
  - Other: Jena, SeSAME, Eclipse, KOWARI
  - Collaboration: planetont.org
- Emerging Semantic Standards for Earth Science
  - SWEET, VSTO, MMI, …


CEDAR Use Case 1
http://dataportal.ucar.edu:2001/useCase1.htm
Google
CEDAR Use Case 1
MLSO Use Case 2
Virtual Solar-Terrestrial...
NCAR
Virtual Solar Terrestrial Observatory
CEDAR Use Case #1
Plot Request
1. Observatory: Millstone Hill
2. Instrument:
3. Start Year:
4. Start Month:
5. Start Day:
8. Plot Type:
Step 2 of 8: Choose Instrument
[?] Instrument: Millstone Hill Fabry-Perot [?]
< Back
Next >
➢Provenance
VSTO portal software version 0.3 © UCAR, all rights reserved.
VSTO Project Web Site

Integrative use-cases:



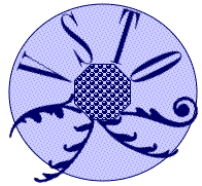
Find data which represents the state of the neutral atmosphere anywhere above 100km and toward the arctic circle (above 45N) at any time of high geomagnetic activity.

Translate this into a complete query for data. Was all the needed information recorded?

Information needs to be inferred (and integrated) from the use-case

What is returned: Data from instruments, indices and models.

