Lessons from Earth Observation

Peter Allan
Rutherford Appleton Laboratory

Other Disciplines

- Other disciplines have similar problems
  - They may have solved them already
- We need to work together to get the most out of GRIDs
- There will be different GRIDs, but we should try to minimize diversity
  - There is only one http

What can we learn?

- Searching for data across distributed systems
  - INFEO
  - NERC metadata project
- Distributed data processing systems
  - GERB ground segment
- "Pre-GRID" systems

What can we offer?

- Data models used in astronomy have wider applicability
- Experience of processing, storage, and access to large and diverse datasets

Distributed Searches

- Based on ISO Z39.50 protocol
- Developed in bibliographic community
- User domains define different 'profiles'
  - First was bib-1
- Defines search criteria in a general way
- Includes concepts like 'contained within'

Z39.50

- What is it?
  - "Z39.50 is a development intended to facilitate networked access to distributed database resources. Basically, Z39.50 itself is no more than a network protocol. The rest of the work is done by your database software or some other tool."

Z39.50 profiles

- Bib-1
- CIP
- GEO
  - Intersection of CIP and GEO
- GILS
- ENRM

CIP

- Catalogue Interoperability Protocol
  - Is a Z39.50 profile
  - Developed by CEOS (Committee for Earth Observation Satellites)
    - So contains EO concepts
  - Practical implementation by CEO (Centre for Earth Observation) under EC Framework IV programme

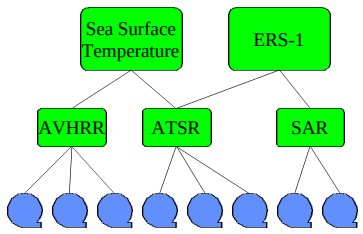
CIP

- Defines about 300 items of metadata in a hierarchical manner
  - Boundary on Earth's surface
    - North, South, East, West bounding rectangle
- Defines capabilities such as search for data and ordering data
  - Ordering may imply a payment

CIP concepts

- Terminal items
  - e.g. a file - something that can be ordered
- Collections
  - of terminal items or of other collections
  - Used to aid navigation around system

Collections



INFEO

- Designed as a distributed system
  - Connected to several data centres
- Initial implementation
  - one "middleware node"
  - several catalogue providers with gateways
- Intend to have to several middleware nodes
  - UK retrieval manager being installed at RAL

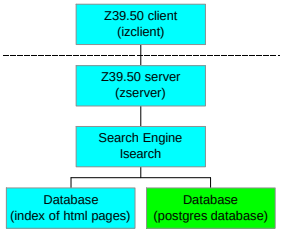
INFEO

- The INFEO system allows you to search all the connected catalogues
  - Meteosat, DLR, ...
  - Environmental, mapping, ...
  - etc
- INFEO provides other facilities
  - announcements, FAQs

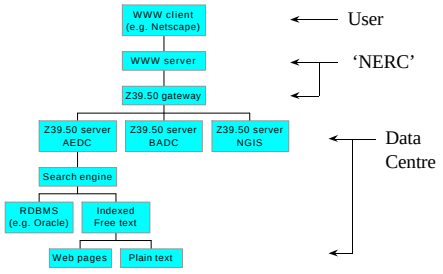
NERC Metadata Project

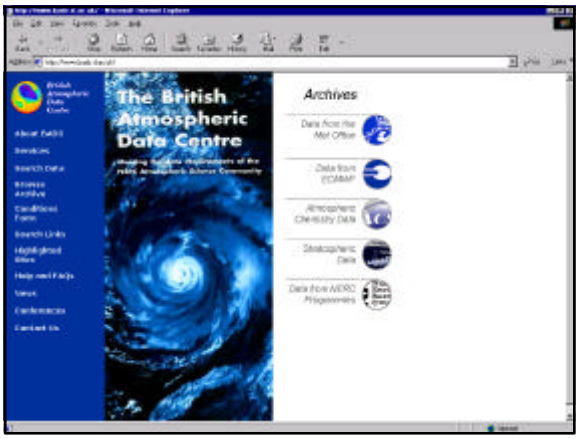
- Purpose:
  - To provide an on-line catalogue of NERC's data holdings at the 7 Designated Data Centres
- Had to handle a wide range of metadata
  - e.g latitude and longitude vs. national grid
- Initially very simple
  - What, when, where

Implemented with Isite



System Architecture





Isite is easy

- Software is freely available
- Can be installed on a Linux system
- I set up an Isite server, database, web gateway and client on an old 486 at home

- Isite can index several types of files
  - Plain text
  - SMGL
    - Most useful way of describing metadata
- Develop XML interface?

Software

- INFEO software is available from CEO
  - Developed with EC funds
- NMP uses Isite software
  - Developed by FGDC ‘clearinghouse’ in USA
  - Currently supported by NASA
  - Small and easy to use
- Other (commercial) software is available

Relevance to Other Areas

- Experience of defining metadata is widely applicable to other areas
- Software for distributed searches is available and can be used directly or modified

Recommendations

- Investigate further how Z39.50 and related tools can benefit astronomy needs
- Investigate incorporation of XML into Isite
- Look at recommendations on metadata from CEO

Distributed Data Processing

- GERB
  - Geostationary Earth Radiation Budget
- Data received in Germany
- Processed
  - RAL -> Belgium -> RAL