

Introduction to GIS and Grid Computing

Andy Turner

MoSeS

Outline

- What is GIS?
- What is special about Geospatial Information?
- UK e-Science Programme
- Grids and emerging e-Infrastructure
- Why GIS and GRID?

What is GIS?

- Originally
 - Geographic Information System
 - Maps in computers
 - Main concern 2D space + attributes
 - Developed mainly since 1980
- Key is that the data is spatially referenced to locations on Earth's surface
- Modern concern 3D space + time + attributes
 - Geospatial

GIS and Grid Computing Agenda Setting Workshop

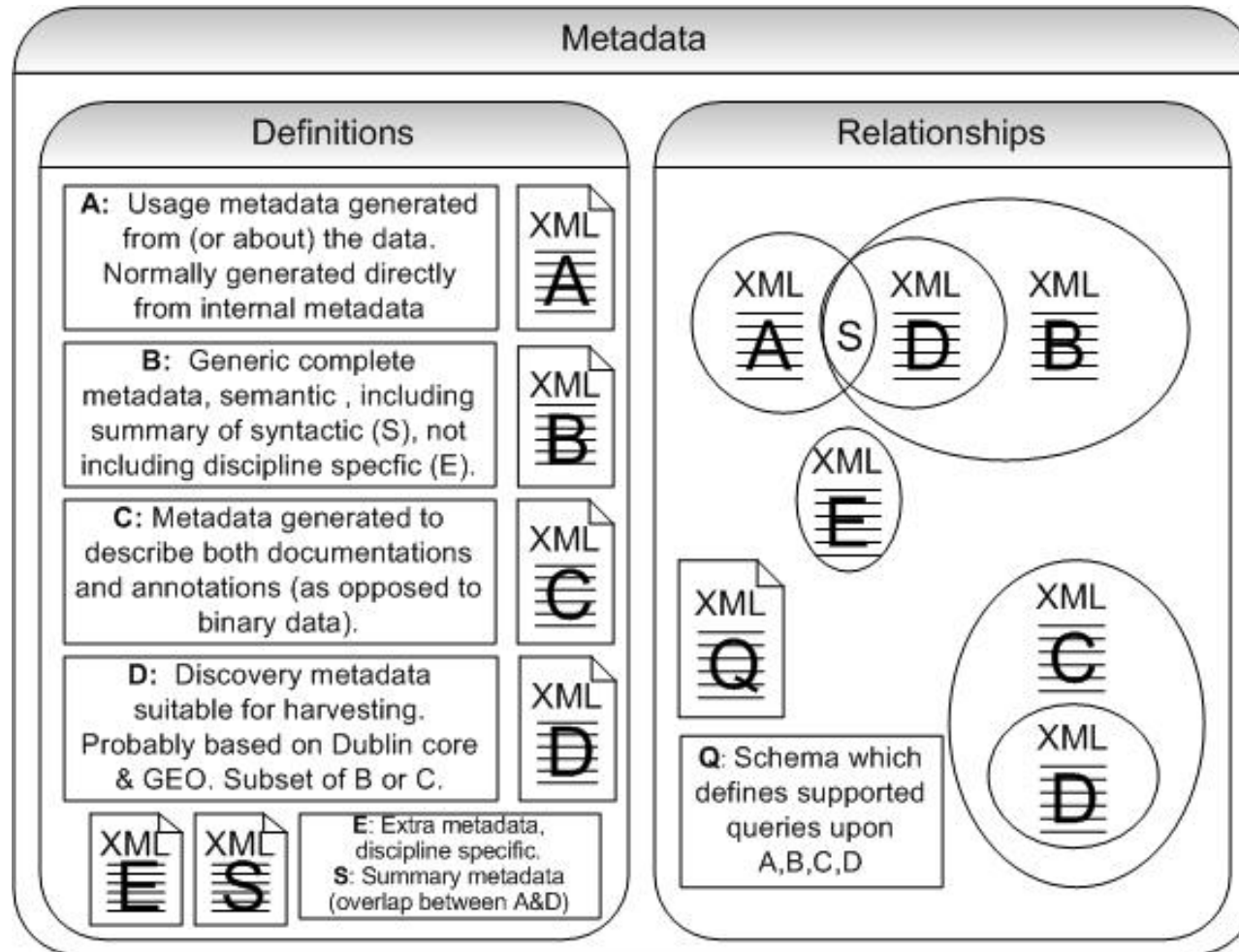


A system of hardware, software, data, people, organizations and institutional arrangements for collecting, storing, analyzing, and disseminating information about areas of the earth (Dueker and Kjerne, 1989)

