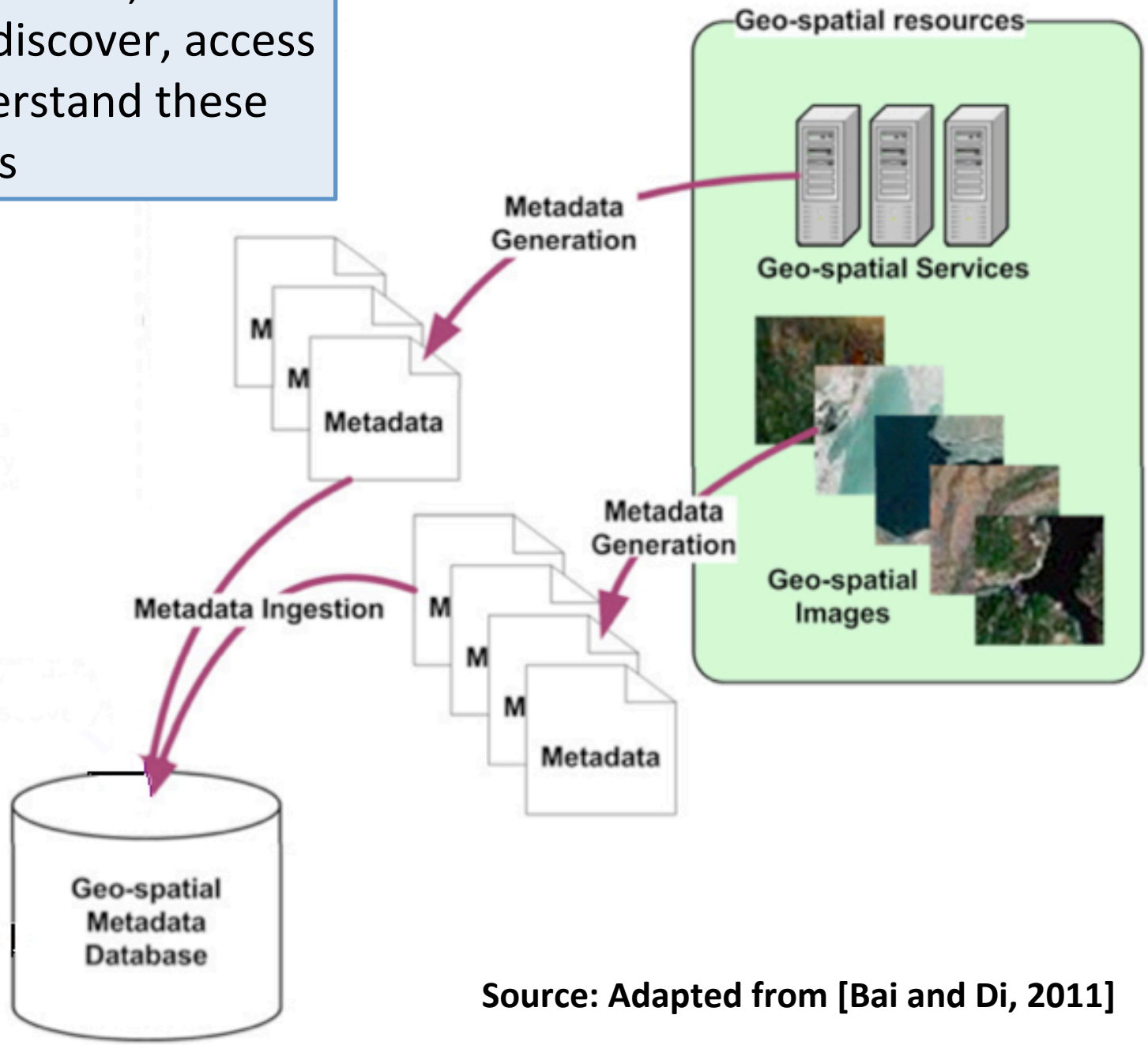
Data Discovery

Lubia Vinhas

Referata GeoInformatica, January 20th, 2012

TECHNICAL BACKGROUND

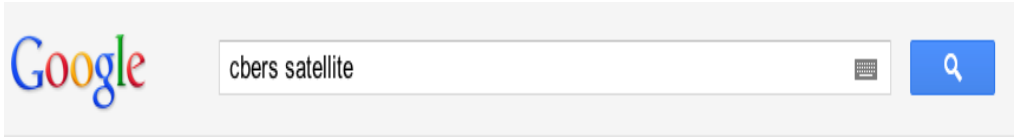
In this scenario, users want to discover, access and understand these resources



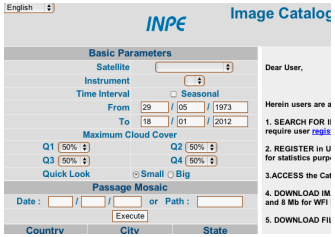
Source: Adapted from [Bai and Di, 2011]

Initial approach

Discover using search engines



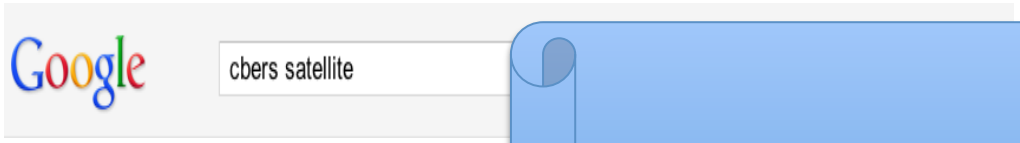
Access through specialized web interfaces



