gCube Metadata Framework (gCubeMF)

Pasquale Pagano
CNR-ISTI
pasquale.pagano@isti.cnr.it

Storage	gCube Storage Management	
Framework Core	Obj2Obj Manager / Metadata Catalog / Col Manager	
Framework Add-on	XML Indexer / Metadata Broker / Annotation	
Services	Full Text Indices / Geo Indices / Feature-based Indices	

gCube
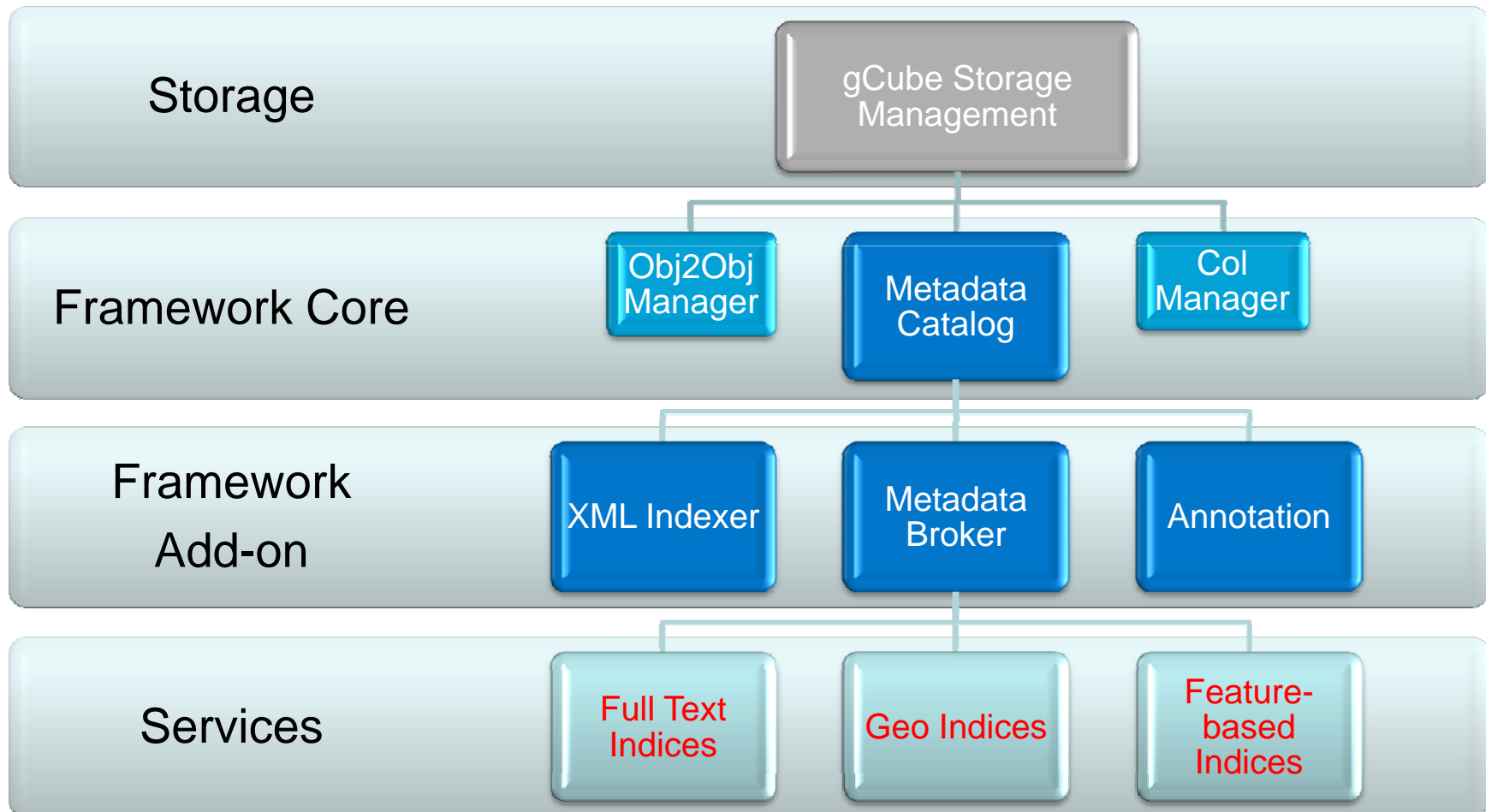
- Manage arbitrary object-to-object relationships

