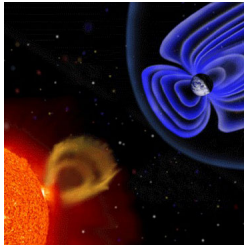


The Right Amount of Glue:

Technologies and Standards Relevant to a Future Solar- Terrestrial Data Environment

Joseph B. Gurman

*NASA Goddard Space Flight Center,
Laboratory for Astronomy and Solar Physics,
Solar Physics Branch,
Solar Data Analysis Center*



The Right Amount of Glue

George Dimitoglou

*L-3 Analytics Corp. and NASA Goddard Space Flight
Center*

Richard S. Bogart and Karen Q. Tian

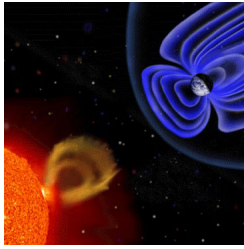
*Stanford University, Center for Space Science and
Astrophysics*

Frank Hill and Stephen Wampler

National Solar Observatory

Piet Martens and Alisdair Davey

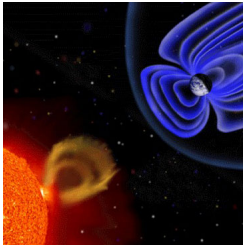
2002 December 6 *Montana State University, Physics Dept.* J.B. Gurman



The Right Amount of Glue

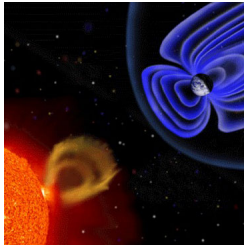
*“Too much glue won’t stick, and too many
words won’t either.”*

