CSML: Confronting the use of GIS for met data from an OGC feature type perspective

Bryan Lawrence[1]

NERC Centres for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Andrew Woolf[2], Roy Lowry[3], Kerstin Kleese van Dam[2], Ray Cramer[3], Marta Gutierrez[1], Siva Kondapalli[3], Susan Latham[1], Dominic Lowe[1], Kevin O'Neill[2], Ag Stephens[1]

[1]British Atmospheric Data Centre
[2]CCLRC e-Science Centre
[3]British Oceanographic Data Centre

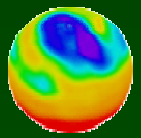
BADC, BODC, CCLRC, PML and SOC

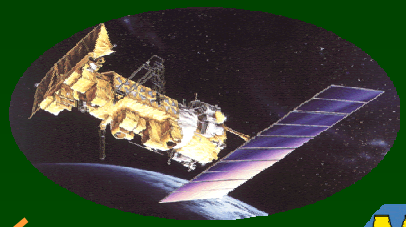
Outline

- Motivation
  - The NERC Datagrid
- Standards
  - OGC Vision
  - Feature Types
- CSML
  - Description
  - Prototyping in MarineXML
  - Round-Tripping
- Feature Type Issues
- Next Steps

Complexity + Volume + Remote Access = Grid Challenge

British Atmospheric Data Centre





NERC DataGrid

PCMDI

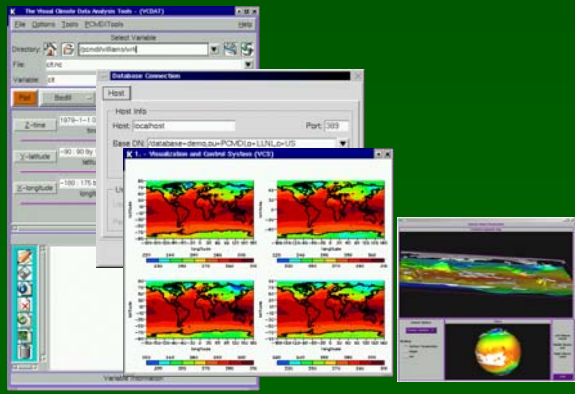
GO-ESSP

M&D Modelle & Daten

PCMDI

NCAR

Solid Earth and Environment GRID



BODC

British Oceanographic Data Centre

http://ndg.nerc.ac.uk

Integration – semantics

- Want semantic access to information, not abstract data
  - getData(***potential temperature*** from ***ERA-40*** dataset in ***North Atlantic*** from ***1990 to 2000***)
  - not: getData("era40.nc", 'PTMP', 20:50, 300:340, 190:200)
  - or even worse:

```
for j=1990:2000
    getData("era40_"+j+".nc", 'PTMP', 20:50,
      300:340)
```

- Lossy is OK!
  - Care less about completeness of representation than semantic unification

OGC Web Services

- Web Map Server
- Web Coverage Server
- Web Feature Server

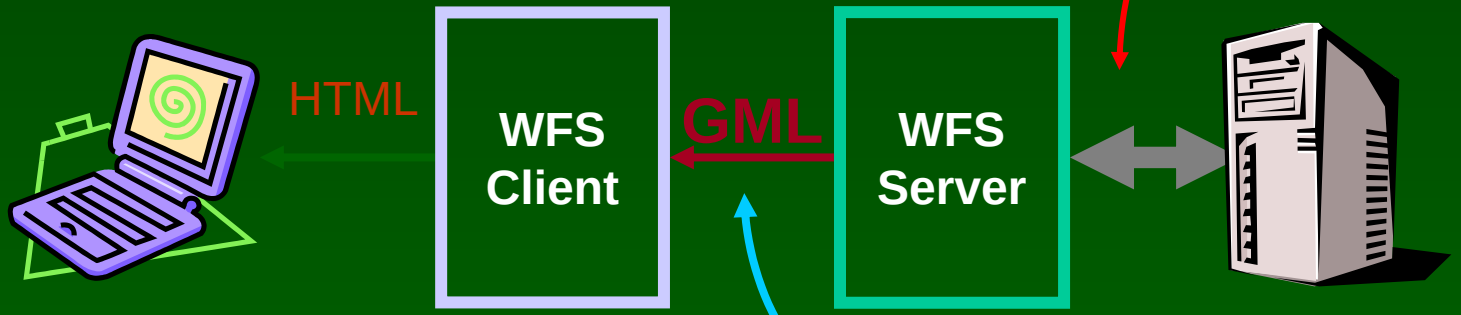
- Web Processing Server
- … and more.

Feature: something you can name,

Coverage, something you have numbers for over some range (A coverage IS a feature!)