- Manage collections of arbitrary metadata and their members

- Can be expanded and plugged with semantic-specific services

- Is a logical layer and rely on an external storage system

- Support replication and distribution

- Support for validation (in case of XML-based metadata)

- Metadata Management
  - creation and removal of Metadata Collections
  - upload, update, removal of Metadata Objects

- Relationships Management
  - creation and removal of nested object-to-object relationships (both at storage time as well as later)
  - discovery and access through navigation over relationships

- Brokerage (optional)
  - transformation and storage (if needed) of objects through pluggable Transformation Programs

- Indexing (optional)
  - indexing of XML Objects and discovery through xPath/xQueries expressions

- Implemented by a set of Web Services over Java technology

- By reference and by value support for bulk operations

- Metadata are returned in (configurable) chunks

- Rely on the gCube Storage Management

- Support GSI security conversations

```xml
<mc:diligentmetadata xmlns:mc="http://diligentproject.org/namespaces/metadatamanagement/mc/model">
  <mc:header>
    <mc:OID>1fa83b10-fb3c-11db-857b-d6a400c8bdbb</mc:OID>
    <mc:targetOID>cbea3ae0-fb37-11db-857b-d6a400c8bdbb</mc:targetOID>
    <mc:memberOf>
      <mc:DiligentID>25ad3c50-fa41-11db-a270-9c01d805f283</mc:DiligentID>
    </mc:memberOf>
    <mc:creation>2007-05-05:09:09:17.CEST</mc:creation>
    <mc:lastUpdate>2007-05-05:09:09:17.CEST</mc:lastUpdate>
  </mc:header>
  <mc:body>
    <DIMAP_DOCUMENT>
      <ORIG_NAME>MER_RR__2P___ALGAL_12005080320050809W01000N3000E04000N450009266212935.dim.xml</ORIG_NAME>
      <DATASET>MER_RR__2P</DATASET>
      <INSTRUMENT>MER</INSTRUMENT>
      <RESOLUTION_TYPE>RR</RESOLUTION_TYPE>
      <PRODUCT_LEVEL>2P</PRODUCT_LEVEL>
      <PRODUCT_NAME>MERIS</PRODUCT_NAME>
      <BAND_USED>___ALGAL_1</BAND_USED>
      <BAND_USED_NORM>ALGAL 1</BAND_USED_NORM>
      <START_YEAR>2005</START_YEAR>
      <START_MONTH>08</START_MONTH>
      <START_DAY>03</START_DAY>
      <END_YEAR>2005</END_YEAR>
      <END_MONTH>08</END_MONTH>
      <END_DAY>09</END_DAY>
      <START_DATE>2005-08-03</START_DATE>
      <END_DATE>2005-08-09</END_DATE>
      <START_DATE_INT>20050803</START_DATE_INT>
      <END_DATE_INT>20050809</END_DATE_INT>
      <LONMIN>-10.00</LONMIN>
      <LATMIN>30.00</LATMIN>
      <LONMAX>40.00</LONMAX>
      <LATMAX>45.00</LATMAX>
      <LONMIN_INT>17000</LONMIN_INT>
      <LATMIN_INT>12000</LATMIN_INT>
      <LONMAX_INT>22000</LONMAX_INT>
      <LATMAX_INT>13500</LATMAX_INT>
      <RESOLUTION_M>9266</RESOLUTION_M>
      <RESOLUTION_KM>9.266212935</RESOLUTION_KM>
      <COVER_REGIONS>World</COVER_REGIONS>
      <OVERLAP_REGIONS>World Europe Bigger_Europe Smaller_Europe  Mediterranean Iberia North_Atlantic Africa North_Africa Middle_East Portugal</OVERLAP_REGIONS>
      <METADATA_FORMAT>DIMAP</METADATA_FORMAT>
      <METADATA_PROFILE>BEAM-DATAMODEL-V1</METADATA_PROFILE>
      <DATASET_SERIES>BEAM-PRODUCT</DATASET_SERIES>
      <DATASET_NAME>algal_1</DATASET_NAME>
      <DATASET_PRODUCER_NAME>Grid on-Demand - ESRIN ESA</DATASET_PRODUCER_NAME>
      <PRODUCT_TYPE>EO GRID</PRODUCT_TYPE>
```

