Data Systems for NASA OCT

& Relations to Solar & Space Physics Decadal Survey subcommittee recommendations

Neal Hurlburt, LMATC

2011 S&SP Decadal Survey

1. Community Input/Control of integrated data environment: There is a need for community oversight of emerging, integrated data systems. This could be through ad-hoc groups such as the NASA Heliophysics Data and Computing Working Group.

2. Incorporating emerging technologies and novel capabilities into existing data systems & infrastructure: Federal support must be agile enough to exploit emerging technologies without investing in development.

    1. Invest in technology demonstration/evaluation as part of existing missions & centers

    2. Support development needed for new platforms, e.g. clustered spacecraft

3. Continue support for development of virtual observatories and similar “middleware” through infrastructure development.

4. Support for integrated data analysis tools. In fields where free, community-developed analysis and display software is not readily available, the emphasis should be on the development of such tools. In other areas, there should be continued support for de facto standard tools.

5. Encourage semantic technologies to integrate with astronomy and geophysics communities

6.The Committee on Space Weather should undertake review and monitoring of agency data policies

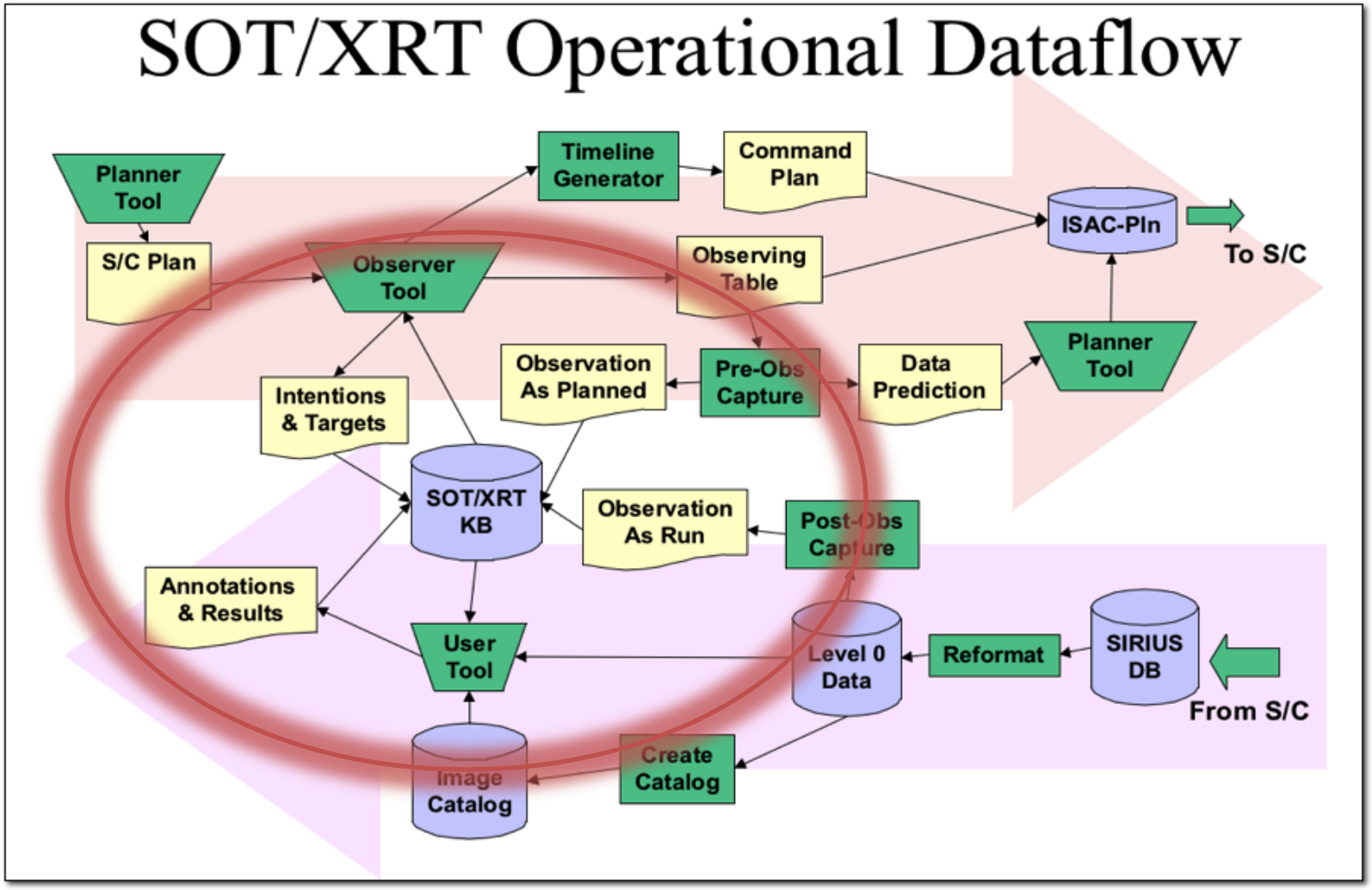
Data Systems (Table 2)