OGC Web Feature Service

public ⇔ private boundary

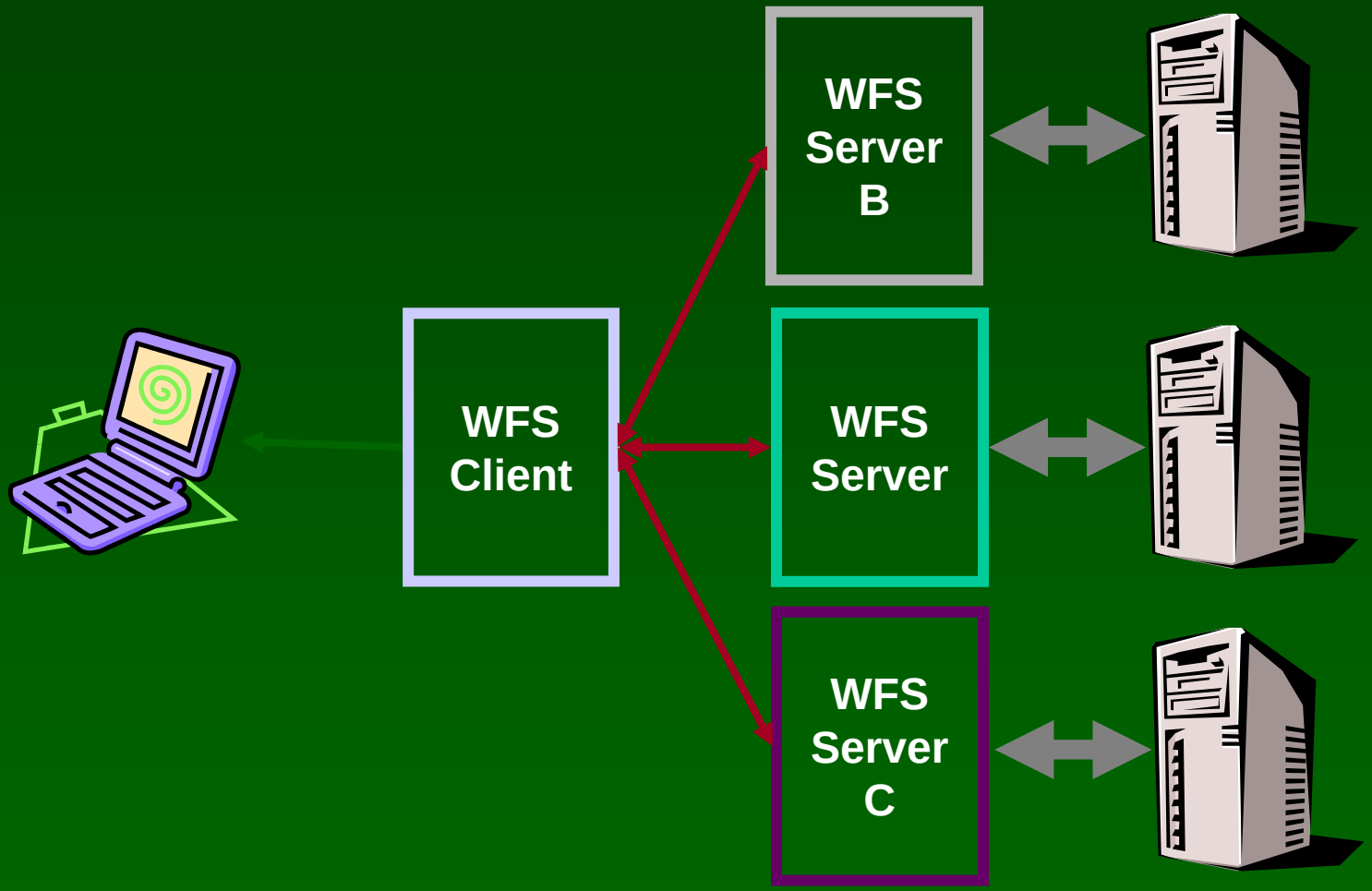


Data-source organised for custodian's requirements

Community-specific **GML application** language

- TigerGML, LandGML, O&M, XMML, CGI-GML, ADX, GPML, CSML etc

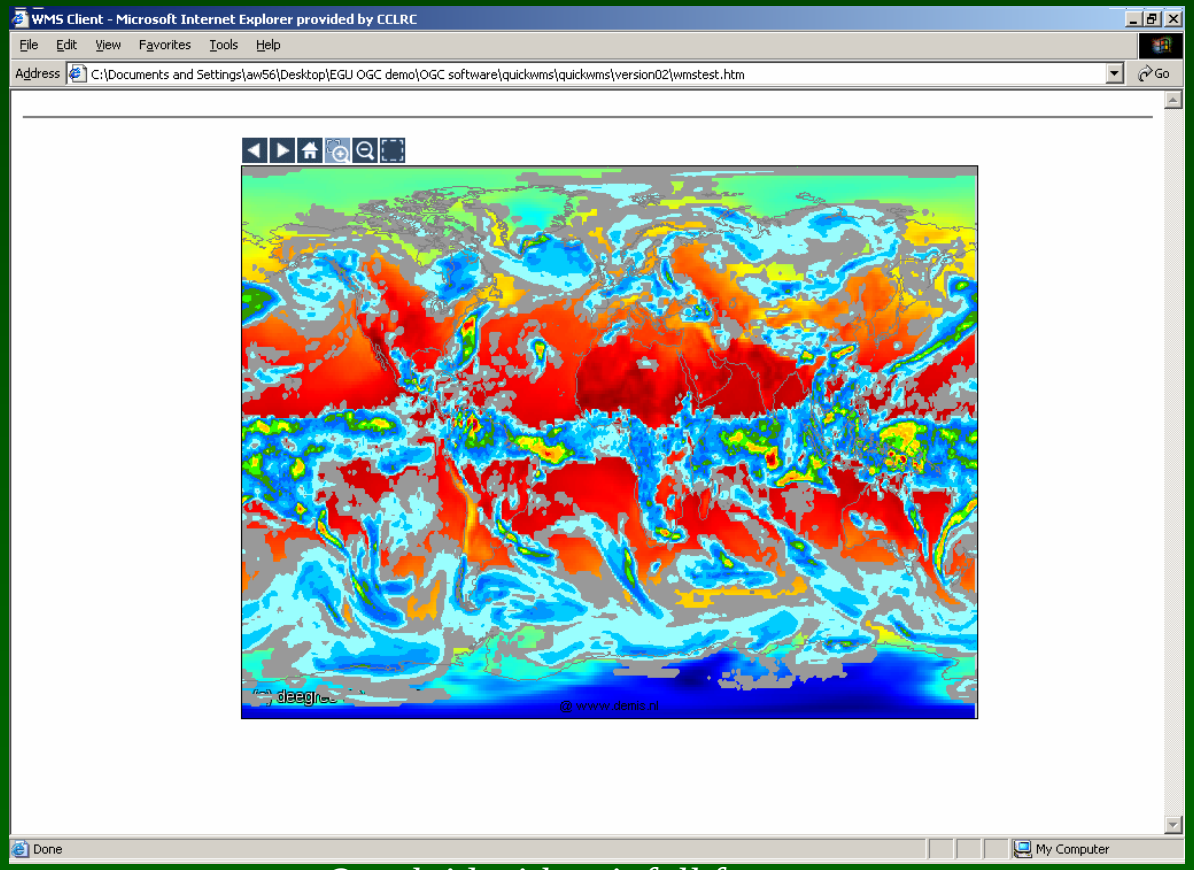
Standard transfer format allows multiple data sources



NetCDF + WMS

e.g.: ERA40 re-analysis surface air temperature, 2001-04-27

- deegree open-source WMS modified with netCDF connector



Overlaid with rainfall from globe.digitalearth.gov WMS server

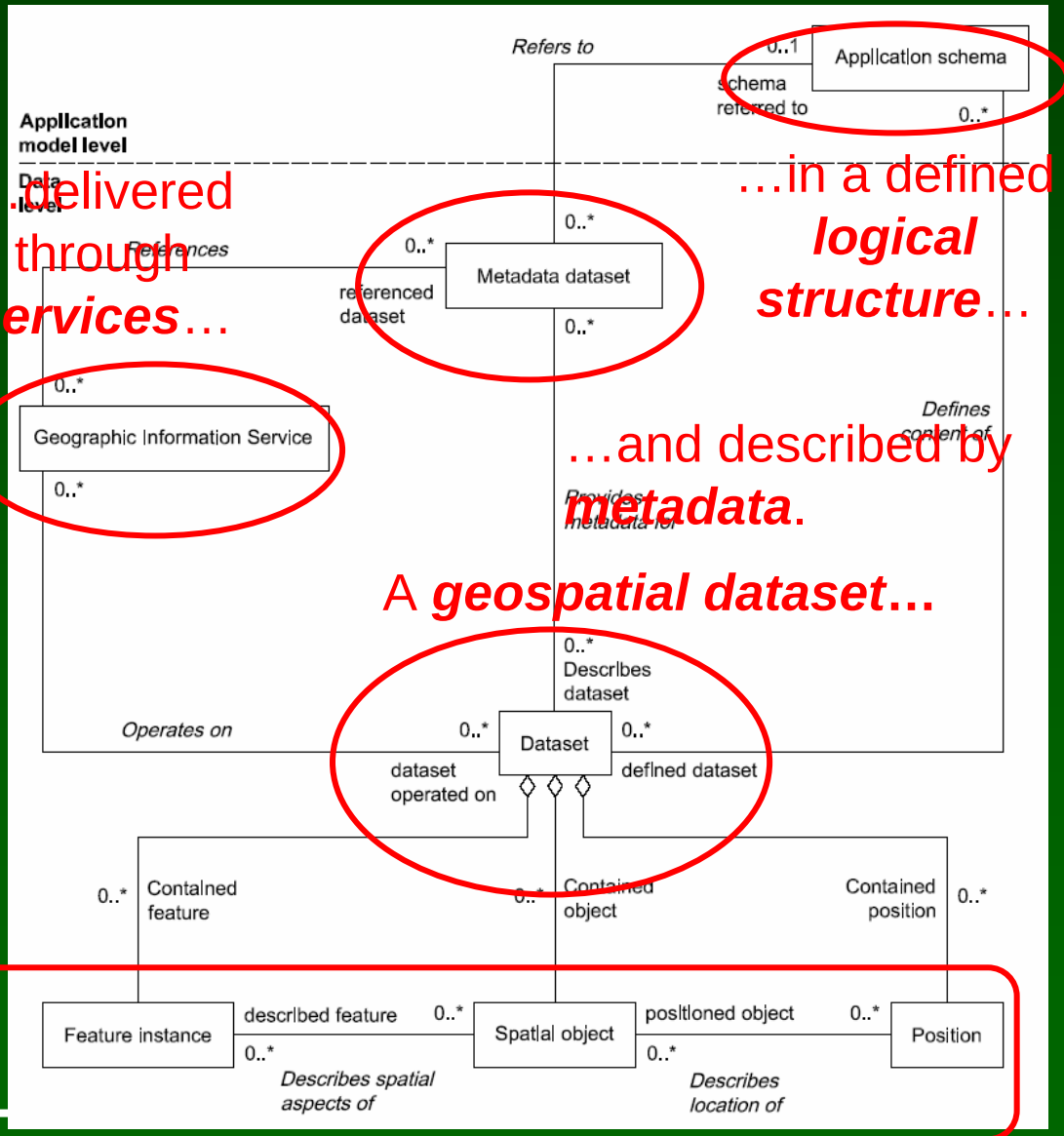
NB: Now using Mapserver for Interoperability experiments