Types of Data

- Coverages
 - Raster (regularly spaced)
 - GeoTiff, netCDF
 - Vector (points, lines, regions)
 - Attributes attached
 - Features
 - Multiple geometries, multiple attributes
 - Can be grouped into Themes
 - GML
- Many different sources
 - Remote sensing
 - Digital maps
 - Census and surveys
- Much data is derived and much does not depict physical objects

GIS and Grid Computing Agenda Setting Workshop



GI Processing

- Data cleaning
 - Readying data for analysis or integration with data from other sources
 - Projecting data into a common spatial framework
- Exploring relationships
 - Scales
 - Resolutions
- Involves a series of (approximately 40) basic processes formed in fairly ad hoc ways
 - Process chains
- Interactive
- Procedural
- Raster calculations are easily parallelised
- Computationally intensive?

GIS application development is addressing generic problems

- Spatial and temporal semantics
- How to integrate information (from different sources) using spatial and temporal location when this is uncertain
- How to store, process and provide access to large volumes of information across networks
- How to take advantage of the data that exists to provide better solutions?
- How to improve data collection, organization and access for research

Internet GIS

- A new distributed processing technology
- Web Map Servers (WMS)
- Web Feature Servers (WFS)
- Web Coverage Servers (WCS)
- Aided by standards
 - Data transfer formats
 - Data access methods
 - Organisations
 - Open Geospatial Consortium (OGC)
 - ISO TC/211
 - W3C
- Google Earth

GIS and Grid Computing Agenda Setting Workshop



- What is meant by e-Science? In the future, e-Science will refer to the **large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet**. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists.
- The Grid is an architecture proposed to bring all these issues together and make a reality of such a vision for e-Science. Ian Foster and Carl Kesselman, inventors of the Globus approach to the Grid define the Grid as an enabler for Virtual Organisations: **‘An infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources.’** It is important to recognize that resource in this context includes computational systems and data storage and specialized experimental facilities.



- 10 Regional Centres
- 7 Centres of Excellence
 - WRG
- 7 other centres
 - NCeSS
 - National Institute for Environmental e-Science
 -



National Grid Service

core production computational and data grid

What are Grids?

- Provide computational resources and a distributed computing infrastructure that supports flexible, secure, coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organisations” (Ghimire et. al., 2005)
- Grids: build “virtual computers” from resources available to collaborations that cross administrative domains (“virtual organisations”) (Mineter, 2005)
- “e-infrastructure” = networks + grids + supporting (people) services
 - In UK, the National Grid Service and Grid Operations Support Centre
 - Europe-wide, EGEE (Mineter, 2005)
- The Open Grid Services Architecture (OGSA) represents the evolution towards a Grid system architecture based on Web services concepts and technologies. (Ghimire et. al., 2005)

Sources:

- Mike Mineter (2005) Future Cities Workshop Presentation at the First International Conference on e-Social Science, September, Manchester
- Ghimire D.R, Simonis I, Wytzisk A (2005) Integration of GRID approaches into the Geographic Web Service Domain

Some Grid Components

- Hardware
 - Computational and operational machines
 - Personal Computers
 - Sensing/control equipment
 - HPC
- Authentication and Authorisation
 - Certification
 - Allowance
 - Access to data
 - Access to computational resource
- Virtual File System
 - For storing data
- Process management

Web Service from Grid perspective

- A Web Service is programmable application logic accessible using standard Internet protocols
- Web service is any piece of code that can be made available over the Internet, so that other applications can invoke it and utilize its functionality
- Web Services Description Language (WSDL) is an XML-based language for describing Web services and how to access them.
- Emerging XML and Web Service specifications is a big bag of worms