Data Intensive Systems Thrust Area	Current Capability	Future Capability
Reference Architectures	Limited reference information	Explicit models for information and technical architecture
Distributed Architectures	Limited distributed infrastructure and data sharing	Highly distributed architectures
Information Architecture	Limited semantic models	Models that capture the semantics in science and mission data
Core Infrastructure	Data management services tightly coupled	Distributed data management services (messaging and metadata/data storage)
Data Processing and Production	Locally hosted clusters and other computational hardware	Wider use of map reduce and other open source capabilities
Data Analysis	Centralized data analysis for computation, tools and data	Separation of computation, tools and data
Data Access	Limited data sharing and software services	Standards-based approaches for accessing and sharing data
Search	Product and dataset-specific searches with form fields.	
	Rich queries, including facet-based, free-text searches, web-service based indexing	
Data Movement	Limited use of parallelized and high throughput data movement technologies	Movement towards higher performing data movement technologies
Data Dissemination	Distribution tightly bound to existing data movement technologies in place.	Distribution of massive data across highly distributed environments.

Top Challenges

- Current data services are not sufficiently interoperable
- Cost of future data systems will be dominated by software development rather than computing and storage
- Uncoordinated development and unpredictable support lifecycle for infrastructure software and data analysis tools
- Need more coordinated approach to data systems software

Technology gaps in the roadmap



Advanced Mission Ops missing connection to data systems

High priority technology areas for NASA

- Data Access (from science modeling):
  - Develop tools to deal with increasing amounts of data
  - Develop fast and transparent access between distributed and remote data storage (bandwidth, firewalls) and simulations
  - Develop standards for data sharing and distribution (format, metadata, naming conventions, ontologies)
  - Should be controlled by user community, not IT
- Data Analysis
- Reference & Information Architectures
- Should avoid investing in technologies being driven by commerce/consumers

NASA's Role

- How do High priority areas align with the NASA's expertise, capabilities, facilities?
  - Support development of virtual observatories and similar "middleware" through infrastructure development.
  - Support for integrated data analysis tools.
- How is NASA's proposed technology development effort competitively placed?
  - Must be linked to missions/community to keep relevant
  - Invest in technology demonstration/evaluation as part of existing missions & centers
  - Incorporating emerging technologies and novel capabilities into existing data systems & infrastructure: NASA support must be agile enough to exploit emerging technologies without investing in development.
  - Can encourage semantic technologies to integrate across disciplines