WCS Sequence

- select source coverage
  ↓
  perform spatial subsetting
  ↓
  perform temporal subsetting
  ↓
  perform range subsetting
  ↓
  perform spatial scaling into result bounding box
  (including interpolation as specified)
  ↓
  perform reprojection into target CRS
  (including interpolation as specified)
  ↓
  perform data format encoding (may involve implicit range interpolation and, hence, accuracy loss, depending on the format chosen)
  ↓
  transmit resulting byte string to target location

Standards

- **ISO 19101: Geographic information – Reference model**



…delivered through ***services***…

…in a defined ***logical structure***…

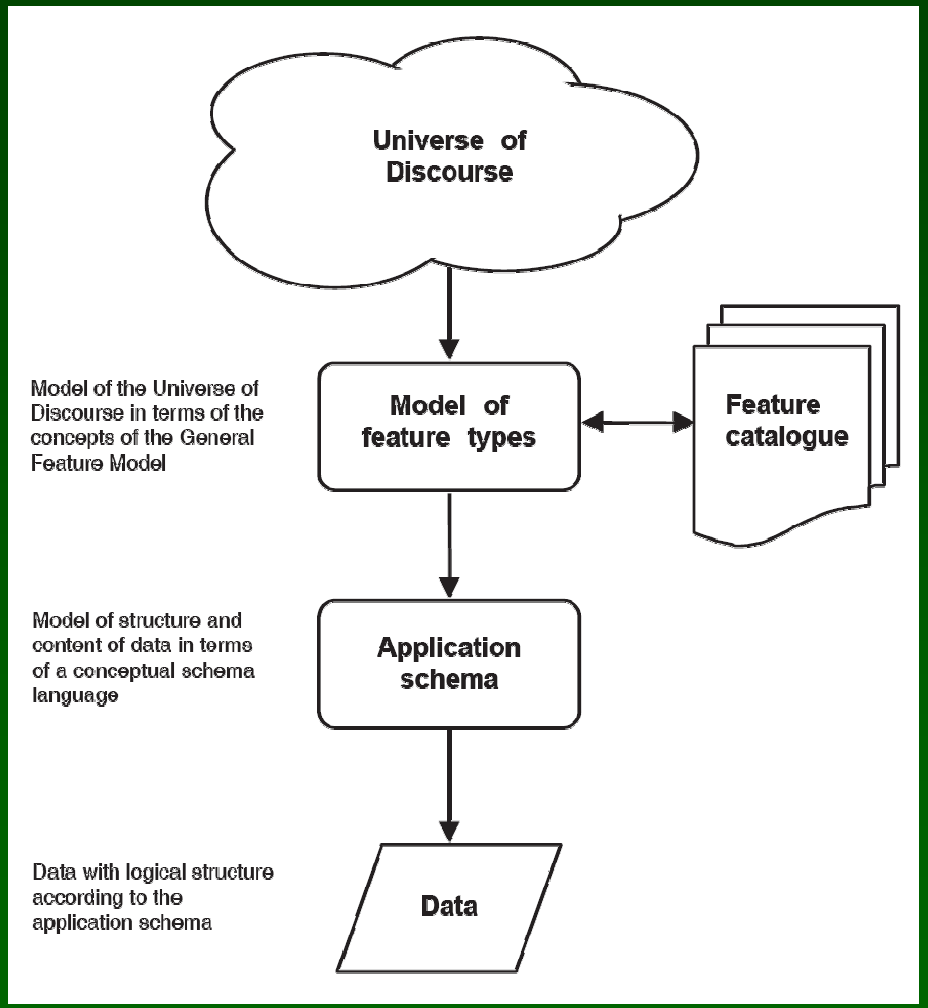
…and described by ***metadata***.

A ***geospatial dataset…***

…consists of ***features*** and related objects…

Standards

- Geographic ‘features’
  - “abstraction of real world phenomena” [ISO 19101]
  - Type or instance
  - Encapsulate important semantics in universe of discourse
  - “Something you can name”
- Application schema
  - Defines semantic content and logical structure
  - ISO standards provide toolkit:
    - spatial/temporal referencing
    - geometry (1-, 2-, 3-D)
    - topology
    - dictionaries (phenomena, units, etc.)
  - GML – canonical encoding

