Integration and visualisation of geospatial data through a common data model and standardised services

Stephen Pascoe (S.Pascoe@rl.ac.uk), A. Woolf (A.Woolf@rl.ac.uk), A. Shaon (A.Shaon@rl.ac.uk), M. Kochan (Martin.Kochan@tessella.com), D. Lowe (D.Lowe@rl.ac.uk), Nicolas Bertrand (nsb@ceh.ac.uk), Quillon Harpham (q.harpham@hrwallingford.co.uk)

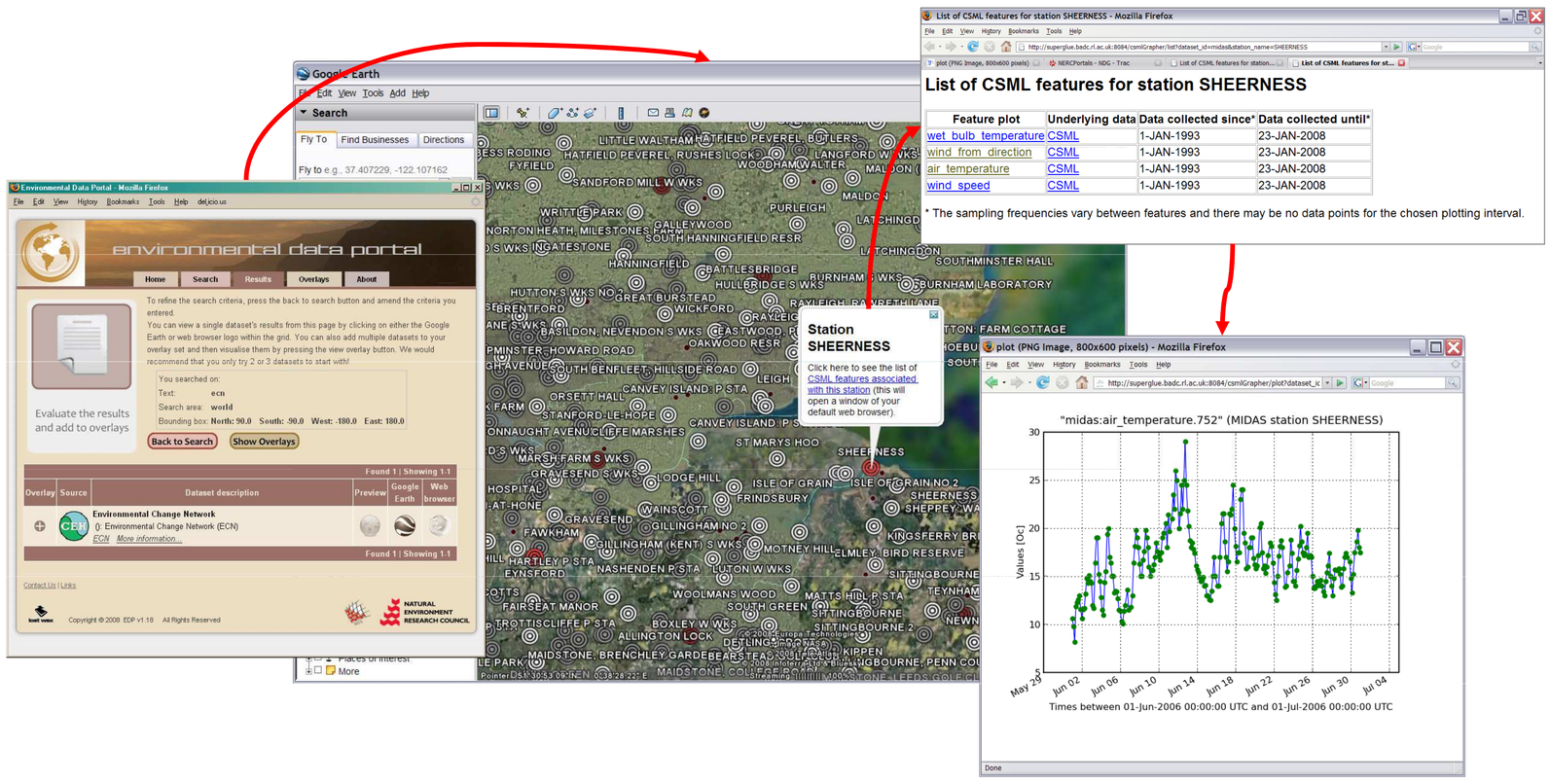
In the environmental data arena, datasets are often held in formats or database schemas specific to each dataset. Developing integrated spatial information services, such as those required by the European INSPIRE directive, will require these data to be represented through a common data model. One such model is provided by the Climate Science Modelling Language (CSML). Here we present a method of efficiently representing environmental data from custom database schemas in CSML and exposing them through the Open Geospatial Consortium (OGC) web services Web Feature Service (WFS) and Web Map Service (WMS). This enables visualisation of two different environmental datasets via GoogleEarth® and a web browser.

