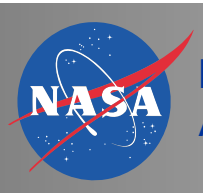


NATIONAL AERONAUTICS
AND SPACE ADMINISTRATION

Semantic Web Infusion Roadmap V1.0 Gap Analysis 1.6

NASA/ESDSWG/TIWG

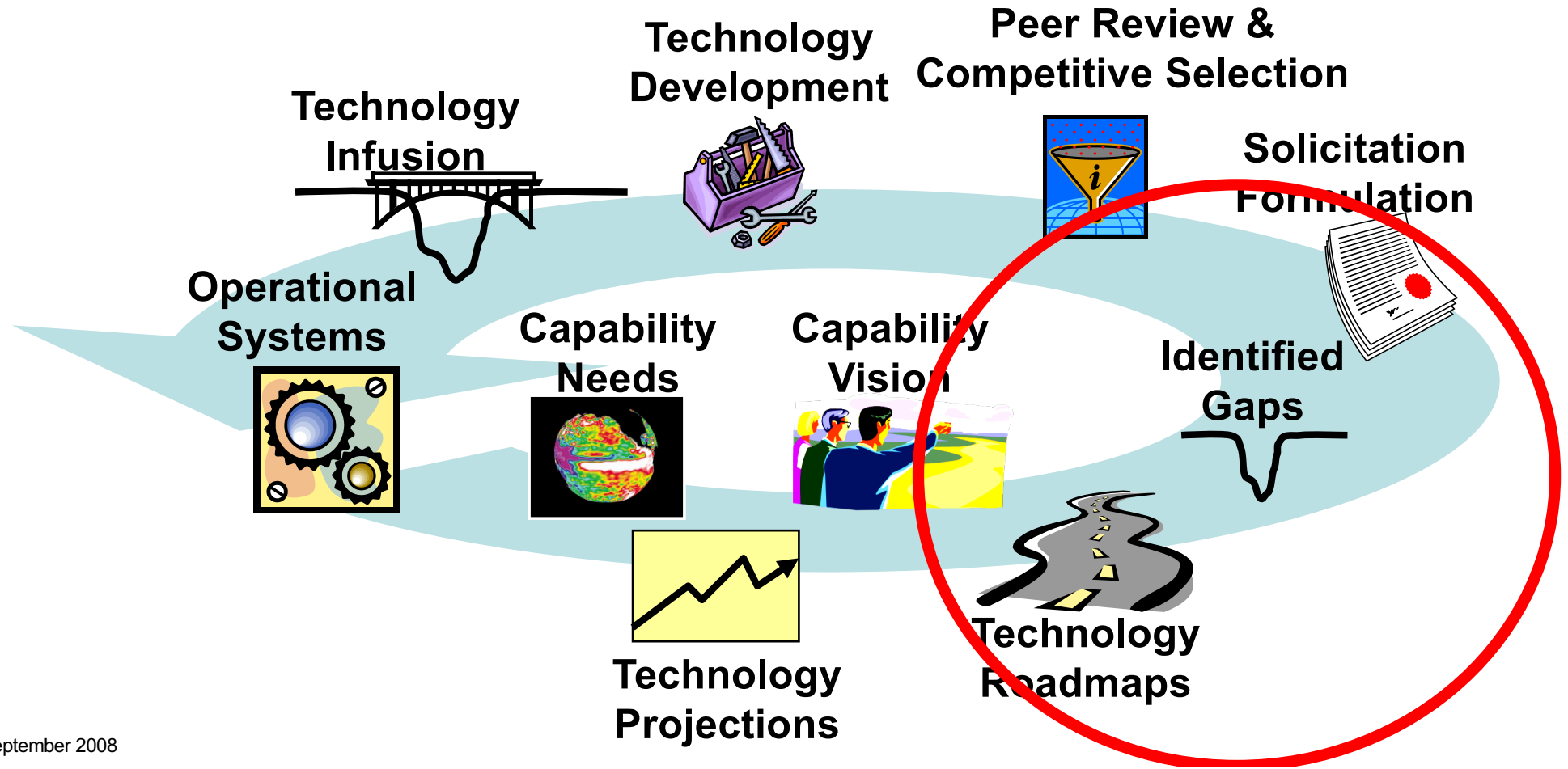
April-November 2008

Semantic Web sub-group

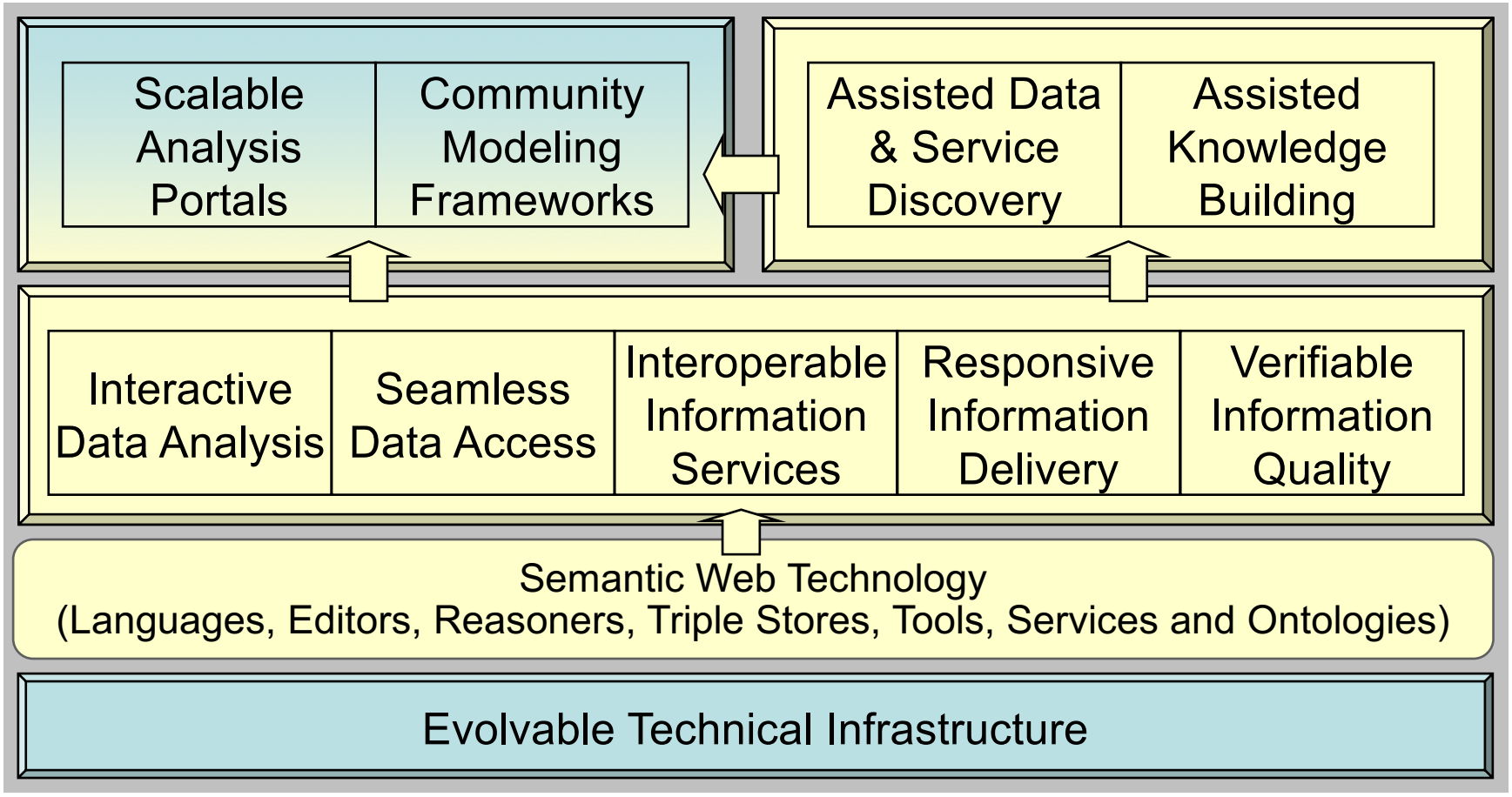
Presented to ESIP-STC April 2018

Background: Technology Infusion Process

- Established a capability vision for Earth science information systems
- Identified Interoperable Information Services as a key capability in the vision
- Identified semantic web as one of the primary supporting technologies
- Currently defining a roadmap for semantic web technology infusion

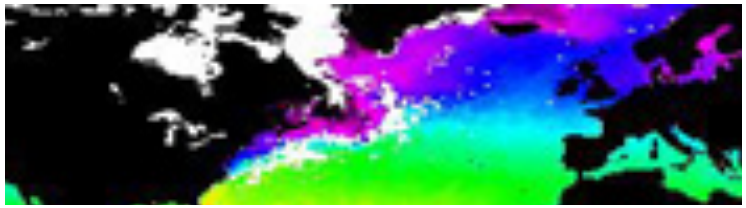


Infusing Semantic Web Will Help Realize the Vision for all Middleware Services and Assisted Capabilities

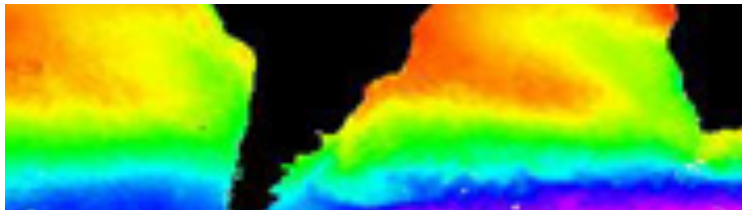


Interactive Data Analysis Inventory

- Current technologies
  - Visual grammars, shared terminologies for visual properties
- Needed
  - Semantically-aware visual programming environments and high-level analysis tools
  - Tagging data properties with metadata, mapping to non-jargon vocabularies