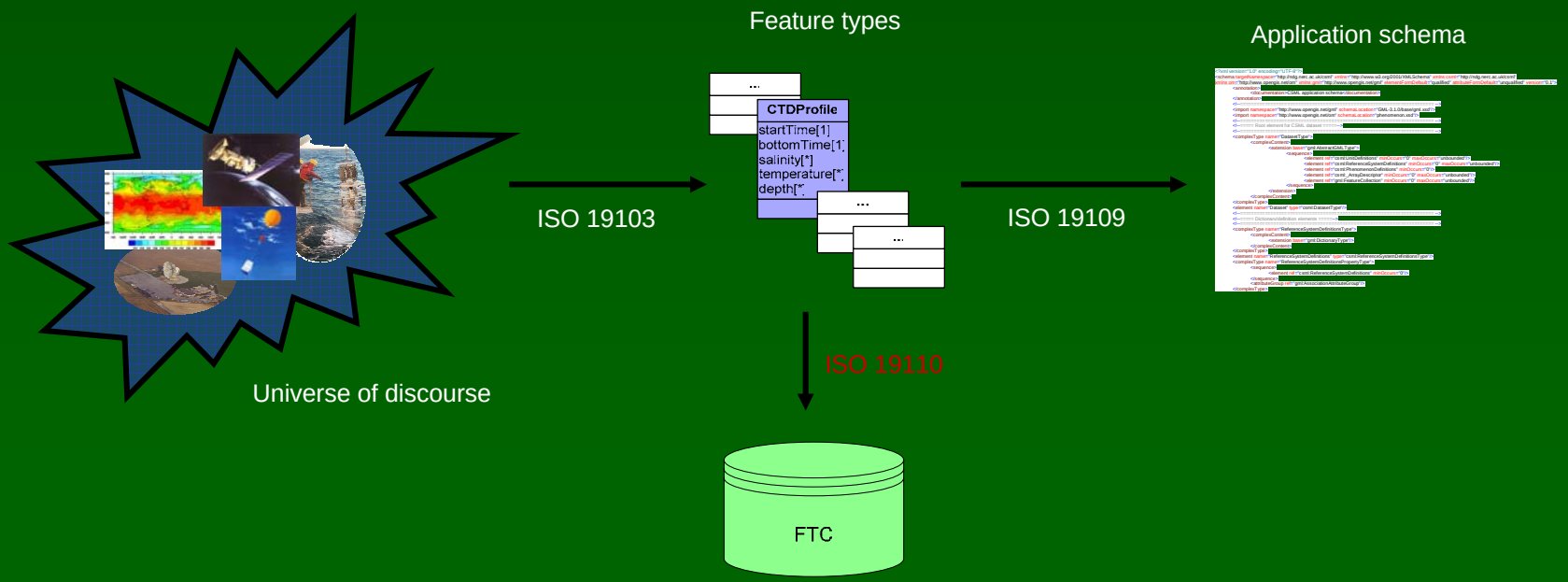


[from ISO 19109 “Geographic information – Rules for Application Schema”]

Standards

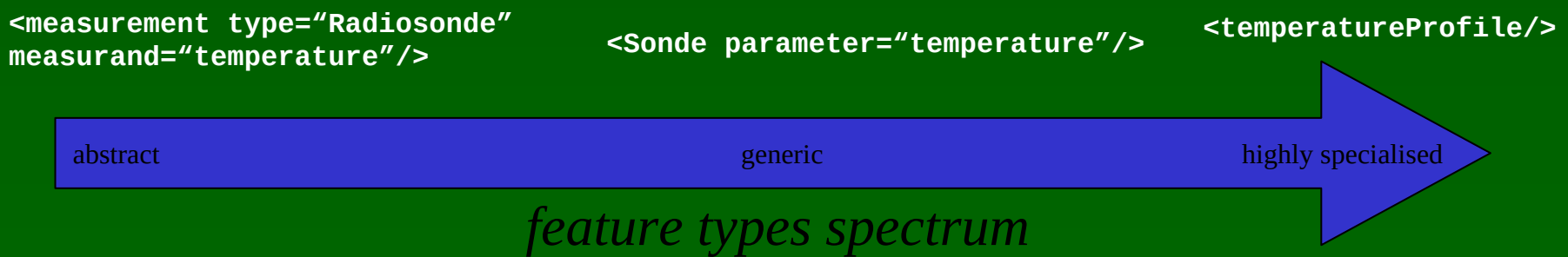
ISO standards

- TC211 – around 40 standards for geographic information
- Cover activity spectrum: discovery → access → use
- Provide a *framework* for data integration



Standards

- The importance of governance
  - Information community defined by shared semantics
  - Need community process to manage those semantics (definitions, models, vocabularies, taxonomies, etc.)
- Governance is the driver for granularity
  - Remit / interest determines appropriate granularity
  - ref. IOC, IHO, WMO



feature types spectrum

NDG-A: Climate Science Modelling Language

- Aims:
  - provide semantic integration mechanism for NDG data
  - explore new standards-based interoperability framework
  - emphasise *content*, not *container*

- Design principles:
  - offload semantics onto *parameter type* ('phenomenon', observable, measurand)
    - *e.g. wind-profiler, balloon temperature sounding*
  - offload semantics onto CRS
    - *e.g. scanning radar, sounding radar*
  - 'sensible plotting' as discriminant
    - *'in-principle' unsupervised portrayal*
  - explicitly aim for small number of *weakly-typed* features (in accordance with governance principle and NDG remit)

Climate Science Modelling Language

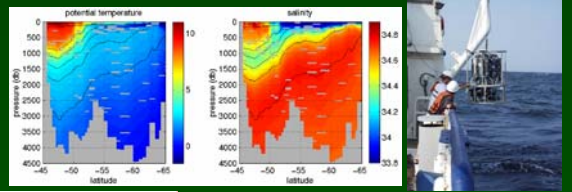
- CSML feature types
  - defined on basis of geometric and topologic structure

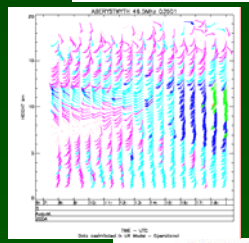
CSML feature type	***Description***	***Examples***
TrajectoryFeature	Discrete path in time and space of a platform or instrument.	ship's cruise track, aircraft's flight path
PointFeature	Single point measurement.	raingauge measurement
ProfileFeature	Single 'profile' of some parameter along a directed line in space.	wind sounding, XBT, CTD, radiosonde
GridFeature	Single time-snapshot of a gridded field.	gridded analysis field
PointSeriesFeature	Series of single datum measurements.	tidegauge, rainfall timeseries
ProfileSeriesFeature	Series of profile-type measurements.	vertical or scanning radar, shipborne ADCP, thermistor chain timeseries
GridSeriesFeature	Timeseries of gridded parameter fields.	numerical weather prediction model, ocean general circulation model

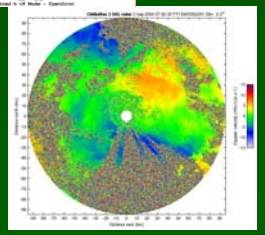
Climate Science Modelling Language

- CSML feature types
  - examples...