-purported Chinese proverb

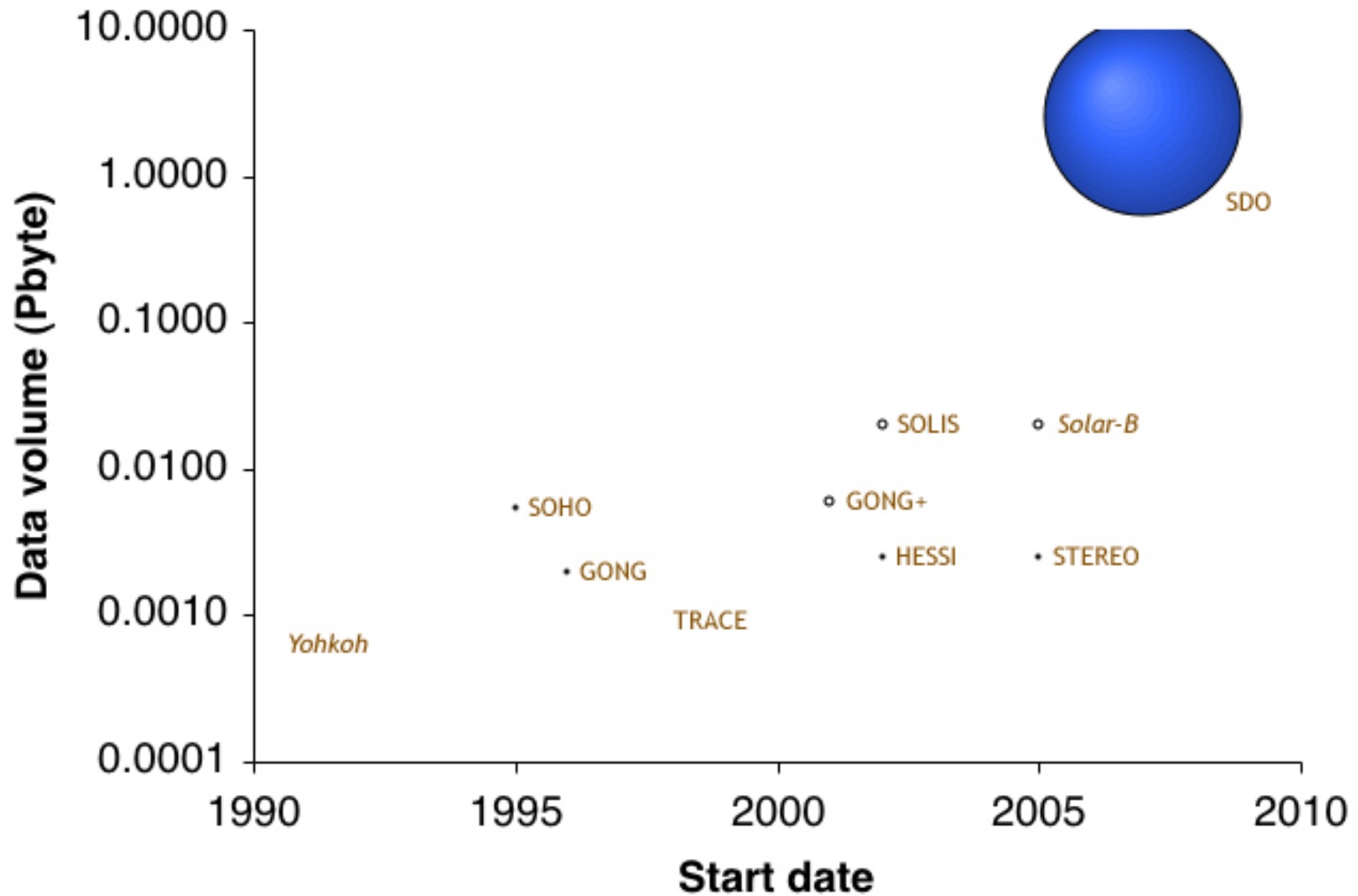


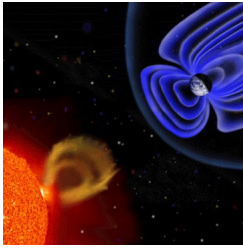
Technology

- “Technology” originated in a Greek word meaning the systematic treatment of an art or craft; literally, the root “ $\tau\epsilon\chi$ ” meant “skill.”
- Now we tend to use “technology” to mean state-of-the-art hardware or software.
- *Axiom: We already have all the technology (of the latter sort) we need to build a universal data access system*
- We must apply all the *skill* at our disposal to make such a system scientifically useful



