WP 5.5 Time Series Data

Chris Perry, RAL

Astrogrid meeting, Cosener's House

17-18th September 2001

Background

- Handling of heterogeneous time series data
- Particular relevance to STP (in-situ multi-point spatial/temporal data)
- Data sets are relatively small ≤TBs
- Catalogues used mainly for data management

Background

- Complexity comes from need to search, extract, manipulate and combine the data themselves

- Uniform metadata crucial for manipulation (e.g. to allow transparent units conversion)

- International perspective important (access to non-UK data sets critical)

Background

- A scientist wishing to study the propagation and effect of a Coronal Mass Ejection might use:-
  - The coronagraph on SOHO
  - Upstream solar wind measurements from ACE
  - Cluster plasma and field measurements near the magnetopause
  - Plasma composition measurements in the mid altitude cusp
  - Ring current enhancements, in-situ, remote sampling and ground based geomagnetic indices
  - Position and timing information

- Data have different locations, query specifications and are returned in different formats

Background

- Current systems typically consist of:-
  - Pre-generated survey plots for rapid browsing
  - Dynamic links to provide access to survey plots of other data
  - Access to actual data limited to simple searches (data set and time), download & quicklook of local holdings
  - No multi-archive search
  - No search on data content
  - No facility for joining and manipulating data

Objectives

- Identify and evaluate implementation options for the efficient query, manipulation and delivery of heterogeneous time series data.
- Implement a test-bed system to assess the ease of integration of heterogeneous archives.
- Implement a mini-web interface to demonstrate capabilities of system to non-Grid experts.
- Evaluate test-bed system and lessons learned during its implementation.

Inputs

- Requirements and use-cases from WP-A1.
- Existing Grid middleware libraries, support software, standards and expertise including outputs from WP-A2, A3 and A4. Inputs from other national and international e-science/Grid activities.
- Heterogeneous data archives (UK Cluster Data Centre and WDC/EISCAT) for federation into the test-bed system.
- Resources - 0.5 s/y

Task 5.5.1

- Develop a metadata translation layer
  - Astrogrid interface expected to be XML based
    - WP A2 to maintain DB of tag specifications?
  - Need to supply STP tags for Astrogrid schema
    - Already some naming standards from ISTP
  - Will need to convert archive specific metadata to the Astrogrid format
  - What to do with time varying metadata?

Task 5.5.2

- Develop a data export layer
  - Translate between archive and compatible Astrogrid formats
    - NASA CDF widely used in STP.
    - Other formats?
  - May use DG's STPDF toolkit
  - Need to support sub-sampling, averaging and possibly streaming of the data
  - Test-bed will only support tabular data

Task 5.5.3

- Implement a simple query layer
  - Support distributed queries and joins
  - Query on data (as well as metadata)
  - Intelligent use of metadata

Task 5.5.4

- Implement an authorisation layer
  - Authentication handled centrally
  - Authorisation usually handled by the archive
  - Need to provide a layer to:-
    - Translate Astrogrid authentication to archive
    - Extract archive data authorisation
    - Translate archive authorisation to Astrogrid

Task 5.5.5

- Integrate system with Grid middleware
  - Interface the archive specific interface layers with the Grid infrastructure
  - Federate the test-bed archives

Task 5.5.6

- Develop simple web-based UI
  - Provide access to federated data sets
  - Simple access so non-expert users can test and provide feedback on the system
  - Quicklook graphics important for PR
  - If time permits support mathematical and domain specific data manipulation

Task 5.5.7

- Evaluate test-bed implementation
  - Gather inputs from developers and users
  - Produce technical note on the lessons learnt
  - Provide inputs to the Phase-B plan