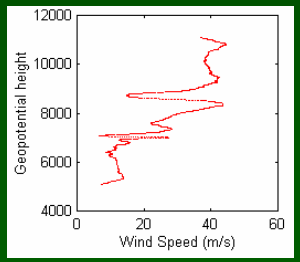
ProfileSeriesFeature



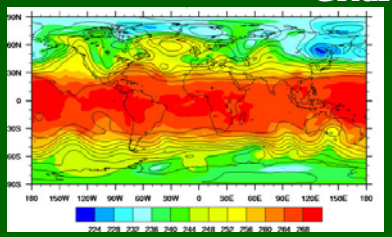




ProfileFeature

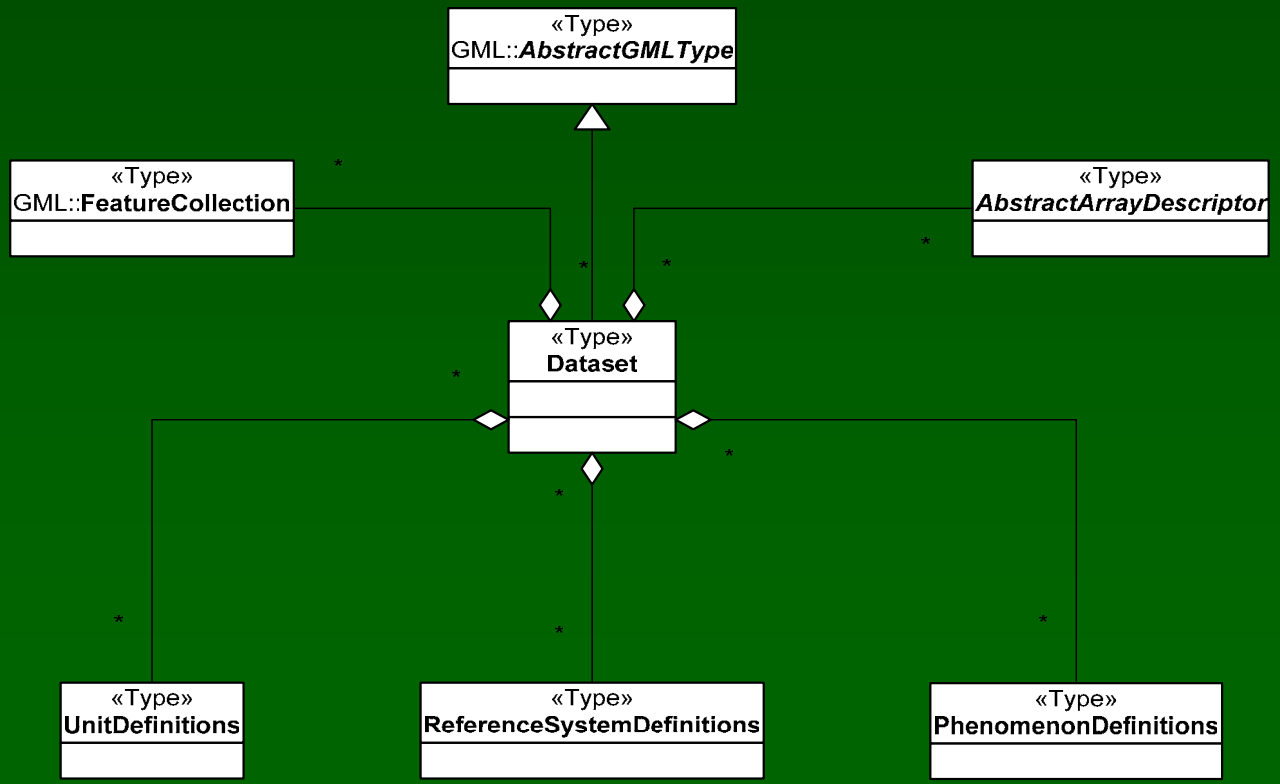


GridFeature



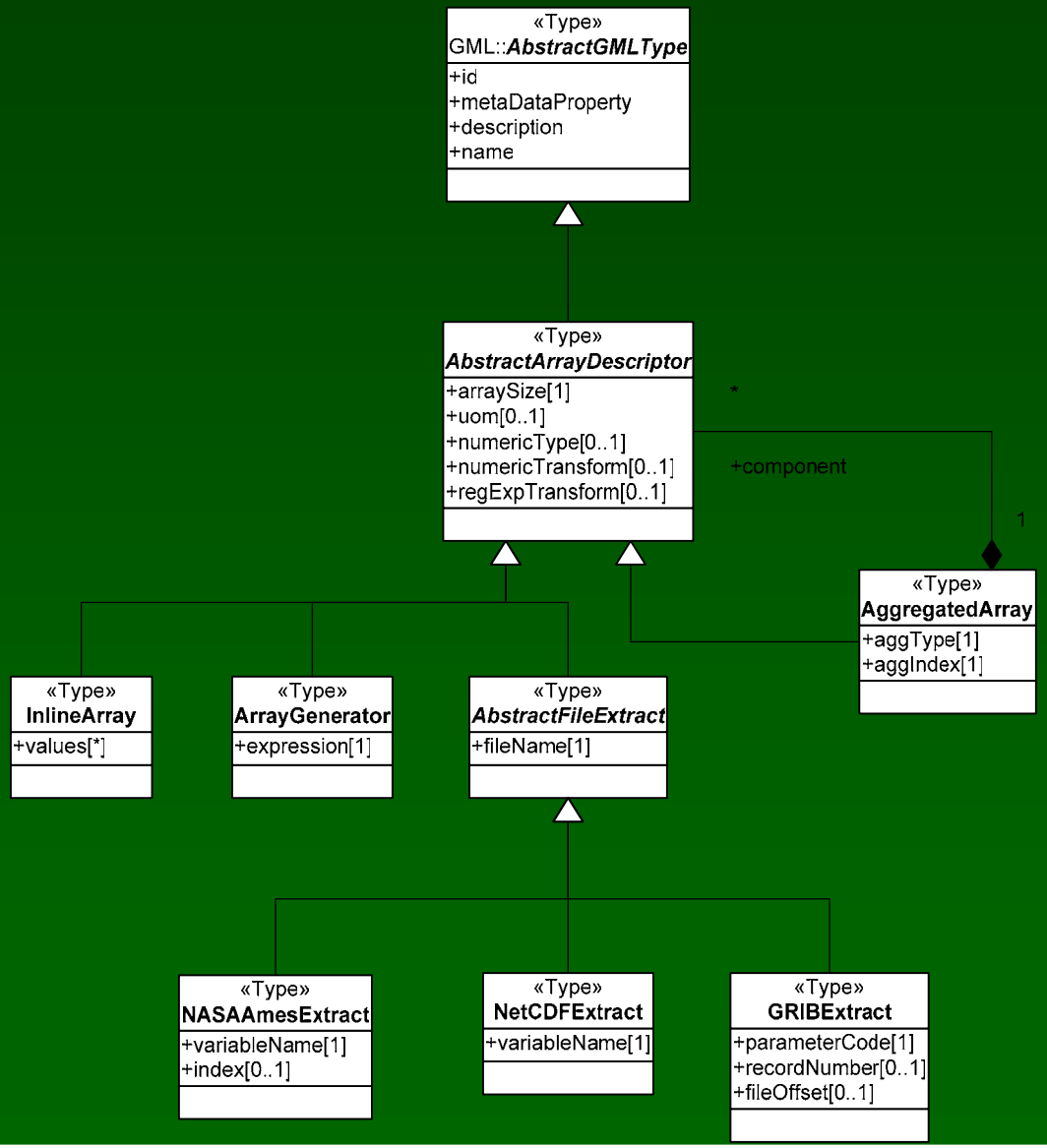
Climate Science Modelling Language

- Application schema
  - logical structure and semantic content of NDG ‘Dataset’
  - Based on GML 3.1