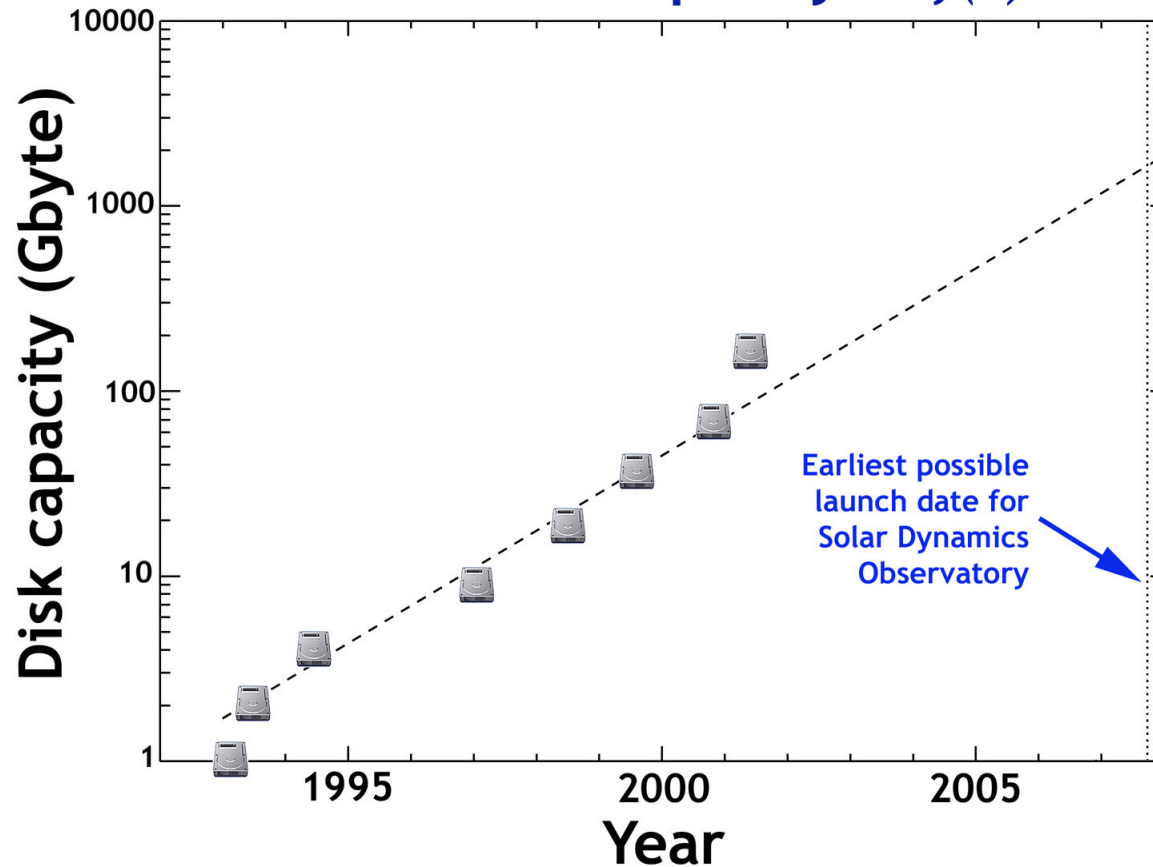
The Technological Challenge

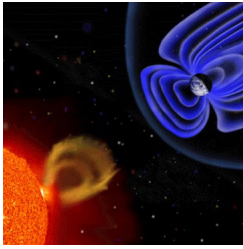




The Technical Solution (I)

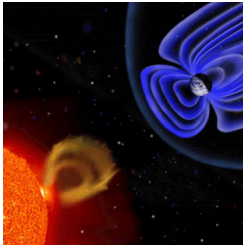
Hard drive capacity as $f(t)$





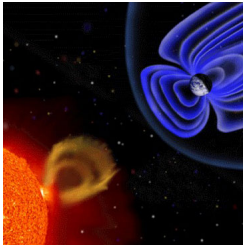
The Technical Solution (II)

- Network-attached storage (reliable RAID systems attached directly to the network) can serve large amounts of data to the Internet (*via* http, ftp) with little or no system admin overhead
- Simple database access schemes based on inexpensive db languages (*e.g.* MySQL) can provide query-based access to the largest databases we are currently envisioning



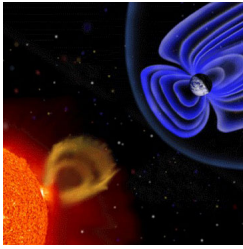
The Real Problem

- *How do we use these massive amounts of data to do new, better, or more “applied” science?*
- We need better *analysis* tools: identifying data from among a large number of sources, discovering whether the data we want are there, employing analysis software that we can understand how to use
- This is where *standards* come in
- Standards really are the glue that hold any data access and data use system together



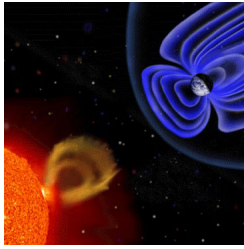
Which glue?

- Challenge: data formats
- Axiom: *Standards that have demonstrated their usefulness to the community tend to be widely adopted (marketplace)*
- Corollary: *Top-down imposition of data formats as standards rarely works (command economy)*
- Problem: Different communities use different formats, or not at all because of negative experiences with the command economy
- Solution (the right glue): XML