6

```xml
<mc:memberOf>
    <mc:DiligentID>25ad3c50-fa41-11db-a270-9c01d805f283</mc:DiligentID>
</mc:memberOf>
<mc:creation>2007-05-05:09:10:02.CEST</mc:creation>
<mc:lastUpdate>2007-05-05:09:10:02.CEST</mc:lastUpdate>
</mc:header>
<mc:body>
    <esObject>
        <title>MERIS - ___ALGAL_1 - 2005-08-20 - 2005-08-26</title>
        <purpose/>
        <descKeys>Chlorophyll distribution product</descKeys>
        <timeFrame>
            <begin>2005-08-20</begin>
            <end>2005-08-26</end>
        </timeFrame>
        <location>
            <width/>
            <centerLat/>
            <centerLon/>
            <westBL>-10.00</westBL>
            <eastBL>40.00</eastBL>
            <southBL>30.00</southBL>
            <northBL>45.00</northBL>
        </location>
        <type>ENVI</type>
        <suppInfo>PRODUCER: Grid on-Demand - ESRIN ESA</suppInfo>
        <suppInfo>SOURCE: BEAM-PRODUCTalgal_1</suppInfo>
        <suppInfo>IDENTIFIER: MER_RR__2P___ALGAL_12005082020050826W01000N3000E04000N450009266212935.dim.xml</suppInfo>
        <description>DATASET: MER_RR__2P</description>
        <description>INSTRUMENT: MER</description>
        <description>RESOLUTION_TYPE: RR</description>
        <description>PRODUCT_LEVEL: 2P</description>
        <description>BAND_USED_NORM: ALGAL 1</description>
        <description>START_YEAR: 2005</description>
        <description>START_MONTH: 08</description>
        <description>START_DAY: 20</description>
        <description>END_YEAR: 2005</description>
        <description>END_MONTH: 08</description>
        <description>END_DAY: 26</description>
        <description>START_DATE_INT: 20050820</description>
        <description>END_DATE_INT: 20050826</description>
        <description>LONMIN: -10.00</description>
        <description>LATMIN: 30.00</description>
        <description>LONMAX: 40.00</description>
        <description>LATMAX: 45.00</description>
```

7

```xml
<mc:diligentmetadata xmlns:mc="http://diligentproject.org/namespaces/metadatamanagement/mc/model">
    <mc:header>
        <mc:OID>87tg3b10-ft4c-09dt-832b-d6a400c8sdrb</mc:OID>
        <mc:targetOID>cbea3ae0-fb37-11db-857b-d6a400c8bdbb</mc:targetOID>
        <mc:memberOf>
            <mc:DiligentID>25ad3c50-fa41-11db-a270-9c01d805f283</mc:DiligentID>
        </mc:memberOf>
        <mc:creation>2007-07-22:10:23:32.CEST</mc:creation>
        <mc:lastUpdate>2007-08-01:11:28:01.CEST</mc:lastUpdate>
    </mc:header>
    <mc:body>
        <ft:fulltext>
```