Understanding based on community knowledge

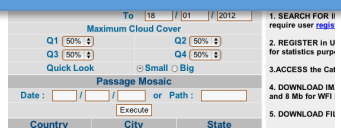
Initial approach

Discover using search engines

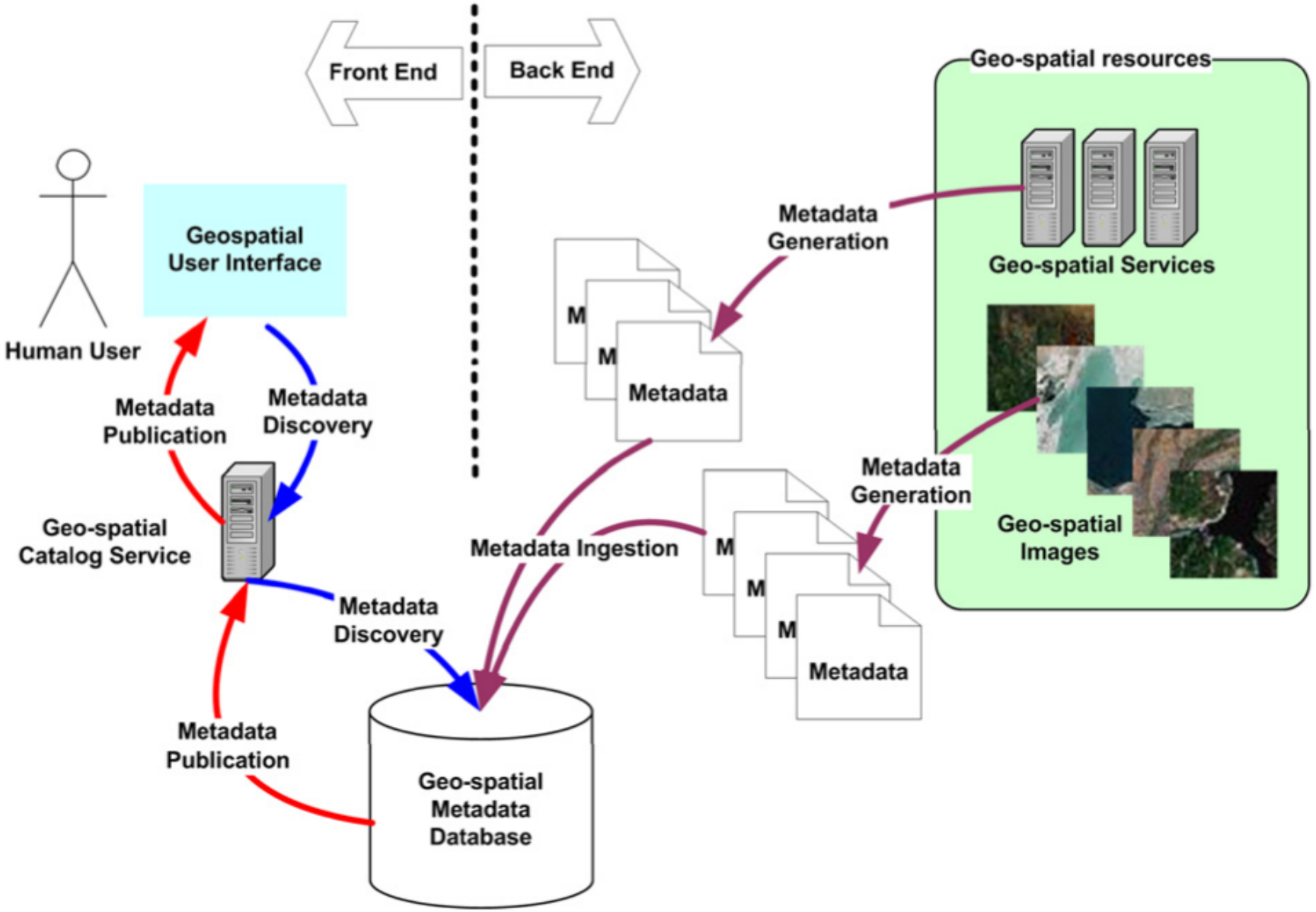