  - Data mediation (units, coordinates)
  - Vocabulary translation for machine-to-machine processing
  - Semantic support for coordinate systems/ projections/ scale factors/ offsets/ special values
  - Support for data quality
  - Support for displaying errors
  - Understanding of minimum/ maximum, color bars (and relation to data)

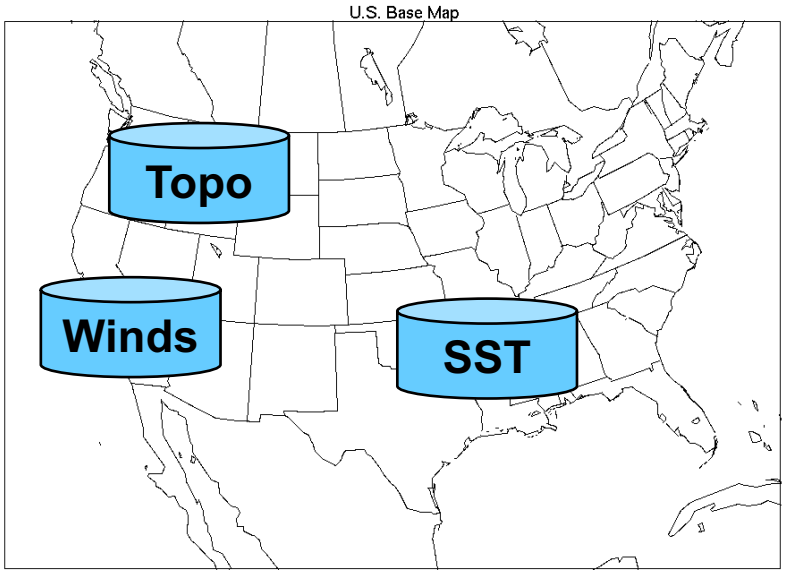


$$\rho C_{Pg} u \frac{\partial T}{\partial x} = \lambda \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) + G$$



Seamless Data Access Inventory

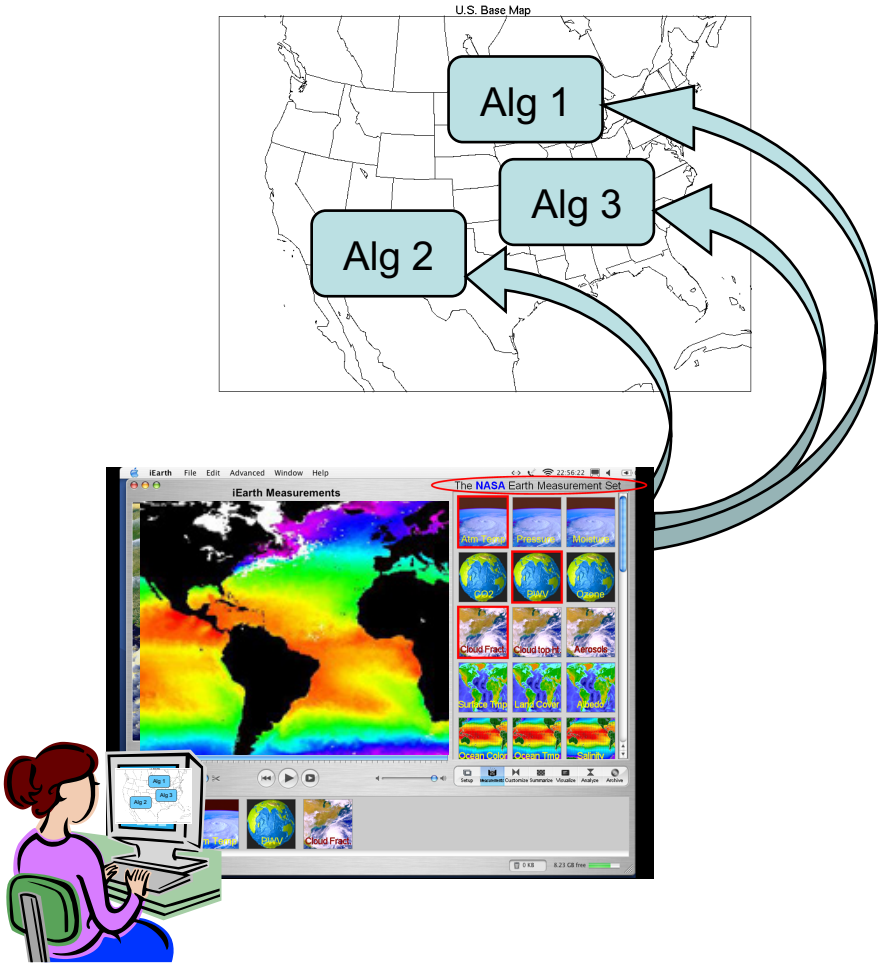
- Current technologies
  - Community network data access protocols (OpenDAP, WMS/WCS, WebDAV, GridFTP)
  - Established data server tools (MapServer, GDS/LAS, ArcWeb) integrate data analysis in local environment from outside sources
- Needed technologies
  - Semantic metadata (OWL-S) to enhanced existing hard-coded services (WxS)
  - Data-type ontology
  - Services ontology (inputs/ outputs) for service chaining
  - Semantic markup for sensors and scheduling/ planning
  - Tools interoperating with other tools at a semantic level