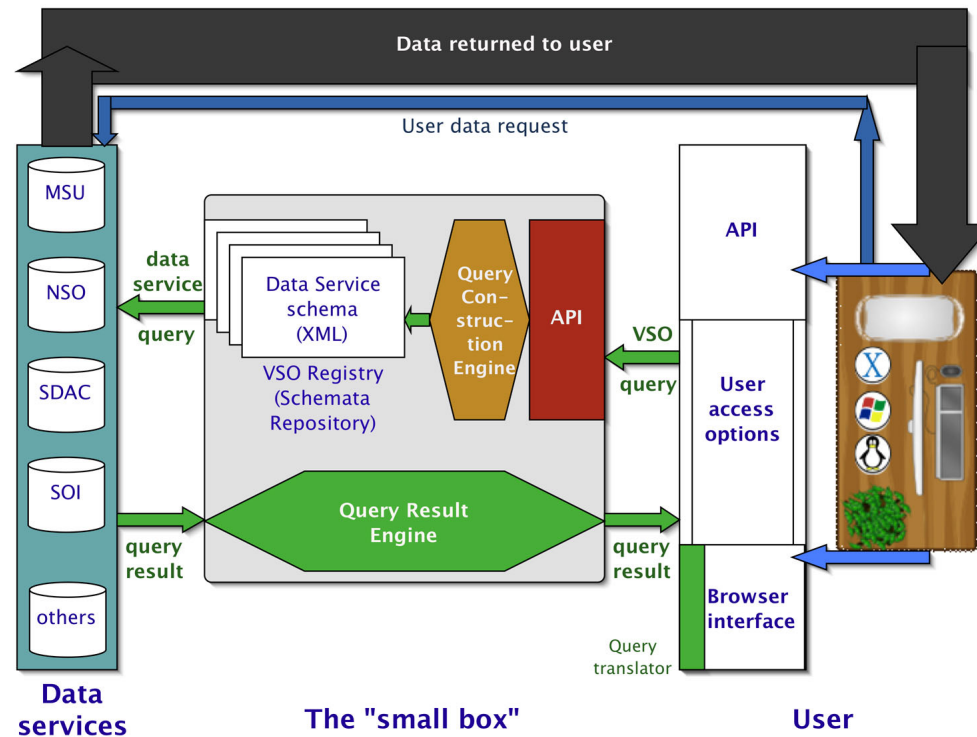
XML as glue

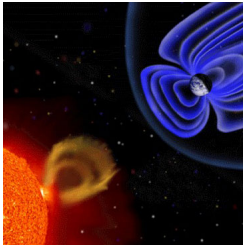
- eXtensible Markup Language
- Can be used to provide both an ASCII (human-friendly but also machine readable) representation of any data object
- D. Han (NSSDC) *et al.* have used XML as the glue in a translation facility that allows CDF, HDF, and FITS users to access each other's data without provoking religious warfare (SH51A-0424)
- Orbit display tools in SSCWeb: XML (Candey *et al.* SH51A-0422)



The Virtual Solar Observatory

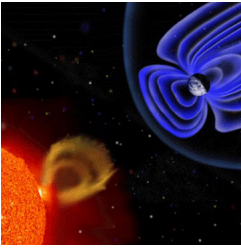
- XML can also be used to describe entire archives, e.g. the proposed architecture for the VSO





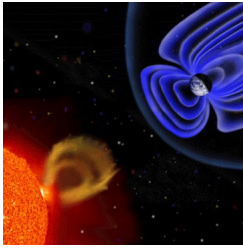
Software as glue

- Just like the English language, software can be used to unite or divide diverse cultures
- OpenDAP (Fox SH51A-0419) – another data interchange tool
- Scientific Resource Access System (Daley *et al.* SH51A-0417) for search and discovery: *It's the metadata, stupid!*
- Collaborative Sun-Earth Connector (Hurlburt SH51A-0420): distributed software
- IDL Solarsoft tree (Freeland and Handy 1998, *Solar Phys.*, **182**, 497)



An IDL Example

- Almost no one would probably chose IDL today: proprietary, “expensive,” “commercial” — but the large code and user bases make it a natural development environment
 - RHESSI synoptic database developed by Zarro, Tolbert, and Dennis
 - (<http://orpheus.nascom.nasa.gov/~zarro/spd/vsol.html>)
 - Includes a wide range of ground- (4 optical, 6 radio) and space-based (4 spacecraft) data sources
 - Accessible through Web browser or IDL session
 - Uses simple socket connections to access ftp and http data services
 - Standards: FITS, TCP/IP sockets, ftp, http.... and IDL



A Standard of Behavior

- The single most important standard is a behavioral one; namely, *a willingness to share one's data with everyone*
- Any other behavioral model retards scientific progress and drives up the costs of data systems, due to access and authentication issues
- The scientifically most successful NASA missions and NSF-funded projects in solar physics, at least, are those that have had completely open data policies