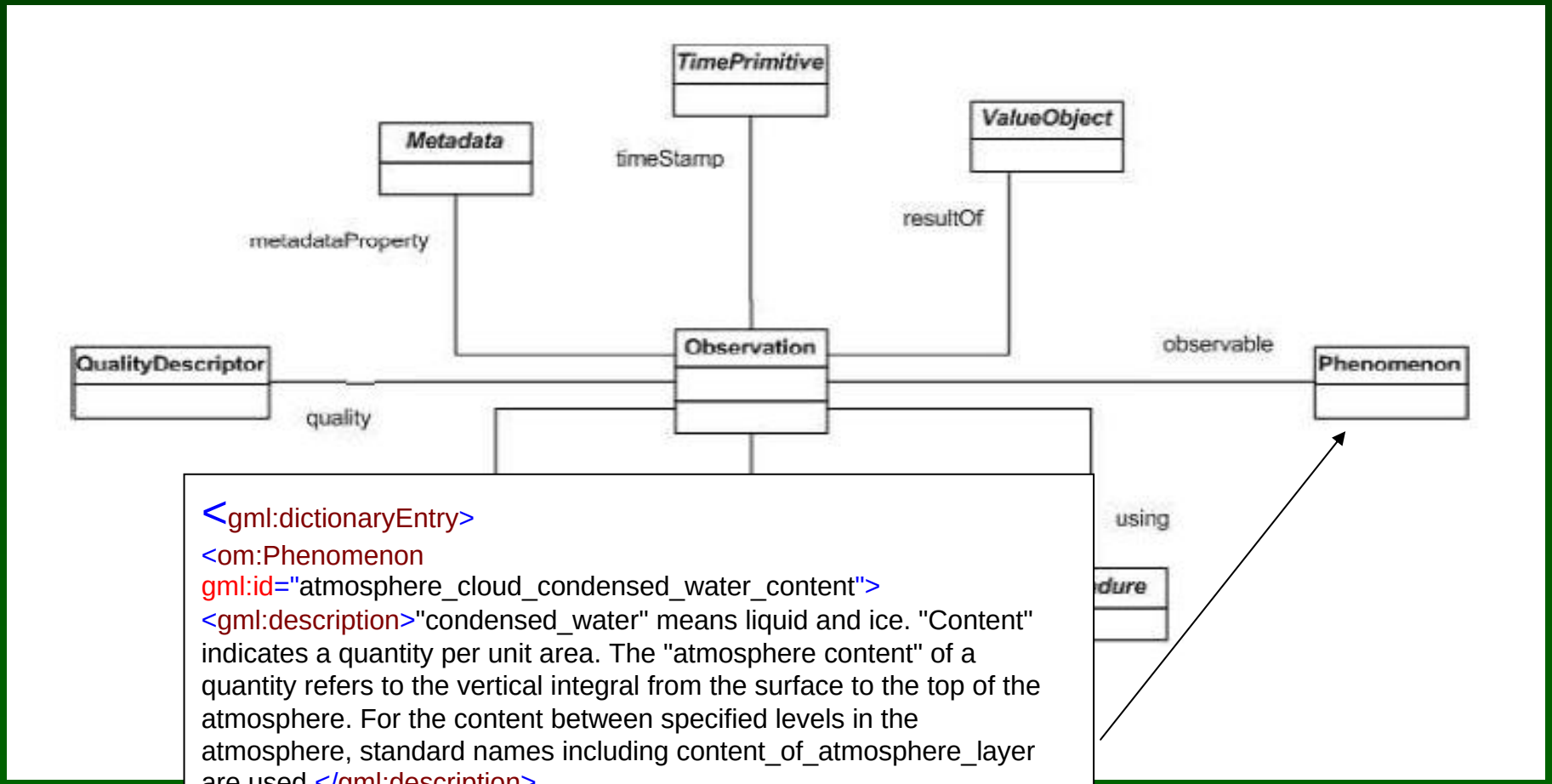


Climate Science Modelling Language

- Numerical array descriptors
  - provides ‘wrapper’ architecture for legacy data files
  - ‘Connected’ to data model numerical content through ‘xlink:href’
- Three subtypes:
  - InlineArray
  - ArrayGenerator
  - FileExtract (NASAAmes, NetCDF, GRIB)
- Composite design pattern for aggregation



CSML Observations




```
<gml:dictionaryEntry>
<om:Phenomenon
gml:id="atmosphere_cloud_condensed_water_content">
<gml:description>"condensed_water" means liquid and ice. "Content"
indicates a quantity per unit area. The "atmosphere content" of a
quantity refers to the vertical integral from the surface to the top of the
atmosphere. For the content between specified levels in the
atmosphere, standard names including content_of_atmosphere_layer
are used.</gml:description>
<gml:name codeSpace="http://www.cgd.ucar.edu/cms/eaton/cf-
metadata/">atmosphere_cloud_condensed_water_content</gml:name>
<gml:name codeSpace="GRIB">76</gml:name>
<gml:name codeSpace="PCMDI">clwvi</gml:name>
</om:Phenomenon>
</gml:dictionaryEntry>
```

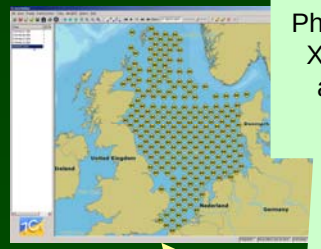
MarineXML Testbed

Data from different parts of the marine community conforming to a variety of schema (XSD)

For each XSD (for the source data) there is an XSLT to translate the data to the Feature Types (FT) defined by CSML. The FT's and XSLT are maintained in a 'MarineXML registry'