Interoperable Information Services Inventory

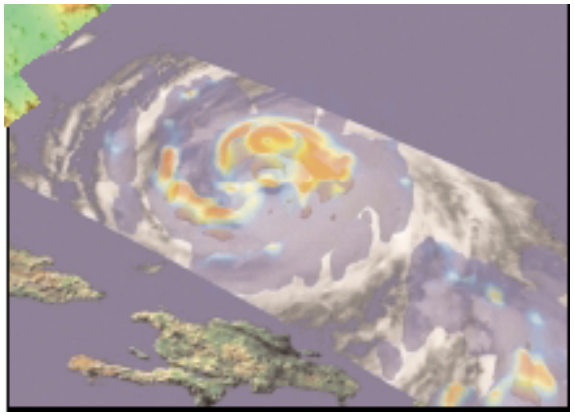
- Current technologies
  - Network service protocols (SOAP, Java RMI, OPeNDAP, CORBA)
  - Utility/grid computing protocols & toolkits (Globus)

- Needed technologies
  - Service and domain ontologies working synergistically
  - Services that know about other services, and their levels
  - Services understand terms, quantities, units, coordinates
  - Semantic service registry
  - Smart service chaining
  - Smart service gap filling



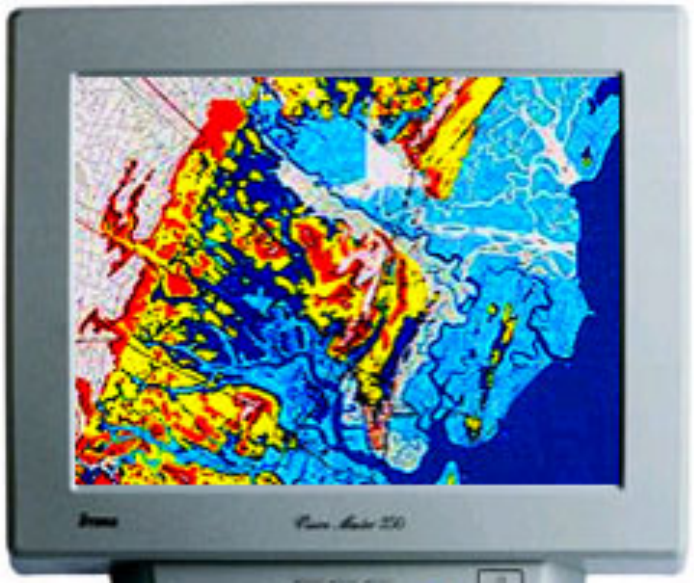
Responsive Information Delivery Inventory

- Current technologies
  - Optical networks (National LambdaRail)
  - Peer-to-peer networks with swarming (Modster)
  - Direct downlink (MODIS/AIRS DDL)
  - Sensor tasking in response to event/ prediction
- Needed technologies
  - Semantic service bindings and groundings include resource estimates, priority scheduling, spectrum of data latency in common terms
  - Late semantic binding for real-time streams
  - Semantic event detection and tagging



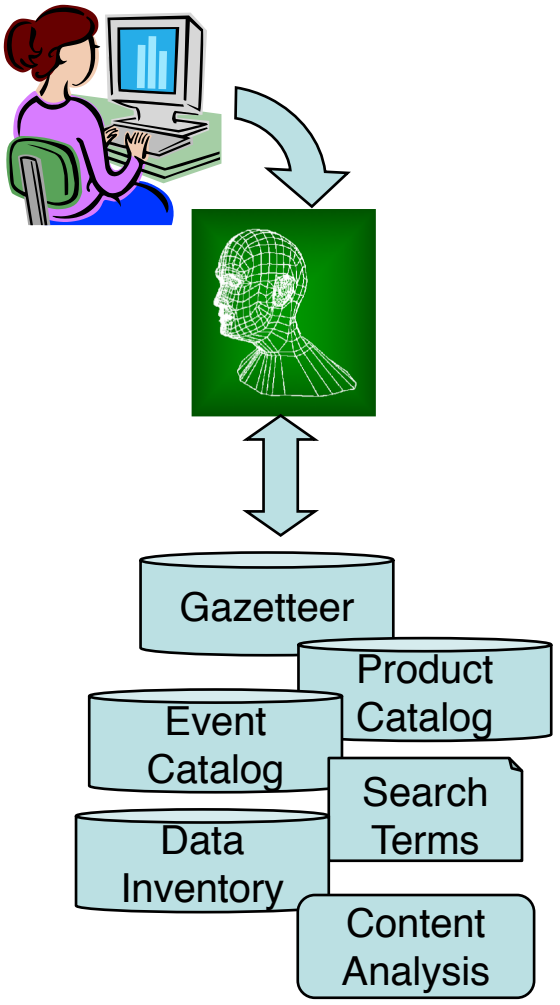
Verifiable Information Quality Inventory

- Current technologies
  - Data pedigree algorithms (Ellis)
  - Machine-readable formats (XML) and common semantic service ontology (OWL-S) including data-types
- Needed technologies
  - Data quality (all dimensions conveyed in meaningful terms), uncertainty and provenance ontologies
  - Domain ontologies include domain and range value restrictions
  - Making provenance from two or more services interoperable
  - Smart quality propagation from two or more information services
  - Rulesets, explanation, trust and proof inferencing



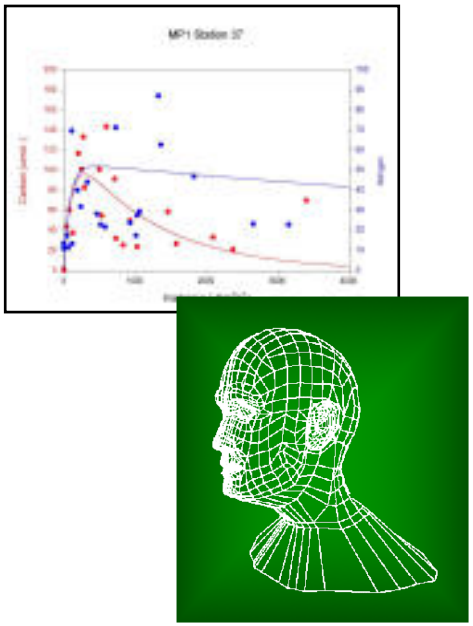
Assisted Data & Service Discovery Inventory