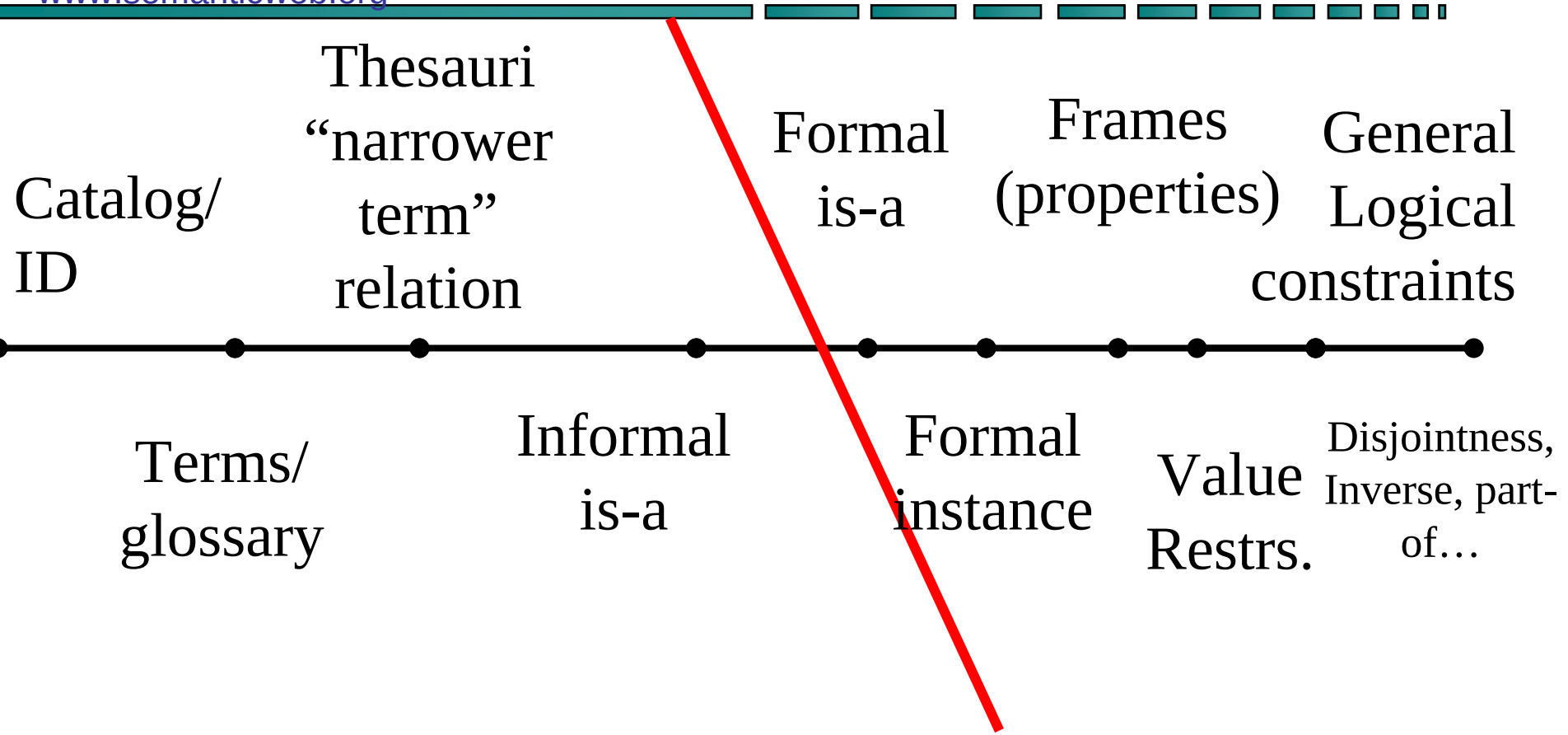
VSTO Progress



- Semantic framework developed and built with a small team in a relatively short time
- Production portal released, includes security, etc. with full community migration (and so far endorsement)
- VSTO ontology version 0.4, (vsto.owl)
- Web Services encapsulation of semantic interfaces being documented
- More use-cases to drive the completion of the ontologies - filling out the instrument ontology

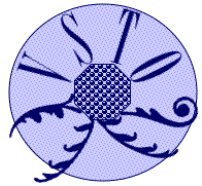
What is an Ontology: A branch of study concerned with the nature and relations of being, or things which exist. A formal machine-operational specification of a conceptualization.

Semantic Web: an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation, www.semanticweb.org



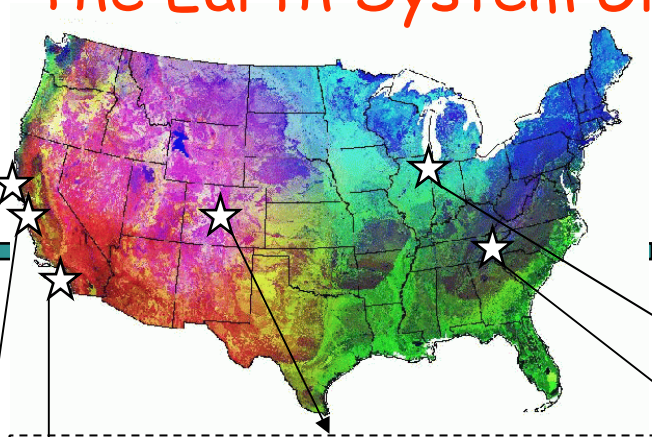
*based on AAAI '99 Ontologies panel – McGuinness, Welty, Ushold, Gruninger, Lehmann

Why we were led to semantics



- When we integrate, we integrate concepts, terms
- In the past we would ask, guess, research a lot, or give up
- It's pretty much about **meaning**
- Semantics can really help find, access, **integrate, use, explain, trust…**
- What if you…
  - could not only use your data and tools but remote colleague's data and tools?
  - understood their assumptions, constraints, etc and could evaluate applicability?
  - knew whose research currently (or in the future) would benefit from your results?
  - knew whose results were consistent (or inconsistent) with yours?…

The Earth System Grid



DATA storage
SECURITY services
METADATA services
TRANSPORT services
ANALYSIS & VIZ services
MONITORING services
FRAMEWORK services

LBNL

gridFTP server/client
HRM
DISK
MySQL
RLS
GSI
CAS client
NERSC HPSS

ANL

Auth metadata
GSI
CAS server

NCAR

SLAMON daemon
TOMCAT
AXIS
GRAM
NCL openDAPg client
LAS server
gridFTP server/client
HRM
openDAPg server
NCAR MSS
DISK
MySQL
RLS
THREDDS catalogs
Xindice
GSI
CAS client
MyProxy client
MyProxy server

LLNL

CDAT openDAPg client
gridFTP server/client
HRM
DISK
openDAPg server
MySQL
RLS
GSI
CAS client

ISI

MCS
MySQL
Xindice
OGSA-DAIS
GSI

ORNL

TOMCAT
SLAMON daemon
gridFTP server/client
HRM
ORNL HPSS
DISK
MySQL
RLS
GSI

The data grid example - data driven

- Earth Sy e system model da (July)
- 220 TB,
- Data gric gregation, bulk file
- Since De s 28TB publishe > 428,000 'files dov
  - > 250
- Gearing