The FTs can then be translated to equivalent FTs for display in the ECDIS system

Phenomena in the XSD must have an associated portrayal



Biological Species

Chl-a from Satellite

Measured Hydrodynamics

Modelled Hydrodynamics

S-57v3 GML

XSD

XML

XML Parser

XSLT

XML

SENC

S52 Portrayal Library

SeeMyDENC

Data Dictionary

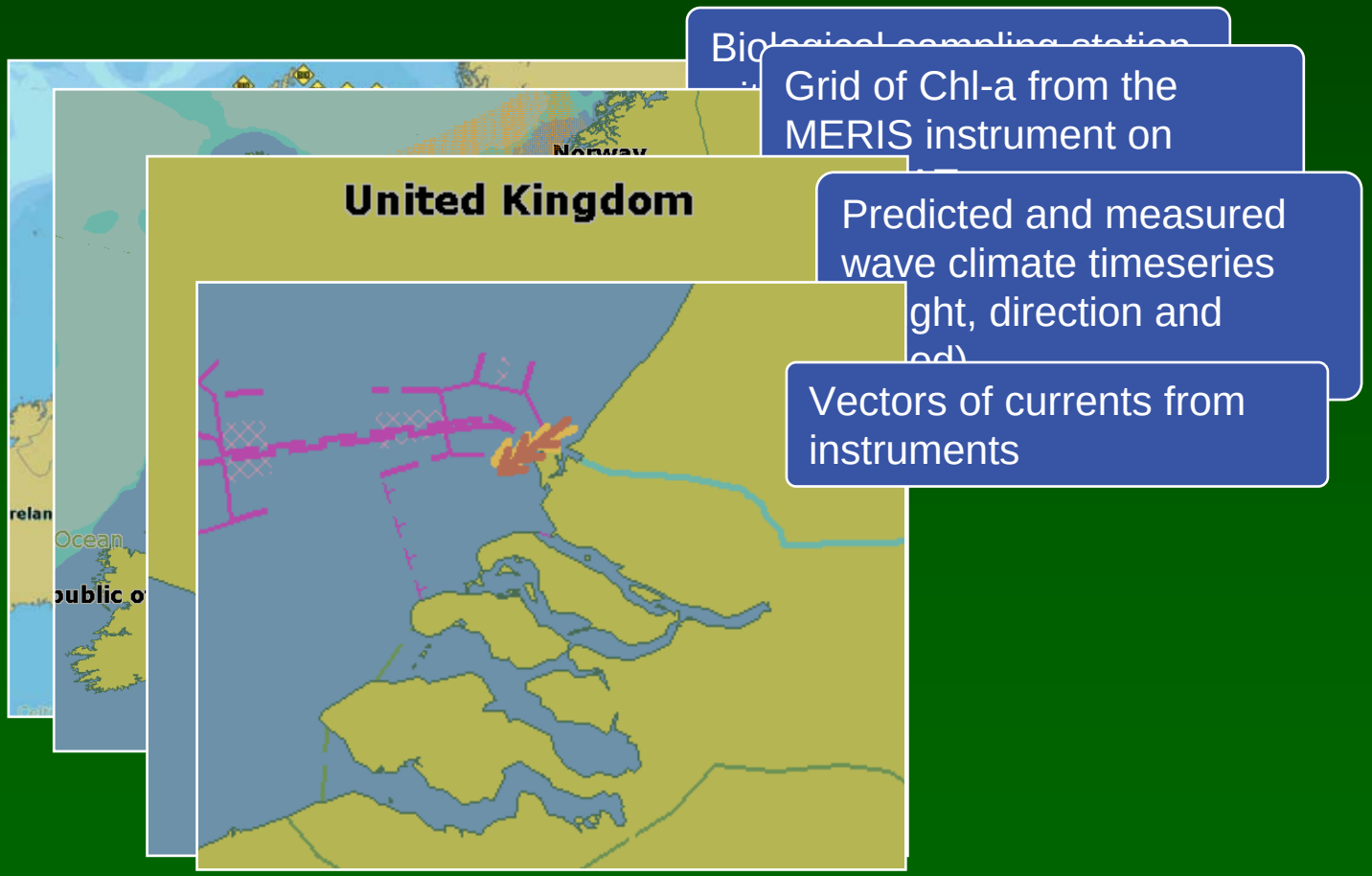
ECDIS acts as an example client for the data.

The result of the translation is an encoding that contains the marine data in weakly typed (i.e. generic) Features

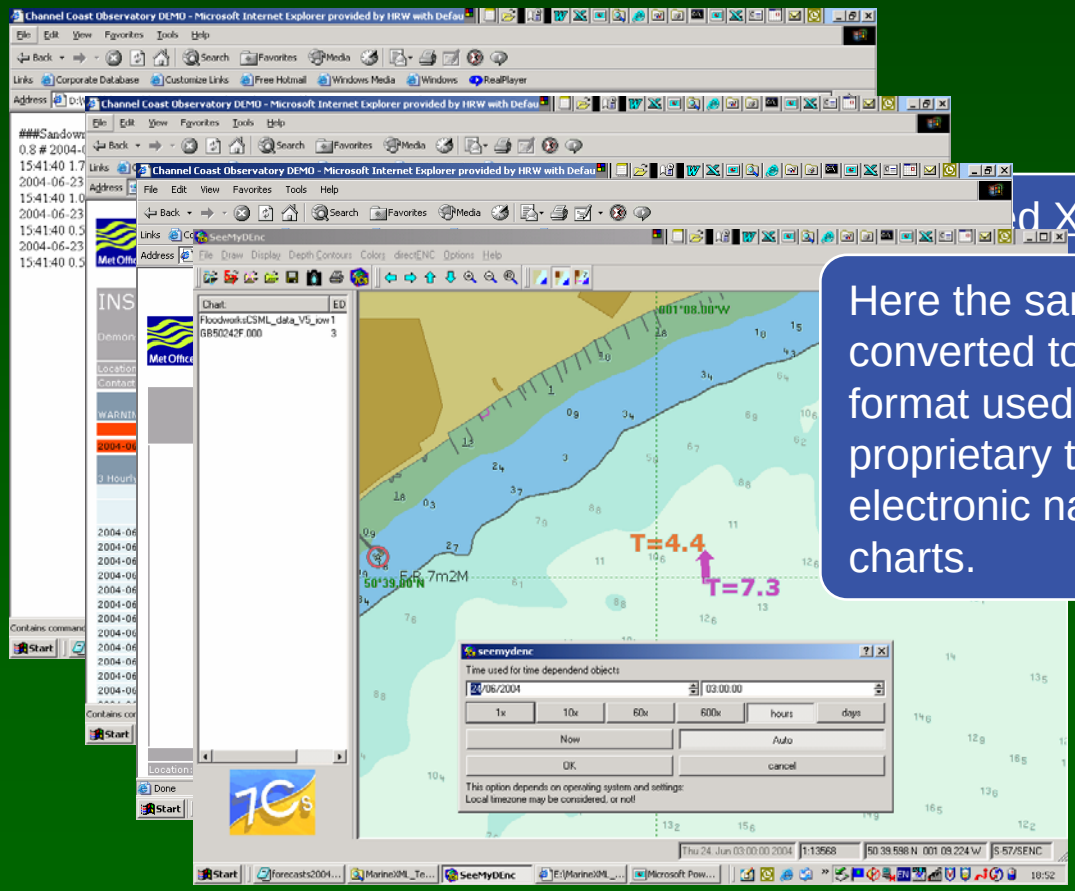
Feature described using S-57v3.1Application Schema can be imported and are equivalent to the same features in CSML'

Features in the source XSD must be present in the data dictionary.

MarineXML Testbed



The Concept of re-using Features



d XML is

Here the same XML is converted to the SENC format used in a proprietary tool for viewing electronic navigation charts.

All this requires agreement on standards

Climate Science Modelling Language

- Status:
  - Initial feature types defined
  - First draft application schema complete
  - Trial software tooling being coded (parser, netCDF instantiation)
  - Initial deployment trial across BODC, BADC datasets
- Future:
  - Separate out wrapper implementation (array descriptors)
  - Disallow ‘internal’ dictionaries
  - More strongly-typed features?
    - Complex feature (vector measurements)
    - Implicit Ensemble Support
    - Swathes
  - Follow (and pursue!) GML evolution, enhance compliance
  - Expand tooling