- Current technologies
  - Smart query with term expansion, narrowing, reasoning, abstracting selection workflows, capturing concepts of discovery, inventory and item/granule -level in finding data
  - Data and service description standards, web service directories, syndication services, topic/ concept maps
  - Established directory services (GCMD, ECHO, THREDDS)
  - Domain ontologies (SWEET, MMI, VSTO, …), rule-based logic, semantic query
- Needed technologies
  - Community standards are needed to avoid wasted and inconsistent efforts
  - Smart mediation among catalogs that are using ontologies and/or standard data model
  - Semantic service registry
  - Data-type and service ontologies
  - Smart discovery of virtual data products (want data but need service to create it for you)
  - Smart crawlers to pull



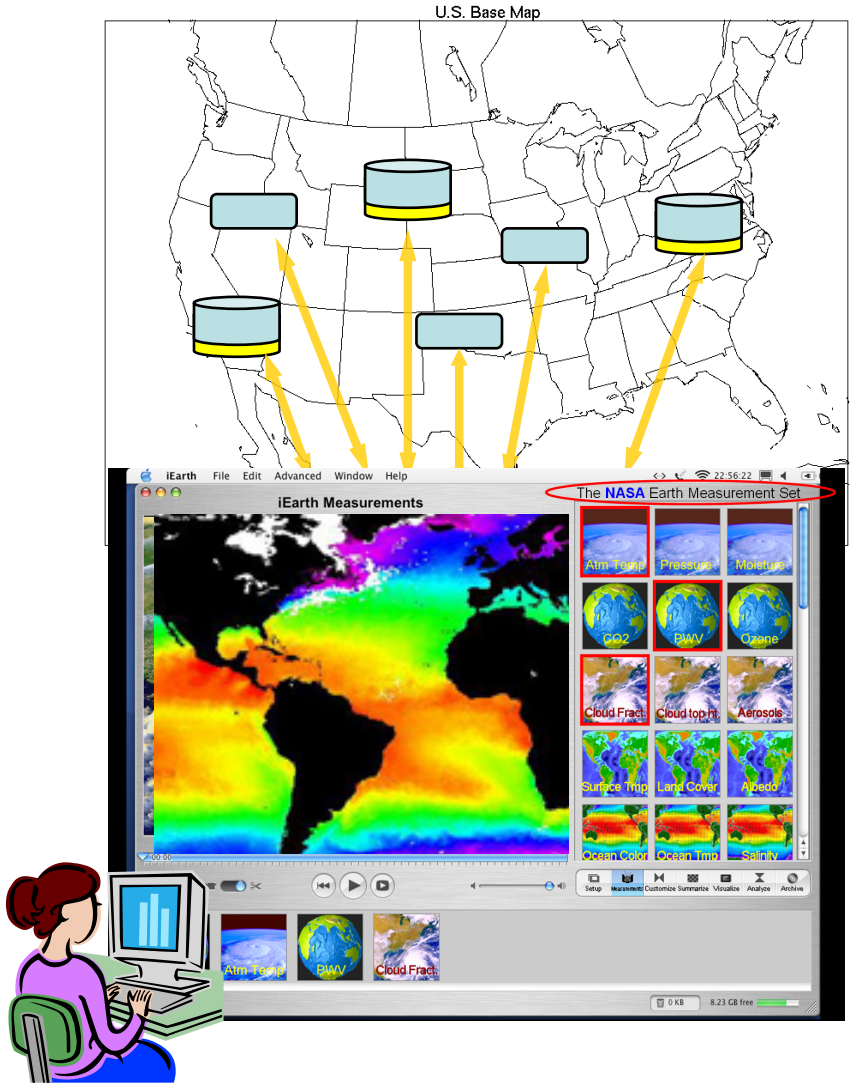
Assisted Knowledge Building Inventory

- Current technologies
  - Data mining algorithms (Support vector machines, independent component analysis, rule induction) and ontologies
  - Data mining toolkits (Adam, D2K, Darwin) and plug-ins (IMAGINE, ENVI, ArcGIS) with semantic annotations
  - Data and service description standards, web service directories, syndication services, topic maps
- Needed technologies
  - Cross-domain data mining and fusion and rule-based smart data mining and fusion
  - Ontologies for visualization and analysis, metadata annotation with ontologies
  - Conversion of folksonomies to ontologies
  - Formalizing the long-term de-facto standards into formal ontologies



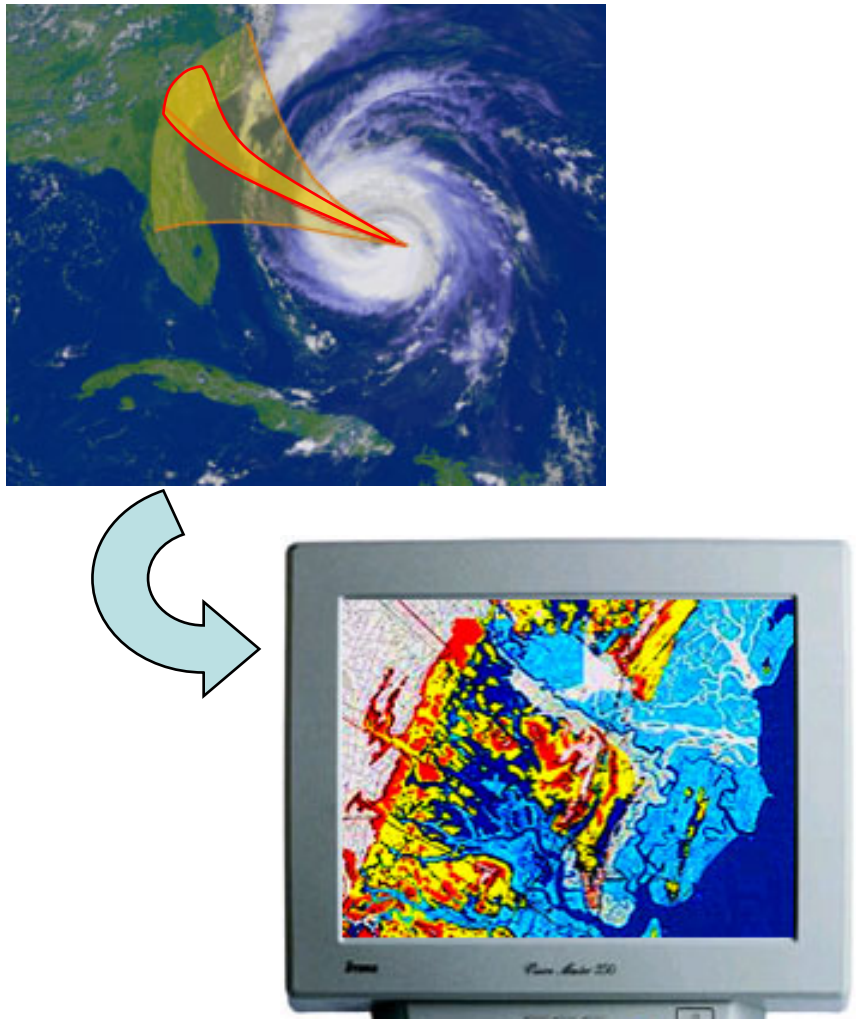
Scalable Analysis Portals Inventory

- Current technologies
  - JSR168/268 porlets
  - Several implementations; Gridsphere, JetSpeed, et c.
- Needed technologies
  - Ontologies for portal modal functions
  - APIs for accessing domain ontologies
  - Capabilities to propagate provenance and other key metadata
  - Ontology to describe analysis functions and results