SAHEL WEATHER AND CROP SITUATION REPORT Report No.1, 13 June 2003 THE RAINY SEASON HAS STARTED GENERALLY ON TIME IN THE CENTRE AND EAST OF THE SAHEL SUMMARY The rainy season started in late April or May in southern Burkina Faso, Chad, Mali, Niger and the extreme south-east of Senegal. Seasonably dry conditions prevail in the rest of Senegal, Cape Verde, The Gambia and Mauritania. This corresponds to the normal pattern in the Sahel, except for Guinea Bissau, where satellite imagery indicates a delay in the onset of the rains, like last year. This situation may affect crop production as the growing season in this country usually starts in late April/early May. Land preparation and planting are in progress following the onset of the rains. Seed availability problems are likely in countries of western Sahel (Cape-Verde, The Gambia, Guinea Bissau, Mauritania and Senegal), where crop production was severely affected by drought last year. The pest situation is calm, although a few Desert Locusts were reported in early May near Gao in Mali. SYNTHESIS SITUATION MAP AS OF 11 JUNE 2003 Mauritania Mali Cape Verde Chad Niger Sénégal Gambia Guinea Bissau Good rains in May Insufficient rains, delayed plantings Legend Burkina Faso No rain up to early June Low numbers of Desert Locusts Global Information and Early Warning System on Food and Agriculture FAO/GIEWS Sahel Report No.1 13 June 2003 Page 1 SITUATION BY COUNTRY BURKINA FASO The rainy season has started on time. The first significant rains were registered in early April in the south and south-west. They progressed northwards in May and covered almost the entire country during the last dekad. Precipitation was generally above average and cumulative rainfall as of late June was above average in most meteorological stations. Sowing of millet and sorghum is underway in the south, west and south-west. Elsewhere, land preparation is underway. No pest activity is reported. Seed availability is generally adequate following the 2002 record harvest and seed distribution to returnees and refugees from Côte d'Ivoire under an FAO Emergency Agricultural Assistance Project. CAPE VERDE Seasonably dry conditions prevail. Planting of maize normally starts in July with the onset of the rains on the main islands. Seed shortages are likely following the 2002 poor harvest. CHAD The cropping season has started on time in the Sudanian zone. Satellite imagery indicates that the rainy season started in late May in the south, although significant rains were registered in mid-April in the extreme south. Planting of coarse grains is underway in the south, in the Sudanian zone. Land preparation is about to start in the Sahelian zone. Seed availibility should be adequate following the 2002 average harvest. THE GAMBIA Seasonably dry conditions prevail. The rains have not yet started and farmers are currently preparing their fields. Planting is expected to start in the weeks ahead with the onset of the rains. Following the 2002 poor harvest, seed shortages are likely. GUINEA-BISSAU The start of the rainy season is delayed. Satellite imagery indicates that the weather remained mostly dry until early June. This situation may affect crop production as the growing season usually starts in late April/early May. The availability of seeds may be limited in the chronically food deficit areas along the northern border with Senegal. Global Information and Early Warning System on Food and Agriculture FAO/GIEWS Sahel Report No.1 13 June 2003 Page 2 MALI The growing season has started in the south. The first significant rains were registered in the extreme south in April. They progressed northwards and remained generally above normal during the two last dekads of May, allowing first planting of millet and sorghum to start in the south. Seed availability is adequate following the 2002 average harvest. As recommended

8

- The Metadata Catalog service manages the lifecycle of Metadata Objects, Metadata Collections and their relationships

- In the Diligent infrastructure:

  - objects uploaded and connected with the objects they describe:
    - over 500.000 in the testing infrastructure
    - 179.712 in the PPS infrastructure (62 distinct metadata formats)

  - upload time (average over 3 months):
    - 750 ms /obj – to upload chunks of 5k docs (generic structure)
    - 1,4 s /obj – to upload and index chunks of 5k docs (generic structure)
    - 430 ms /obj – to upload and index chunks of 200k docs (predefined structure)

  - retrieval time (average over 3 months)
    - 100 ms /obj – to return the first chunk of metadata objects
    - 11 ms /obj – to return subsequent chunks

9

- The Metadata Broker service can transform Metadata Objects coming
  - from external locations
    - single metadata records or entire chunks passed by reference or by value
  - from internal locations
    - Metadata Collections passed by reference

- In the Diligent infrastructure:

- Collections > 100k objects have been transformed

- 18 Transformation Programs have been defined
  - (DIMAP -> DC), (DC -> ES), (ISO19115 -> Rowset), (DIMAP -> Rowset), (DIMAP -> ES), (TEI -> DC), etc.