Access through sp

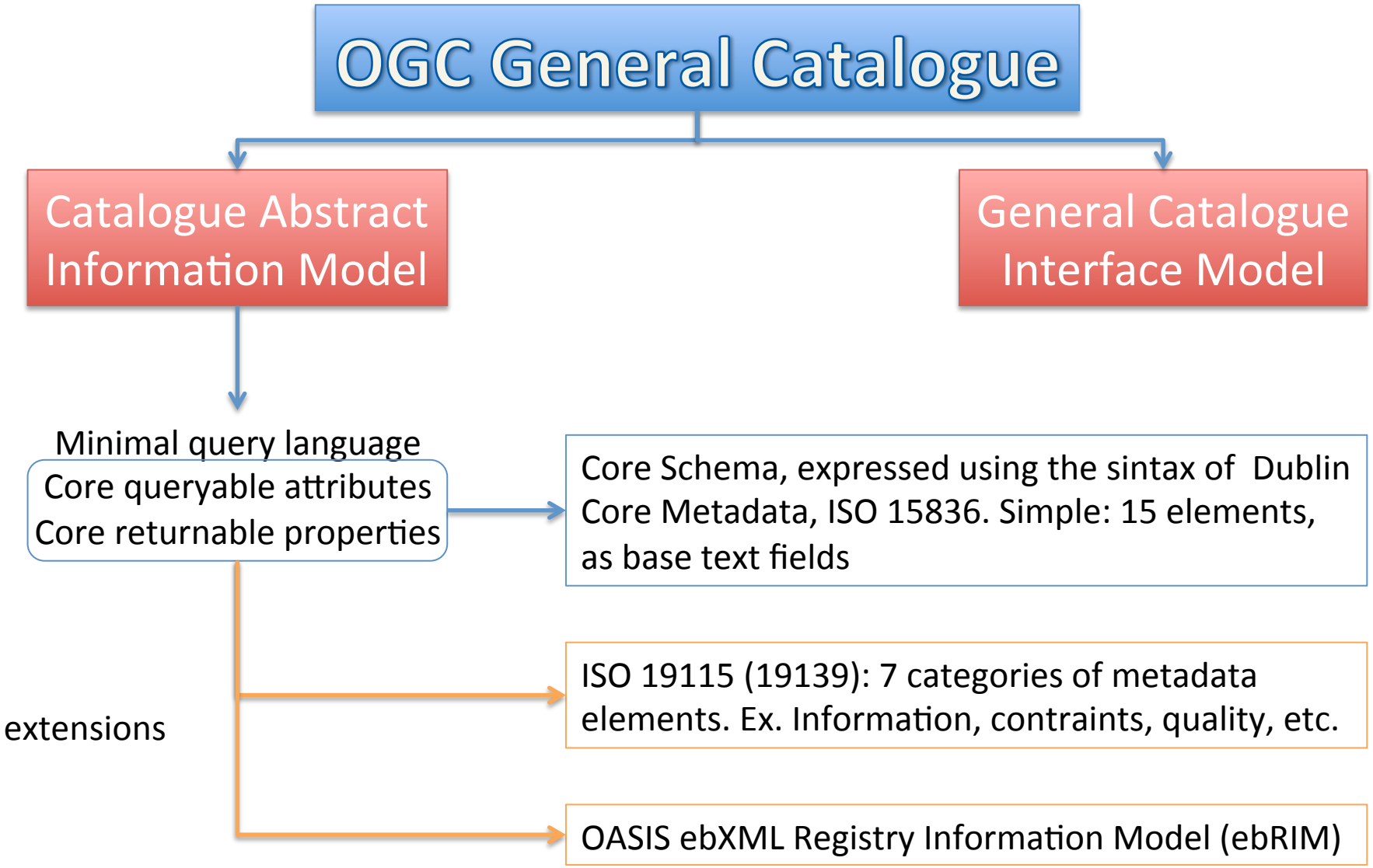
Standards, SOA and semantics are the evolution of this approach aimed at interoperability!