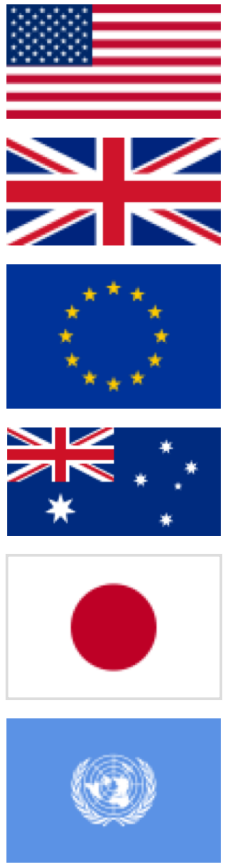


Community Modeling Frameworks Inventory

- Current technologies
  - Frameworks: ESMF, SPMF, etc.
  - Models: WRF, ROM, etc.
  - Earth System Curator
- Needed technologies
  - Ontologies for model component and framework functions
  - APIs for accessing domain ontologies
  - Capabilities to generate provenance and other key metadata
  - Use of data-type ontologies, units (point back to lower layer)
  - Semantics of assumptions, especially among models
  - Semantics to facilitate data assimilation

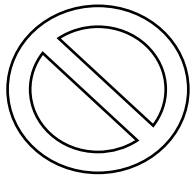


Keys to notations in gap analysis



Keys to notations in gap analysis

- Symbols:
  - Warning = gap



- Color scheme (unless otherwise noted)

Yellow - okay, or some effort, not proven
Orange - fair, definite gap, effort needed
Red - none or poor, serious gap, effort required

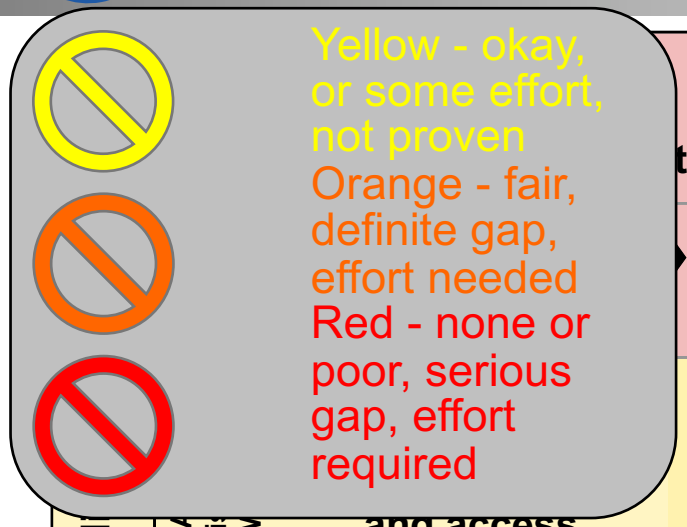
- Distance measure:
  - NASA =
  - Earth science =
  - Country = flag

- Examples:
  - outside US
  - in US, outside NASA and ES
  - in ES, outside NASA
  - within NASA and ES

Semantic Web Roadmap

		Current	Near Term (0-2 yrs)	Mid Term (2-5 yrs)	Long Term (5+ yrs)
Results	Outcome	♦ Improved Information Sharing	♦ Increased Collaboration & Interdisciplinary Science	♦ Acceleration of Knowledge Production	♦ Revolutionizing how science is done
	Output	♦ Geospatial semantic services established	♦ Geospatial semantic services proliferate	♦ Scientific semantic assisted services	♦ Autonomous inference of science results
Capability	Assisted Discovery & Mediation	♦ Some common vocabulary based product search and access	♦ Semantic geospatial search & inference, access	♦ Semantic agent-based searches	♦ Semantic agent-based integration
	Interoperable Information Infrastructure	♦ Local processing + data exchange	♦ Basic data tailoring services (data as service), verification/ validation	♦Interoperable geospatial services (analysis as service), results explanation service	♦ Metadata-driven data fusion (semantic service chaining), trust
Technology	Vocabulary	♦ SWEET core 1.0 based on GCMD/CF	♦ SWEET core 2.0 based on best practices decided from community	♦ SWEET 3.0 with semantic callable interfaces via standard programming languages	♦ Reasoners able to utilize SWEET 4.0
	Languages/ Reasoning	♦ RDF, OWL, OWL-S	♦ Geospatial reasoning, OWL-Time	♦ Numerical reasoning	♦ Scientific reasoning

Semantic Web Roadmap - Gap Analysis



Capabil	Dis… M…		♦ Increased Collaboration & terdisciplinary Science	♦ Acceleration of Knowledge Production	♦ Revolutionizing how science is done
			♦ Geospatial semantic services proliferate	♦ Scientific semantic assisted services	♦ Autonomous inference of science results
		and access	Sem…ntic geo…patial s…earch & infe…ence access	♦ S…mantic agent-ba…ed s…aches	♦ Se…antic …gent-bas…d int…ation
	Interoperable Information Infrastructure	♦ Local processing + data exchange	♦ B…sic data tai…ing s…rvices (d…ta a… s…rvice), …erific…ion/ validation	♦Int…perable geos…tial …ervices …nalysis as se…ice). …esults explanation service	♦ Met…data driven d…ta fusi…n (sem…antic …rvice chai…s), trust
Technology	Vocabulary	♦ SWEET core 1.0 based on GCMD/CF	♦ SWEET core 2.0 based on best practices decided from community	♦ SW…EET …3.0 with se…a…tic c…allable interf…ces …ia standard program…ing languages	Reas…ners …ble to …tilize SWE… 4.0
	Languages/ Reasoning	♦ RDF, OWL, OWL-S	♦ Ge…spatia… reasonin…, OW…Time	♦ …u…erical r…ason…n…	♦ …Sc…enti…c r…ason…ng

Current | Near Term (0-2 yrs) | Mid Term (2-5 yrs) | Long Term (5+ yrs)

Semantic Web Roadmap - Gap Analysis

Distance measure:
- NASA =
- Earth science =
- Country =

		Current	Near Term (0-2 yrs)	Mid Term (2-5 yrs)	Long Term (5+ yrs)
Capability	Ass… Disco… Med…	product search and access	…emantic …atial search & inference, access	♦ Semantic agent-based searches	♦ Semantic agent-based integration
	Interoperable Information Infrastructure	♦ Local processing + data exchange	♦ Basic data tailoring services (data as service), verification/ validation	♦Interoperable geospatial services (analysis as service), results explanation service	♦ Metadata-driven data fusion (semantic service chaining), trust
Technology	Vocabulary	♦ SWEET core 1.0 based on GCMD/CF	♦ SWEET core 2.0 based on best practices decided from community	♦ SWEET 3.0 with semantic callable interfaces via standard programming languages	♦ Reasoners able to utilize SWEET 4.0
	Languages/ Reasoning	♦ RDF, OWL, OWL-S	♦ Geospatial reasoning, OWL-Time	♦ Numerical reasoning	♦ Scientific reasoning
			…reased …oration & …inary Science	♦ Acceleration of Knowledge Production	♦ Revolutionizing how science is done
			…ial semantic proliferate	♦ Scientific semantic assisted services	♦ Autonomous inference of science results