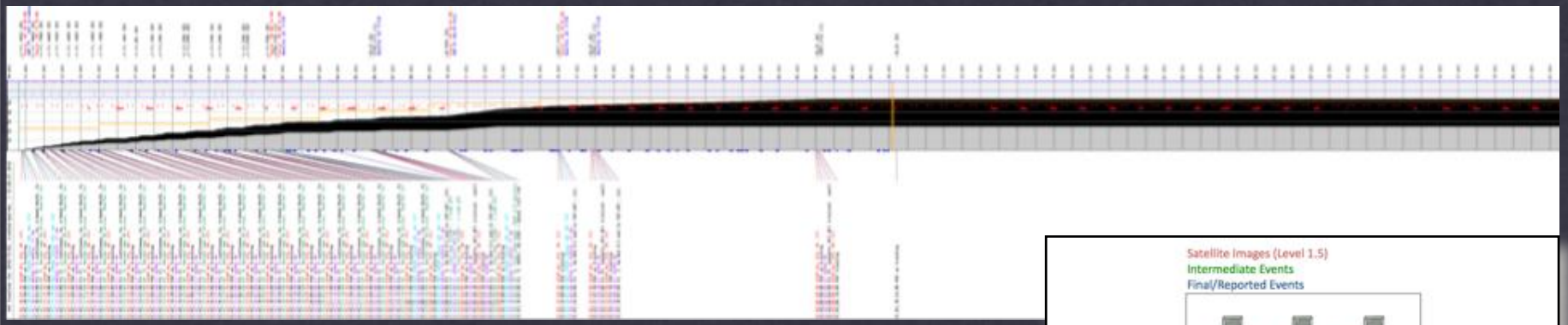
Reference Architectures

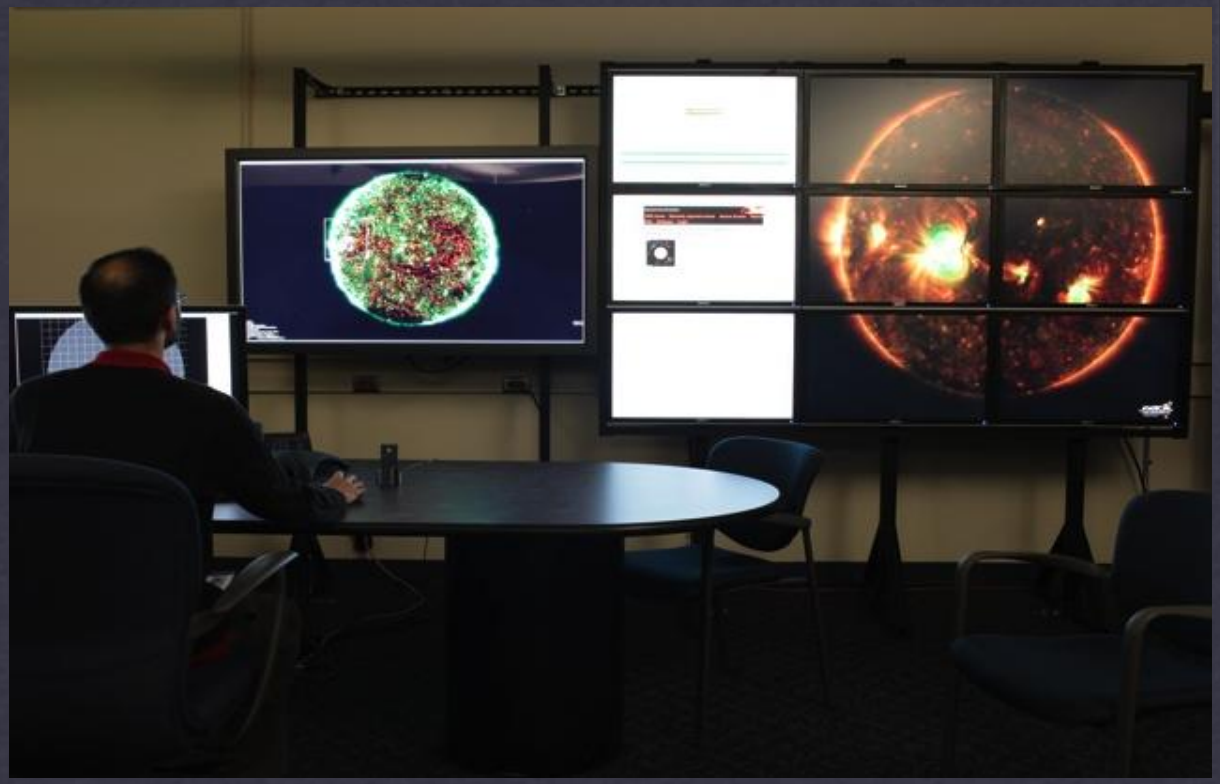
Adapted from Science modeling:

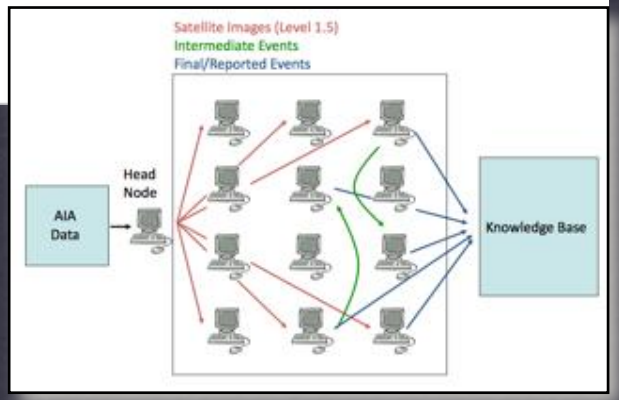
- Development of software standards and interoperability standards
- Development of software engineering tools to facilitate a transparent adaptation to architecture evolution
- Establishment of data systems testbeds and transition support