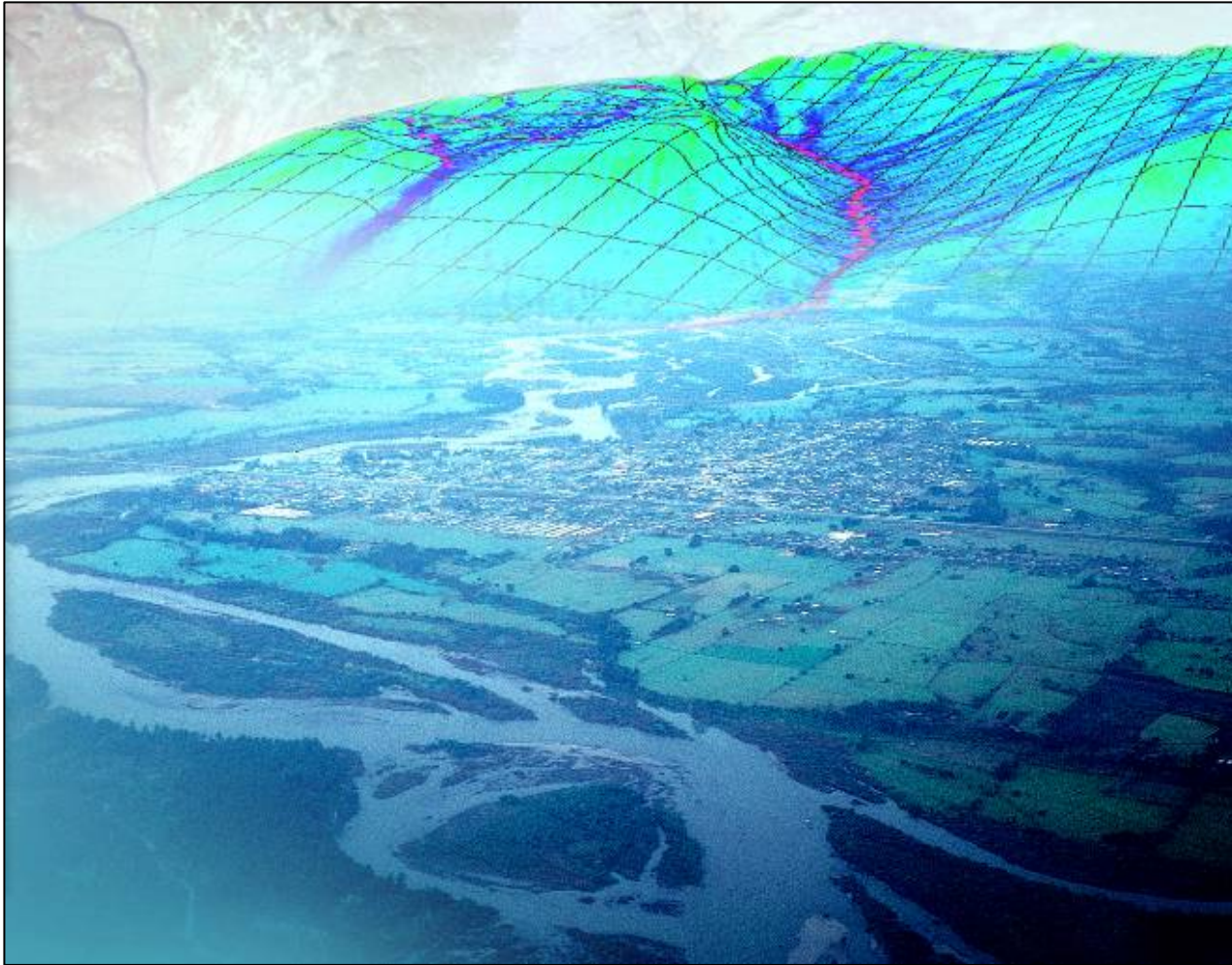
Web Services Resource Framework (WSRF)

- Allow Web and Grid communities to move forward on a common base
- separates '**service**' and '**stateful resource**' upon which service acts
- WS-Resource:
 - identity (WS-Addressing)
 - lifetime (WS-ResourceLifetime)
 - state/properties (WS-ResourceProperties)
- publish-subscribe mechanism for state changes (WS-Notification)

Why GIS and Grids?

- As a user the GI industry is well placed to demonstrate the potential of GRID
 - A mature web-services culture
 - Functioning, widely used, reasonably OK interoperability data standards
 - A process orientated environment
 - A set of problems that need addressing
 - An IT literate user base in academia and industry
 - The sexiest data

GIS and Grid Computing Agenda Setting Workshop



Source: Phil James Geo-Processing Frameworks Issues and Opportunities Presentation at the Grid Middleware and GeoSpatial Standards Workshop, Edinburgh, September 2005

Some Problems/Challenges

- Range of application domains, lots of software providers and many institutions involved
 - Agreement on standards
 - E.g. What do we expect from a projection service?
 - OGC
- Data quality
 - Need for good meta data
 - Need for standard formats
- Developing Spatial Data Infrastructures (SDIs)
 - Need for education
- Security and DRM issues
- How do OGC 'web services' map into mainstream Web Services?