Semantic Web Roadmap (expanded capability)

Capability	Current	Near Term (0-2 yrs)	Mid Term (2-5 yrs)	Long Term (5+ yrs)
Assisted Discovery & Mediation	♦ Some common vocabulary based product search and access	♦ Semantic geospatial search & inference, access	♦ Semantic agent-based searches	♦ Semantic agent-based integration
Assisted Knowledge Building	♦ Some metadata and limited provenance available	♦ Ontologies for data mining, visualization and analysis emerging/ maturing	♦ Common terminology captured in ontologies, crossing domains	♦ Provenance/ annotation with ontologies in user tools
Verifiable Information Quality	♦ Verification is manual with minimal tool support	♦ Ontologies for information quality developed	♦ Domain and range properties in ontologies used in tools	♦ Service ontologies carry quality provenance
Responsive Information Delivery	♦ Services must be hardwired and service agreements established	♦ Services annotated with resource descriptions	♦ Dynamic service discovery and mediation, and data scheduling	♦ Semantic markup of data latency (time lags) which adapt dynamically
Interoperable Information services	♦ Local processing + data exchange	♦Basic data tailoring services (data as service), verification /validation	♦ Interoperable geospatial services (analysis as service), results explanation service	♦ Metadata-driven data fusion (semantic service chaining), trust
Interactive Data Analysis	♦ Limited metadata passed to analysis applications	♦ Tag properties, non-jargon vocabulary for non-specialist use	♦ Shared terminology for the visual properties of interface objects and graph types...	♦ Semantic fields to describe tag key modal functions.
Seamless Data Access	♦ Access mediated by agreed standard vocabularies, hard-wired connections	♦ Access mediated by common ontologies	♦ Mediation aided by services with domain/ range properties	♦ Key data access services are semantically mediated

Roadmap - getting from near-term to mid-term

first priority, | second priority, | third priority

Capability	Near Term (0-2 yrs)		Mid Term (2-5 yrs)
Assisted Discovery & Mediation	♦ Semantic geospatial search & inference, access	-> requires agent development and vocabulary for agent characterization	♦ Semantic agent-based searches
Assisted Knowledge Building	♦ Ontologies for data mining, visualization and analysis emerging/ maturing	-> requires mature (domain and data-type) ontologies with community endorsement and governance and a robust integration framework	♦ Common terminology captured in ontologies, crossing domains
Verifiable Information Quality	♦ Ontologies for information quality developed	-> requires mature quality and uncertainty ontologies with domain and range properties added and populated	♦ Domain and range properties in ontologies used in tools
Responsive Information Delivery	♦ Services annotated with resource descriptions	-> requires semantic service (ontology) registry	♦ Dynamic service discovery and mediation, and data scheduling
Interoperable Information services	♦ Basic data tailoring services (data as service), verification/ validation	-> requires service to implement v/v, new descriptions of analyses, developing explanation	♦ Interoperable geospatial services (analysis as service), results explanation service
Interactive Data Analysis	♦ Tag properties, non-jargon vocabulary for non-specialist use	-> requires development of portal modal function vocabulary and ontology, link to domain context and data structure	♦ Shared terminology for the visual properties of interface objects and graph types...
Seamless Data Access	♦ Access mediated by common ontologies	-> requires adding properties to classes in ontologies and populating instances with expert agreement	♦ Mediation aided by services with domain/ range properties

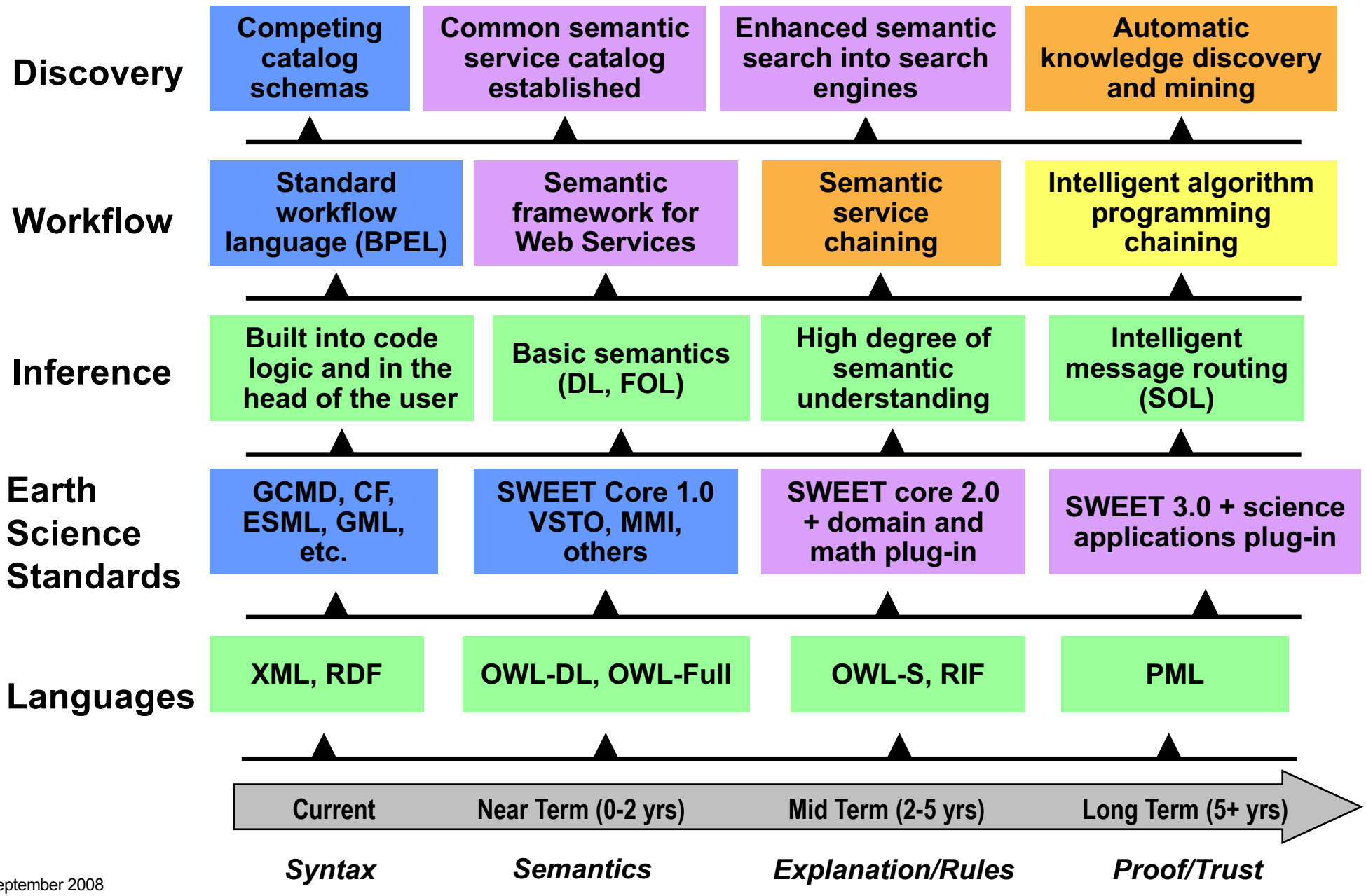
Roadmap - getting from near-term to mid-term