CSML Round Tripping - 1

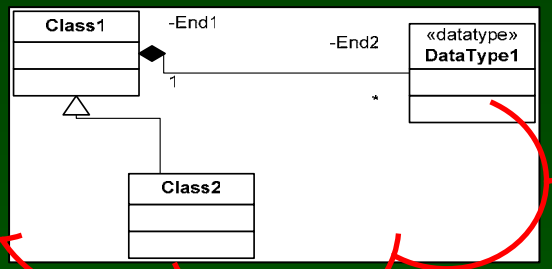
Managing semantics

New Dataset

101
010

Conforms to

conceptual model

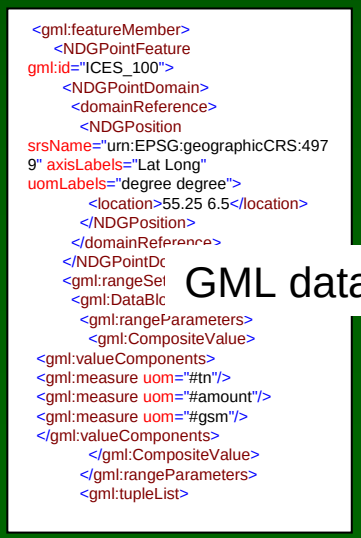


UGAS

XML

GML app schema

instance



GML dataset

produces

Application

Under Development

parser

CSML Round Tripping - 2

Managing data - 1

Under Development

XML

GML app schema

CF Dataset

101
010

scanner

CF

produces

Application

Under Development

parser

```
<gml:featureMember>
    <NDGPointFeature
gml:id="ICES_100">
      <NDGPointDomain>
        <domainReference>
          <NDGPosition
srsName="urn:EPSG:geographicCRS:497
9" axisLabels="Lat Long"
uomLabels="degree degree">
            <location>55.25 6.5</location>
          </NDGPosition>
        </domainReference>
      </NDGPointDc
      <gml:rangeSel
        <gml:DataBlc
          <gml:rangeParameters>
            <gml:CompositeValue>
<gml:valueComponents>
<gml:measure uom="#tn"/>
<gml:measure uom="#amount"/>
<gml:measure uom="#gsm"/>
</gml:valueComponents>
            </gml:CompositeValue>
          </gml:rangeParameters>
          <gml:tupleList>
```

GML dataset

instance

Managing Data 2

CF Dataset

CF Dataset

101
010

scanner

```
<gml:featureMember>
  <NDGPointFeature
gml:id="ICES_100">
    <NDGPointDomain>
      <domainReference>
        <NDGPosition
srsName="urn:EPSG:geographicCRS:497
9" axisLabels="Lat Long"
uomLabels="degree degree">
          <location>55.25 6.5</location>
        </NDGPosition>
      </domainReference>
    </NDGPointDo
    <gml:rangeSe
      <gml:DataBlo
        <gml:rangeParameters>
          <gml:CompositeValue>
<gml:valueComponents>
<gml:measure uom="#tn"/>
<gml:measure uom="#amount"/>
<gml:measure uom="#gsm"/>
</gml:valueComponents>
          </gml:CompositeValue>
        </gml:rangeParameters>
        <gml:tupleList>
```

GML dataset

Define Dataset

DECISION PROCESSES

XSLT

XML

ISO19115

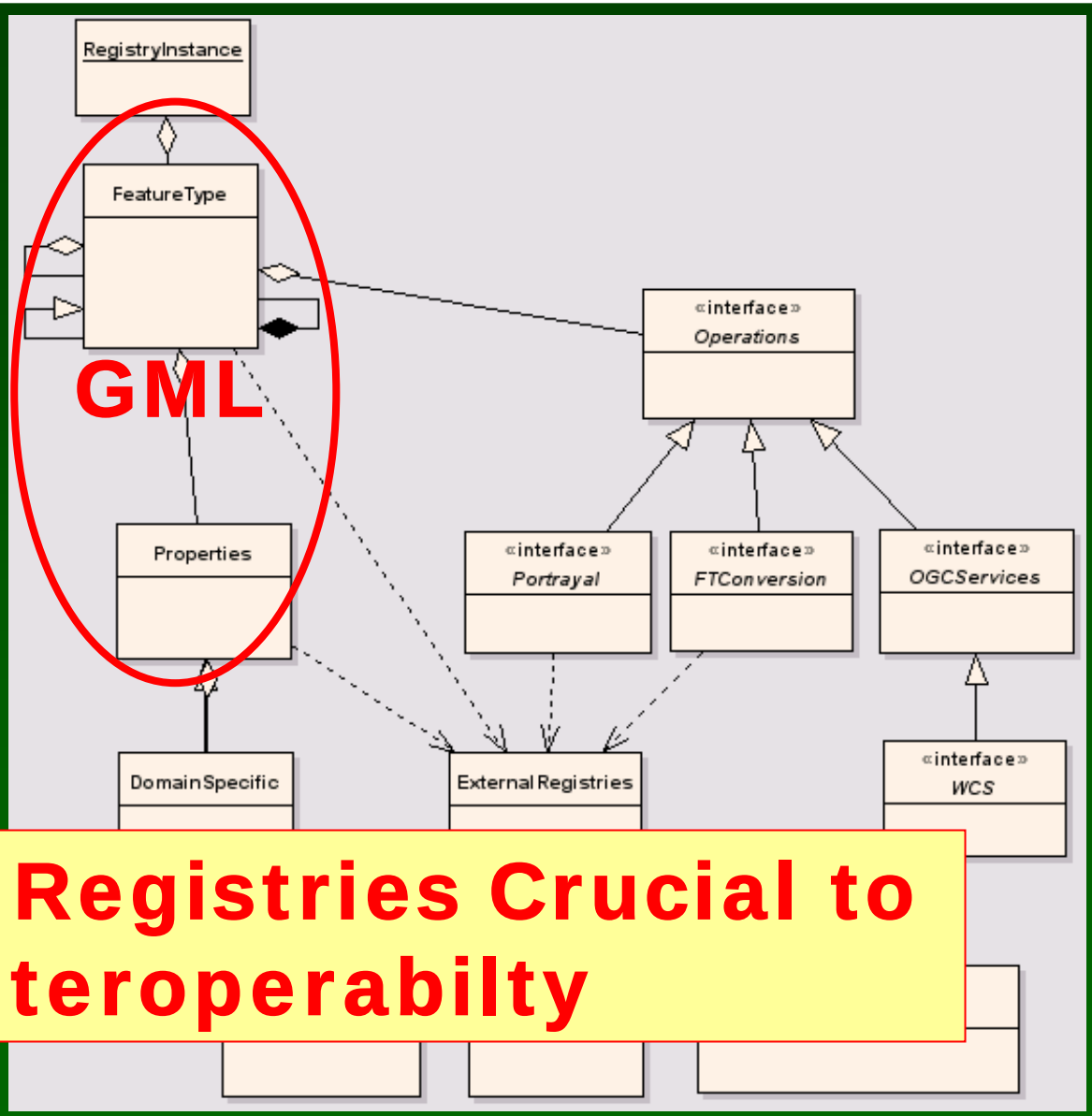
Add Information

PUBLISH

Registry Relationships

Feature Types, need

- relationships
- operations



Managed Registries Crucial to Interoperabilty

Catalogues or Registries?

The terms *catalogue* and *registry* are often used interchangeably, but in OGC speak:

"a registry is a catalogue that exemplifies a formal registration process (e.g., such as those described in ISO 19135 or ISO 11179-6)."

A registry is typically maintained by an authorized registration authority who assumes responsibility for complying with a set of policies and procedures for accessing and managing metadata items within some application domain.

Summary Points

- Need clear distinction in mind between inter- and intra-operability
  - When to use GML etc, and when not to!
- Need clear governance principles at a number of levels for
  - Dataset definitions (data provider, e.g. BADC)
  - Feature-type definitions (community, e.g. WMO)
  - Interoperability (frameworks, e.g. OGC)
- CSML
  - Provides some geometric features based on portrayal
  - Can and will be extended
  - Provides generalisation of specific examples, e.g. METARS