- transformation time (average over 3 months):
  - 3 ms /obj

- The XMLIndexer service can index XML Metadata Objects with arbitrary schemas

- In the Diligent infrastructure:

  - xQueries response time (average):
    - 180 ms /obj to retrieve the 1st chunk
    - 16 ms /obj to retrieve subsequent chunks

  - served XQueries/XPath expressions per day (average over 3 months) :
    - ~ 8120 – 1 query every 10,64 s

  - indexing time (average over 3 months):
    - 700 ms /obj per chunk of 4.5k objects

- Annotation Management is a domain-specific service built on top of the Metadata framework
  - designed to enrich data and metadata object with additional information, e.g.
    - data provenance
    - process information
  - composed by an Annotation Back-end and an optional to use Annotation Front-end pluggable in any JSR168 Portal
  - generates metadata objects managed by the lower layer services of the gCubeMF
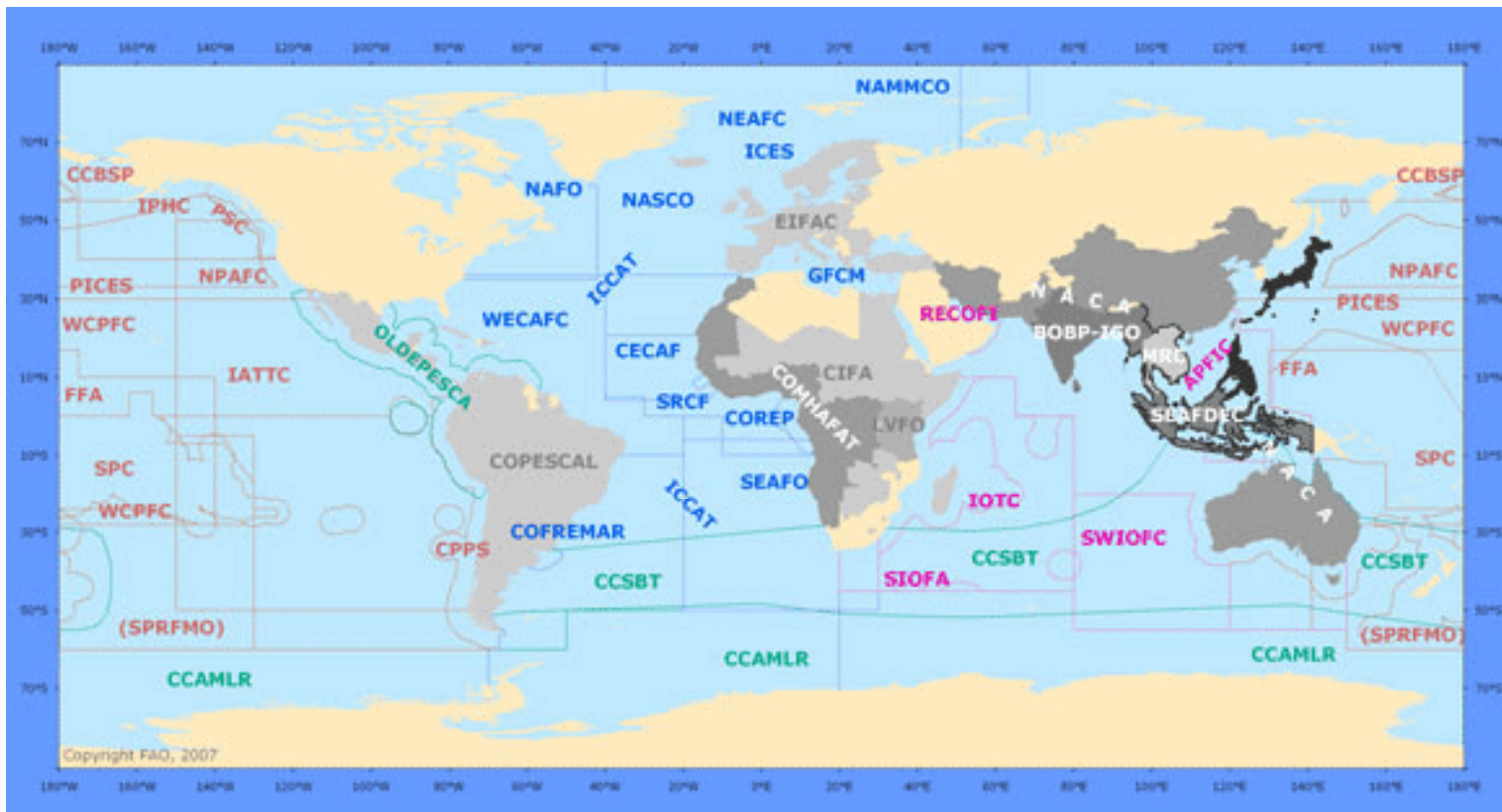
- In the Diligent infrastructure:
  - used to create human generated annotations treated as specialized metadata object with is-annotated-by relationships
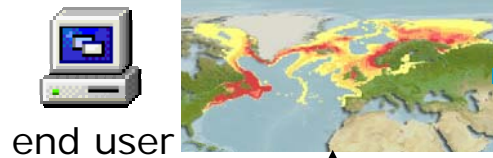
gCube