first priority, second priority, third priority

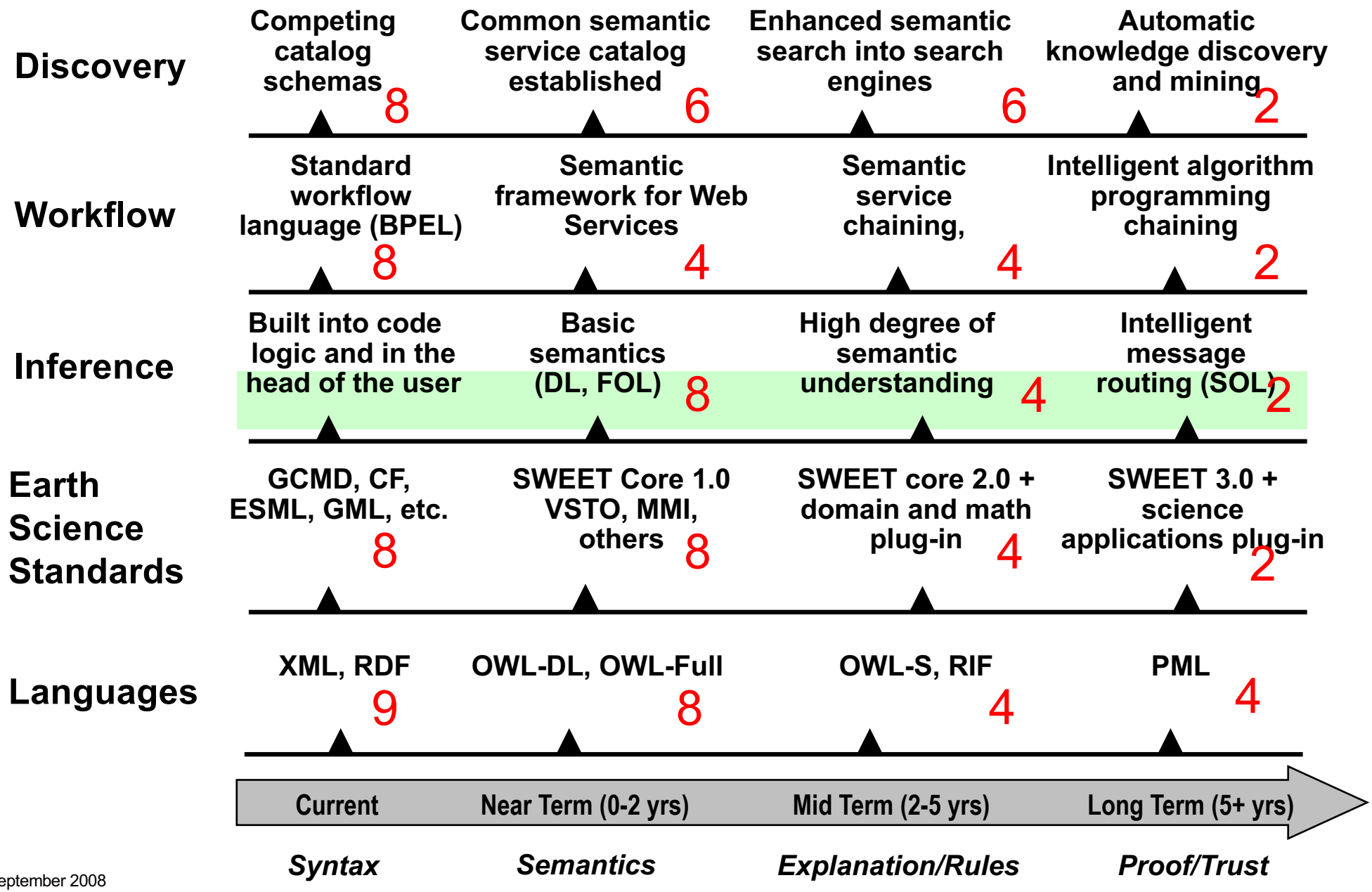
Capability	Near Term (0-2 yrs)		Mid Term (2-5 yrs)
Assisted Knowledge Building	♦ Ontologies for data mining, visualization and analysis emerging/ maturing	-> requires mature (domain and data-type) ontologies with community endorsement and governance and a robust integration framework	♦ Common terminology captured in ontologies, crossing domains
Responsive Information Delivery	♦ Services annotated with resource descriptions	-> requires semantic service (ontology) registry	♦ Dynamic service discovery and mediation, and data scheduling
Verifiable Information Quality	♦ Ontologies for information quality developed	-> requires mature quality and uncertainty ontologies with domain and range properties added and populated	♦ Domain and range properties in ontologies used in tools
Interoperable Information services	♦ Basic data tailoring services (data as service), verification/ validation	-> requires service to implement v/v, new descriptions of analyses, developing explanation	♦ Interoperable geospatial services (analysis as service), results explanation service
Seamless Data Access	♦ Access mediated by common ontologies	-> requires adding properties to classes in ontologies and populating instances with expert agreement	♦ Mediation aided by services with domain/ range properties
Interactive Data Analysis	♦ Tag properties, non-jargon vocabulary for non-specialist use	-> requires development of portal modal function vocabulary and ontology, link to domain context and data structure	♦ Shared terminology for the visual properties of interface objects and graph types...
Assisted Discovery & Mediation	♦ Semantic geospatial search & inference, access	-> requires agent development and vocabulary for agent characterization	♦ Semantic agent-based searches

Semantic Web: Roadmap Details

first priority, | second, | third, | done by others (comp. sci.), | in place



Semantic Web: Roadmap Details – TRL estimate



Technology Inventory for Semantic Web

- Languages
  - OWL - Web Ontology Language (W3C Recommendation), OWL 2 coming soon
  - RDF - Resource Description Framework (W3C Recommendation)
  - OWL-S/SWSL/SWSM/SAWSDL - Web Services (W3C Submission) – not standard (2018: ***)
  - SWRL - Semantic Web Rule Language (W3 Working Draft) – unlikely to be recommended
  - Rule Interchange Format (RIF) (W3 Submission)!!! (2018: W3 Recommendation)
  - PML - Proof Markup Language – defacto standard (2018: PROV superseded PML and OPM)
  - ODM/MOF - Ontology Definition Metamodel/Meta Object Facility (OMG) (2018: Gone)
- Editors: Protégé, SWOOP, Medius, SWeDE, CMAP/COE (2018: Most gone)
- Reasoners
  - Pellet, Racer, Medius KBS, FACT++, fuzzyDL, KAON2, MSPASS, QuOnto
- Query Languages
  - SPARQL (W3 Recommendation), XQUERY, SeRQL, OWL-QL, RDFQuery
- Other Tools for Semantic Web
  - Search: SWOOGLE swoogle.umbc.edu
  - Collaboration: www.planetont.org
  - Ontology repository/ registry: needed (three efforts now: BIOPORTAL, OMV and OOR)
  - Other: Jena, SeSAME/SAIL, Mulgara, Eclipse, KOWARI
  - Semantic wiki: OntoWiki, SemanticMediaWiki
  - Semantic content; Drupal,
  - Inference Web (IW; iw.rpi.edu)
- Semantic Standards for Earth Science
  - SWEET, VSTO, MMI, GeoSciML (all have governance models)
  - Need to promote domain and ES-specific service ontology development/ governance