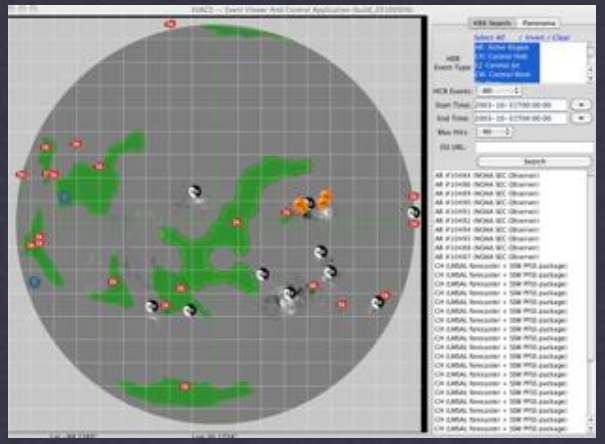
Critical Points

- "Game Changing Technology"
  - Cloud-computing could shift costs for data and analysis systems
- Tipping points
  - wide-spread use of consistent metadata/semantic annotation





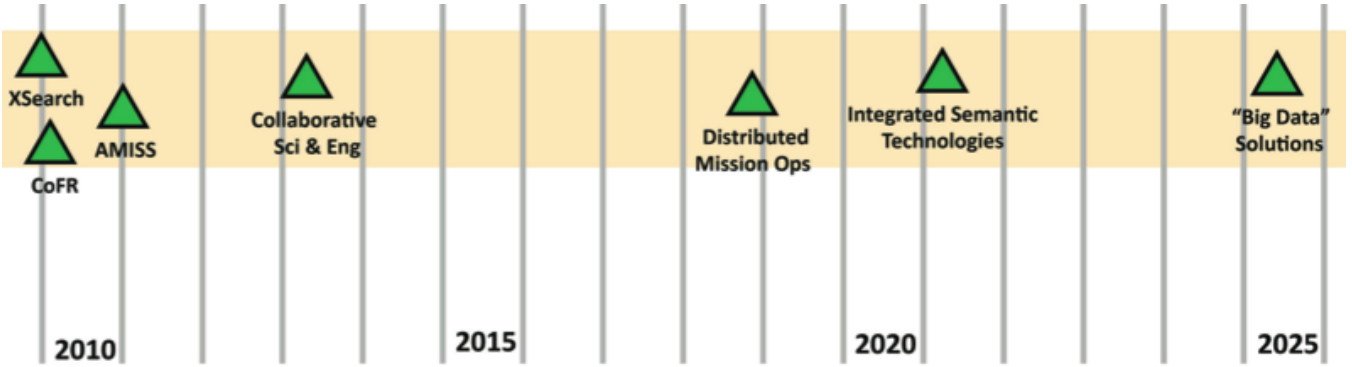




Capturing Solar Events As metadata

- Planning Inputs, Daily Visual inspection, automated detection, External contributions

Insertion Plan



- Timelines too far to the right.
  - Big data solutions are already underway (e.g. SDO, ATST) why delayed 14 years?
  - Distributed Ops already supported by some programs
- Much out of NASA's control

Parallel Development for SDO

<table>
<tr><th>Year</th><th>HPDE</th><th>Stanford</th><th>LM</th><th>LASP</th></tr>
<tr><td>Pedigree</td><td>ISTP/SOHO</td><td>MDI</td><td>TRACE/Yohkoh</td><td>UARS/SOURCE</td></tr>
<tr><td>2002</td><td>TR&T; VSO</td><td>HMI proposal</td><td>AIA Proposal</td><td>EVE proposal</td></tr>
<tr><td>2003</td><td>HPDCWG</td><td>HMI CSR</td><td>CoSEC</td><td>EVE CSR</td></tr>
<tr><td>2004</td><td rowspan="3">VxOs</td><td rowspan="3">JSOC
DRMS/SUMS</td><td>AIA CSR</td><td></td></tr>
<tr><td>2005</td><td rowspan="2">Hinode HCR</td><td rowspan="2"></td></tr>
<tr><td>2006</td></tr>
<tr><td>2007</td><td colspan="2" rowspan="2">VSO/(net)DRMS/SUMS</td><td rowspan="2">HEK</td><td rowspan="2"></td></tr>
<tr><td>2008</td></tr>
<tr><td>2009</td><td colspan="3" rowspan="2">VSO/(net)DRMS/SUMS/HEK/Helioviewer/HELIO...</td><td></td></tr>
<tr><td>2010</td><td></td></tr>
</table>