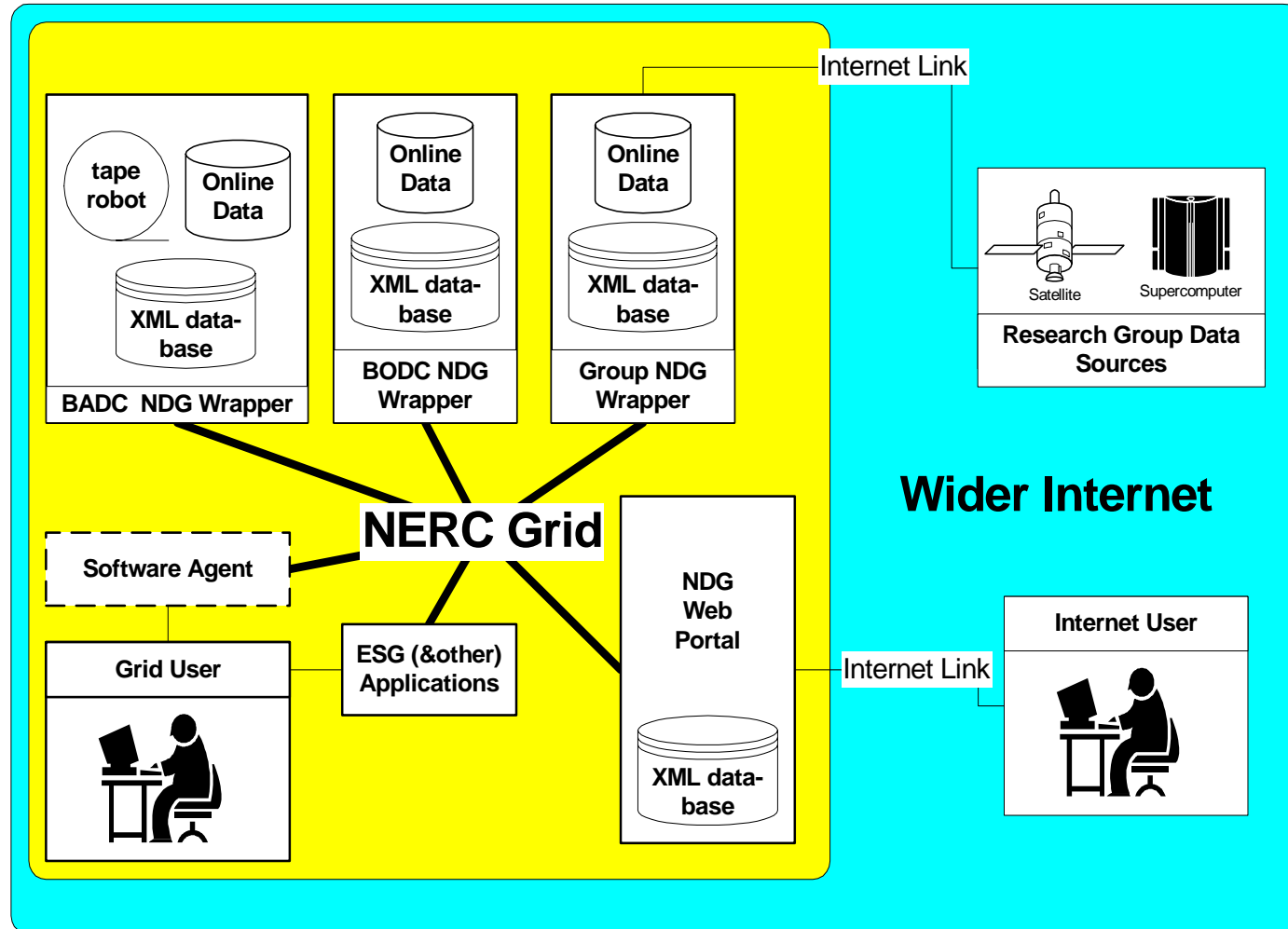
Important Organisations

- Open Geospatial Consortium
- ISO
- W3C
- National Institute for Environmental e-Science
- Open source community

Summary

- GIS is maturing to support Earth Systems Science in 5D
- GIS has a mature web services culture
- GIS can benefit from Grid technology as good GIS software in some ways provides a Grid environment
- GIS has much to offer Grid computing and vice versa
- There can be confusion of terminology when experts in GIS and Grid computing meet
- Spatial Data Infrastructure is key to supporting many of the applications we want
- Work is ongoing at national and international levels?
- What are the special needs in terms of Grid GIS for e-Social Science?

GIS and Grid Computing Agenda Setting Workshop



Source

Bryan Lawrence Presentation on NERC Data Grid at Workshop on Grid Middleware and Geospatial Standards for Earth System Science Data, NeSC, Edinburgh, September 2005

G

- Geographical, Geographic, Geospatial, Grid

I

- Information, Informatics, Integration, Intelligence

S

- System, Science, Service

Acknowledgements

- Some of these slides have used images ripped from the web