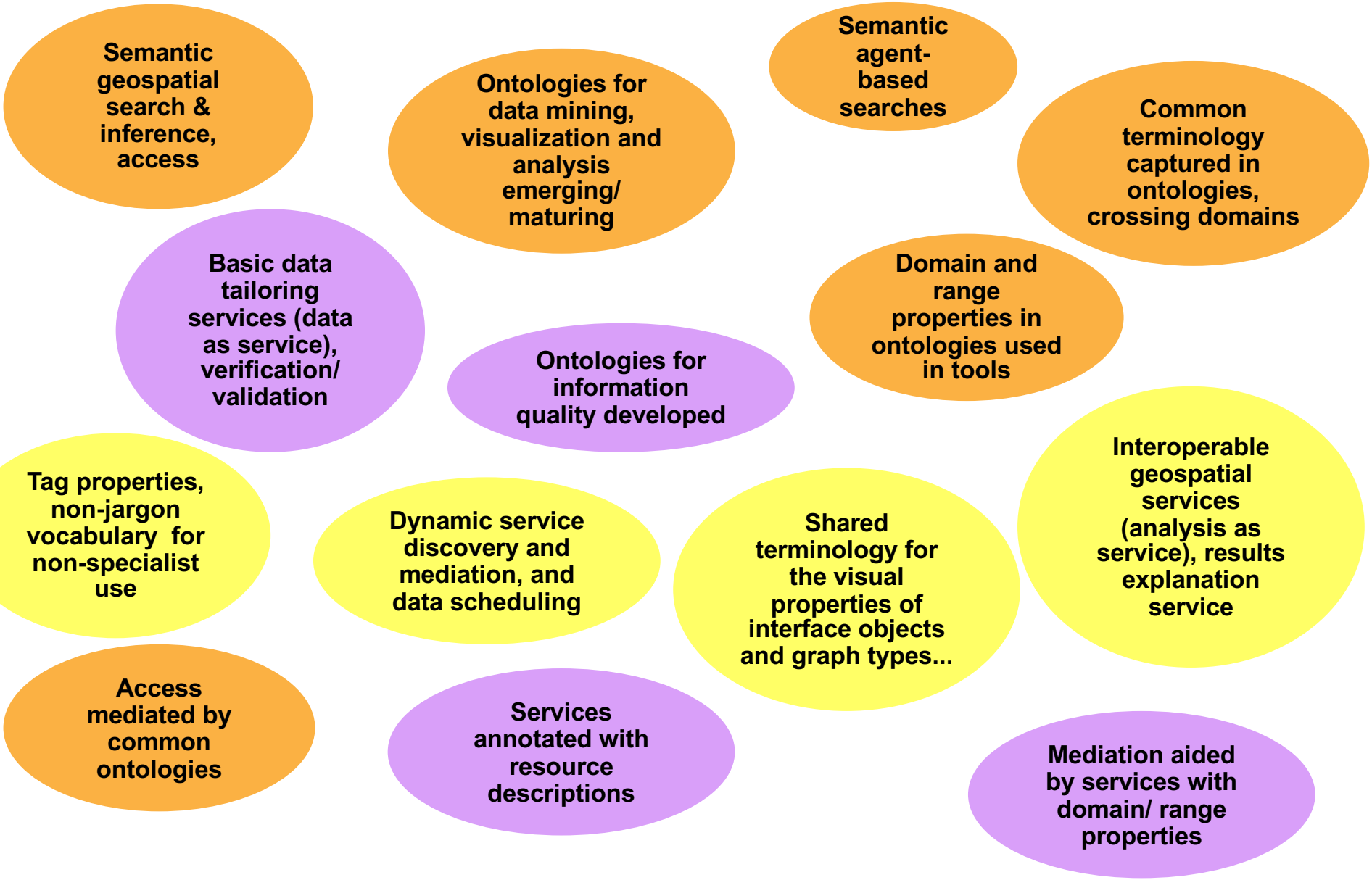
What NASA could fund now – priorities

first priority, second priority, third priority

- Semantic agent-based searches
- Common terminology captured in ontologies, crossing domains
- Semantic geospatial search & inference, access
- Ontologies for data mining, visualization and analysis emerging/ maturing
- Domain and range properties in ontologies used in tools
- Basic data tailoring services (data as service), verification/ validation
- Ontologies for information quality developed
- Interoperable geospatial services (analysis as service), results explanation service
- Tag properties, non-jargon vocabulary for non-specialist use
- Dynamic service discovery and mediation, and data scheduling
- Shared terminology for the visual properties of interface objects and graph types...
- Access mediated by common ontologies
- Services annotated with resource descriptions
- Mediation aided by services with domain/ range properties

What NASA could fund now – priorities

first priority, second priority, third priority



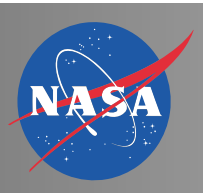
What NASA could fund now – priorities

first priority, second priority, third priority

- Basic data tailoring services (data as service), verification/ validation
- Ontologies for information quality developed
- Services annotated with resource descriptions
- Mediation aided by services with domain/ range properties

- Access mediated by common ontologies
- Ontologies for data mining, visualization and analysis emerging/ maturing
- Domain and range properties in ontologies used in tools
- Common terminology captured in ontologies, crossing domains
- Semantic geospatial search & inference, access
- Semantic agent-based searches

- Shared terminology for the visual properties of interface objects and graph types...
- Interoperable geospatial services (analysis as service), results explanation service
- Dynamic service discovery and mediation, and data scheduling
- Tag properties, non-jargon vocabulary for non-specialist use



Ongoing Steps

- Baseline metrics for current capabilities
  - TRLs, number of implementations, effort to add new data/ service, development time, new results
  - How to measure? TRL in/out, Technology Readiness Assessment, project-defined metrics, impacts/ nuggets
- Communicate findings to NASA
  - awareness of gaps - need to communicate where there are gaps and the implications for NASA
  - recommendations for solicitation wording that would encourage research to fill gaps
- Recommendations
  - NASA to ensure coverage in domain application areas of semantics
  - Identify ways to encourage technical/ vocabulary progress to support capability progress
  - Leverage non-NASA, non-ES, and non-US efforts