- Metadata brokerage has allowed the run time creation of specialized Indices like:

  - Geo-Indices
    - o geo Rowset format generated from
      - Dimap
      - ISO 19115
      - EIDB

  - Interoperable Indices
    - o Rowset format generated from any other

- D4Science project will put in production the gCubeMF
  - "*The project will deploy, progressively consolidate and expand the e-Infrastructures built so far by the EGEE and DILIGENT projects so that they address the needs of several new scientific communities affiliated with the broad disciplines of* **Environmental Monitoring** *and* **Fishery Resources Management**"
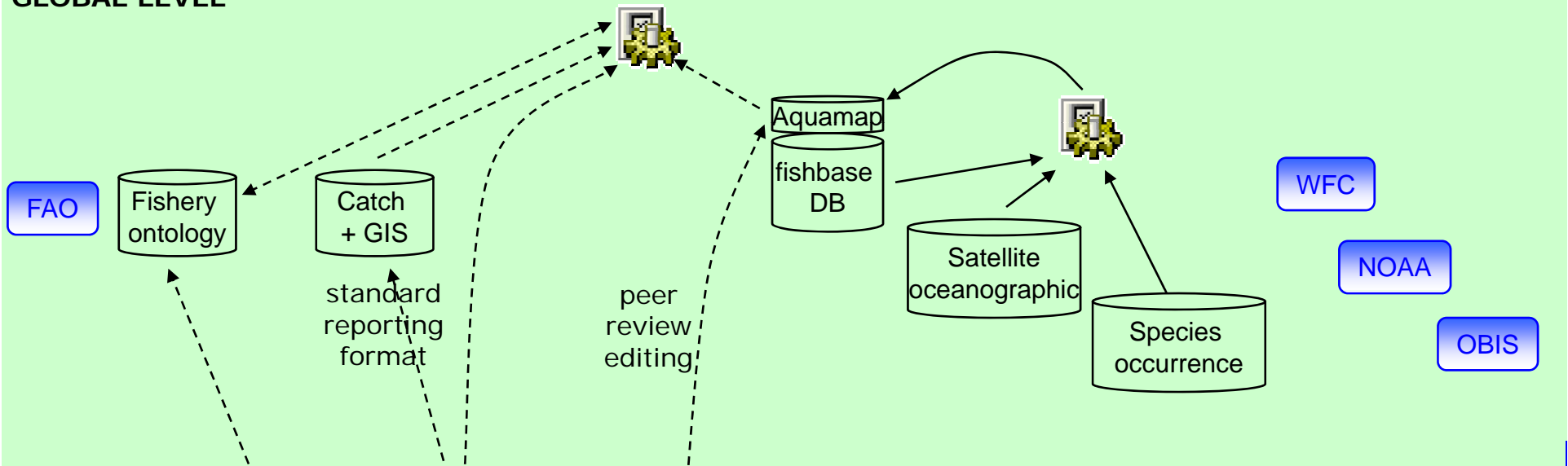
gCube

- *Fishery Resources Management*
  - **Regional RFB's**

end user

Fishery Resources Management
- Organigram of possible workflow

GLOBAL LEVEL

FAO

Fishery ontology

Catch + GIS

standard reporting format

peer review editing

Aquamap

fishbase DB

Satellite oceanographic

Species occurrence

WFC

NOAA

OBIS

Reference system

Catch + GIS

RFBs

REGIONAL LEVEL

16

QUESTIONS ?