NERC Portals Project

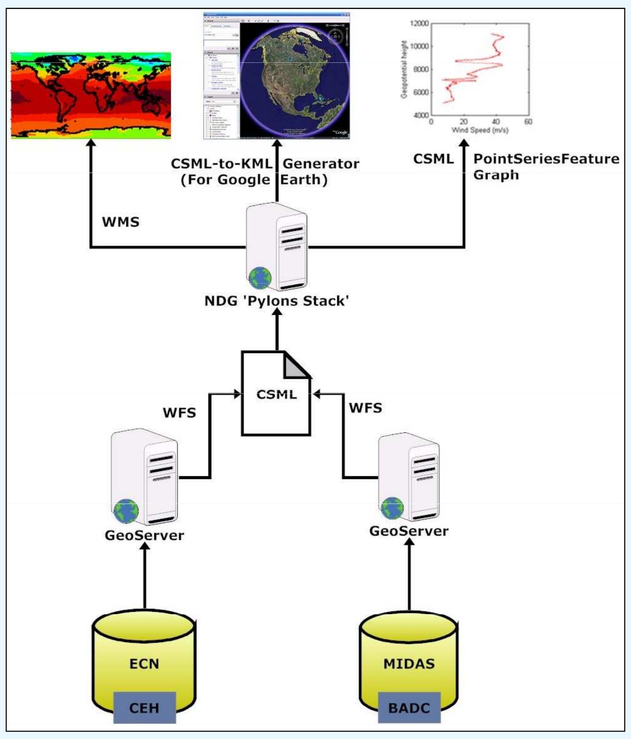
This work is part of the NERC Portals project to demonstrate how data can be visualised and overlaid using a truly distributed environment.

Metadata (data about data and the service available for the data) are stored centrally but the actual data are stored at the participating organisations.

Once discovered, the data can all be displayed in either Google Earth and/or through a map-based web browser. Links to the data holders are shown.



Architecture



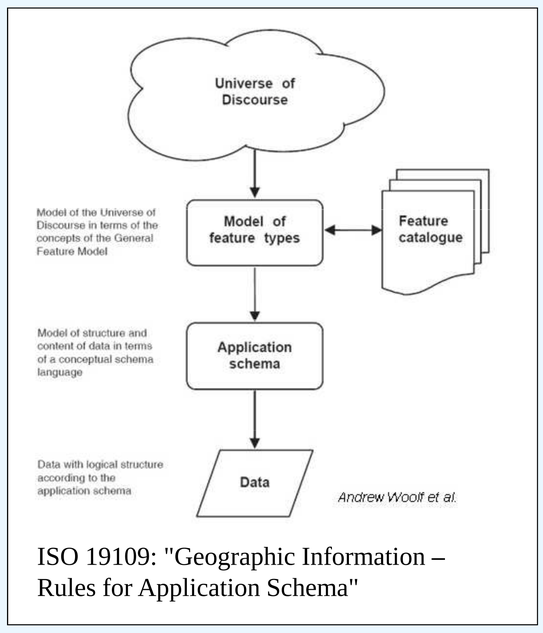
The architecture of this work is primarily comprised of two instances of GeoServer deployed as the OGC compliant Web Feature Services (WFS) over two contrasting datasets residing at the Centre for Ecology & Hydrology (CEH), Lancaster and the British Atmospheric Data Centre (BADC) respectively. For this project, we have used the "Community Schema" version of GeoServer, which enables representing data from a relational database in an application schema (e.g. CSML) that is defined independently of the underlying database structure. Both GeoServers have been configured to produce a list of data collecting stations for their respective datasets in GML format, and the data associated with each station as CSML PointSeriesFeature. In addition, we have implemented the three following methods of visualisation of the data served up by two GeoServers:

- An OGC compliant Web Map Service (WMS), which enables visualisation of CSML data using any standard WMS client.
- A tool for transforming CSML data into KML format (called CSML-to-KML generator), which can be used to visualise CSML data in Google Earth.
- A graphing tool for presenting CSML PointSeriesFeature data as time-value graphs.

All of these data visualisation methods have been implemented in Python programming language and deployed on the NERC Data Grid's Pylons OGC Services Stack, which is an integrated web application framework, also written in Python.

Glossary

- **GeoServer**: an Open Source web server that publishes Geospatial information on the web using open standards, such as the OGC standards.
- **INSPIRE:** The European Union directive establishing an infrastructure for spatial information.
- **OGC**: The Open Geospatial Consortium is a international standards organisation developing standards for geospatial services including.
  - **GML:** Geography Markup Language is an XML grammar written in XML Schema for the description of application schemas as well as the transport and storage of geographic information.
  - **WMS:** Web Map Service for serving raster images of maps.
  - **WFS:** Web Feature Service for retrieving and updating geospatial data encoded in GML.
- **Pylons:** A web application framework for the Python programming language.



ISO 19109: "Geographic Information – Rules for Application Schema"

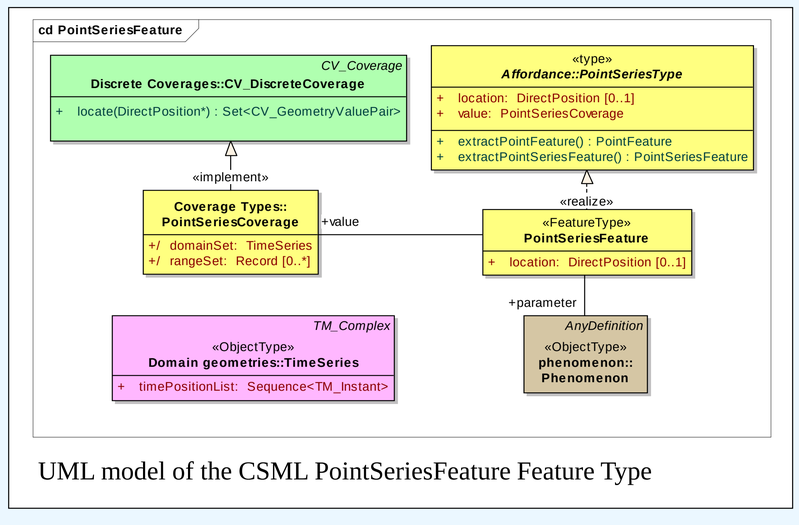
CSML

The Climate Science Modeling Language is a GML application schema that provides a standards-based semantic model and encoding for representing a range of conceptual information classes of relevance to climate science. These classes may be employed to build intelligent services for data subsetting, aggregation, processing, etc. As well, CSML provides a "wrapper" mechanism to encapsulate legacy file-based data, exposing them instead through the conceptual view.

In this work we use CSML to integrate heterogeneous datasets from two different data providers. The datasets are conceptually similar (observation data) so can be harmonised at the conceptual modelling level using CSML feature types and harmonised at the implementation level using the CSML Application Schema and software.

CSML is

- Based on the TC111 ISO standards for Geographic Information
- A Conceptual Model defining Climate Science Feature Types
- A Geography Markup Language (GML) Application Schema
- Software and tooling



UML model of the CSML PointSeriesFeature Feature Type

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

Science & Technology Facilities Council
e-Science

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

HR Wallingford
Working with water

Tessella
Scientific software solutions