What is the Grid?

- In silico experiments
  - Information harvesting & PSE
  - Dynamically forming virtual organisations to solve problems.
  - Describing, searching for and weaving resources: people, applications, db, content, instruments
- Orchestrating resources
- Support for scientific method: provenance, argumentation, opinion contextualisation etc
Grid is metadata based middleware

Astronomy Sky Survey Data Grid

1. Portals and Workbenches
2. Knowledge & Resource Management
3. Metadata View, Data View, Catalog Analysis, Bulk Data Analysis
4. Grid Security, Caching, Replication, Backup, Scheduling
5. Information Discovery, Metadata delivery, Data Discovery, Data Delivery
6. Catalog Mediator, Data mediator
7. Compute Resources, Derived Collections, Catalogs, Data Archives
Grid is...
dynamic marshalling of resources

- Needs describing the resources, mapping between resources.
- Resources include BIG databases - the instances will not be on the web, they will be inside applications and databases (there isn’t much difference).
- Semantic web technologies for shared meaning (through ontologies) and shared models (e.g. exporting results through RDF and using inference over them).
Isn't information all computationally accessible already?

- Document publishing paradigm.
- Descriptive knowledge.
- Ontologies for controlling content already used.
- Evolving, non-predictive schemas
- XML is king.
Grid is ... services, services, services

- The first generation of Grid was protocol based.
- Second generation is service based: Open Grid Service Architecture.
- Semantic Web description and annotation technologies core to service sophisticated service description and processing.
- Descriptions => Automated discovery & search, selection, (imprecise) matching, composition & interoperation, invocation, execution monitoring
- Reasoning is darn handy
The Semantic Grid is ...

knowledge management

Q: What ATPase superfamily proteins are found in mouse?
1. P21958 (from Swiss-Prot).
2. InterPro is a pattern database and could tell you if you had permission and paid.
3. Attwood’s lab expertise is in nucleotide binding proteins
4. Jones published a new paper on this in Nature Genetics two weeks ago
5. Smith in your lab already asked this question...
Remarks

- Semantic Web is a part of the Grid vision?
- Semantic Web technologies should be relevant for Grid metadata at all levels.
- There isn’t one Grid, there are collections of Grids for communities - might be a more tractable model for the Semantic Web.
- Most facts will stay in databases. Metadata about the (scientific) process and facts could be in RDF.
- E-Science (everyone?) loves XML and ignores RDF
  - Annotations sit in other (non RDF) databases.
- Reliability, scalability, performance, explanation, longevity, evolution ...
The Road Ahead: Scientific Data Integration with the Semantic Web !?

From Bertram Ludäscher, SDSC
The Road Ahead: Scientific Data Integration with the Semantic Web !?
The Road Ahead: Scientific Data Integration with the Semantic Web !?

From Bertram Ludäscher, SDSC
The Road Ahead: Scientific Data Integration with the Semantic Web !?
The Road Ahead: Scientific Data Integration with the Semantic Web!

From Bertram Ludäscher, SDSC
The Road Ahead: Scientific Data Integration with the Semantic Web

From Bertram Ludäscher, SDSC
The Road Ahead: Scientific Data Integration with the Semantic Web!

Scientific Data

From Bertram Ludäscher, SDSC
The Road Ahead: Scientific Data Integration with the Semantic Web

Integrated Data Views

Scientific Data

RDB
ORDB
XMLDB

RDF
WSDL
DAML-S

DOOD rules
XQuery

DAML+S

DAML+OIL

XML

Logic

descrip

subs

infer

Internet2

Data-Grid

From Bertram Ludäscher, SDSC