Understanding based on community knowledge



Service scenario for a geospatial catalogue. Source: [Bai and Di, 2011]

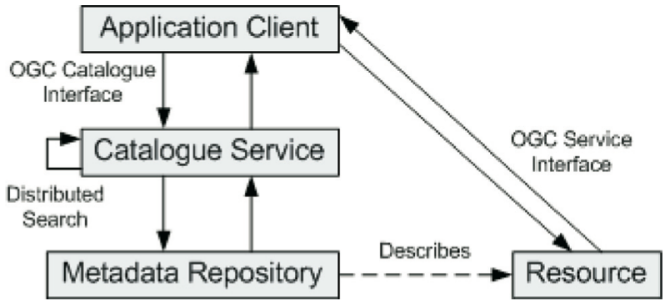

OGC General Catalogue
Catalogue Abstract Information Model
General Catalogue Interface Model
Minimal query language
Core queryable attributes
Core returnable properties
Core Schema, expressed using the sintax of Dublin Core Metadata, ISO 15836. Simple: 15 elements, as base text fields
extensions
ISO 19115 (19139): 7 categories of metadata elements. Ex. Information, contraints, quality, etc.
OASIS ebXML Registry Information Model (ebRIM)

OGC General Catalogue

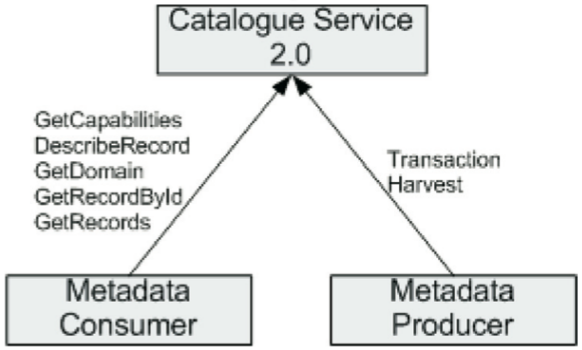
Catalogue Abstract Information Model

General Catalogue Interface Model

Reference Model Architecture



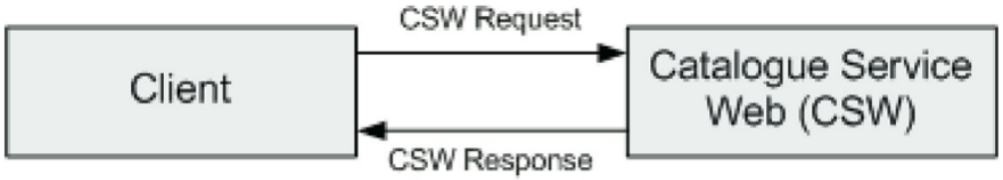
Conceptual Architecture



General interfaces can be bound to several application protocols

HTTP
Z39.50
CORBA

HTTP protocol binding (Catalogue Services for the Web, CSW)



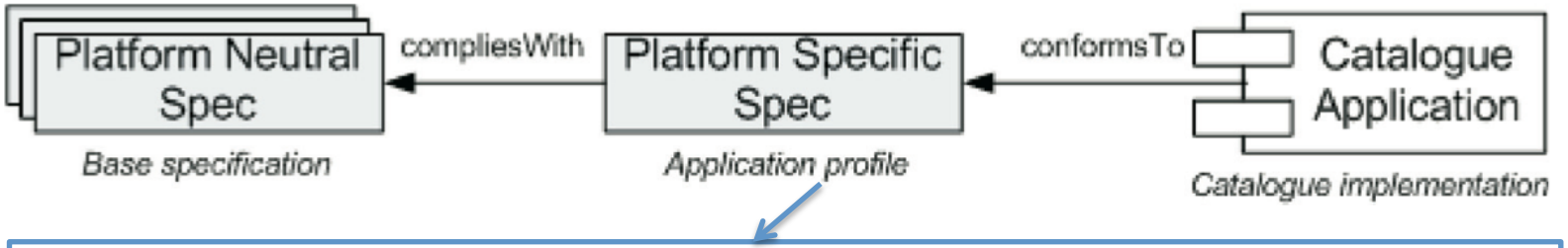

```
http://localhost:8080/deegree-csw/services?REQUEST=GetCapabilities&version=2.0.2&
service=CSW
```

```
http://localhost:8080/deegree-csw/services?REQUEST=DescribeRecord&version=2.0.2&
service=CSW&TypeName=gmd:MD_Metadata
```

CSW (CS/W)

```
<?xml version="1.0" encoding="UTF-8" ?>
<csw:GetRecords service="CSW" version="2.0.2" outputFormat="application/xml"
 outputSchema="http://www.isotc211.org/2005/gmd" resultType="RESULTS"
 xmlns:csw="http://www.opengis.net/cat/csw/2.0.2"
 xmlns:ogc="http://www.opengis.net/ogc"
 xmlns:apiso="http://www.opengis.net/cat/csw/apiso/1.0">
 <csw:Query typeNames="gmd:MD_Metadata">
     <csw:ElementSetName>full</csw:ElementSetName>
     <csw:Constraint version="1.1.0">
       <ogc:Filter>
         <ogc:PropertyIsEqualTo>
           <ogc:PropertyName>apiso:title</ogc:PropertyName>
           <ogc:Literal>Cadastral municipalities</ogc:Literal>
         </ogc:PropertyIsEqualTo>
```

Application Profiles



To address particular needs or requirements
Specify conformance tests
Use of an application-layer protocol such as HTTP

OpenGIS Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile (1.0.0)
CSW-ebRIM Registry Service – Part 1: ebRIM profile of CSW (1.0.1)
Catalogue Services Standard 2.0 Extension Package for ebRIM Application Profile: Earth Observation Products

Some OS Implementations for CSW

Web based geographic metadata catalogue application. It implements the ISO19115/19139 Geographic Metadata, Z39.50, CSW 2.0.2 and OGC WMS standards among others

Software package that implements the OGC Catalogue Service Implementation Specification 2.0.2 and ISO 19115/19119 Application Profile 1.0.0

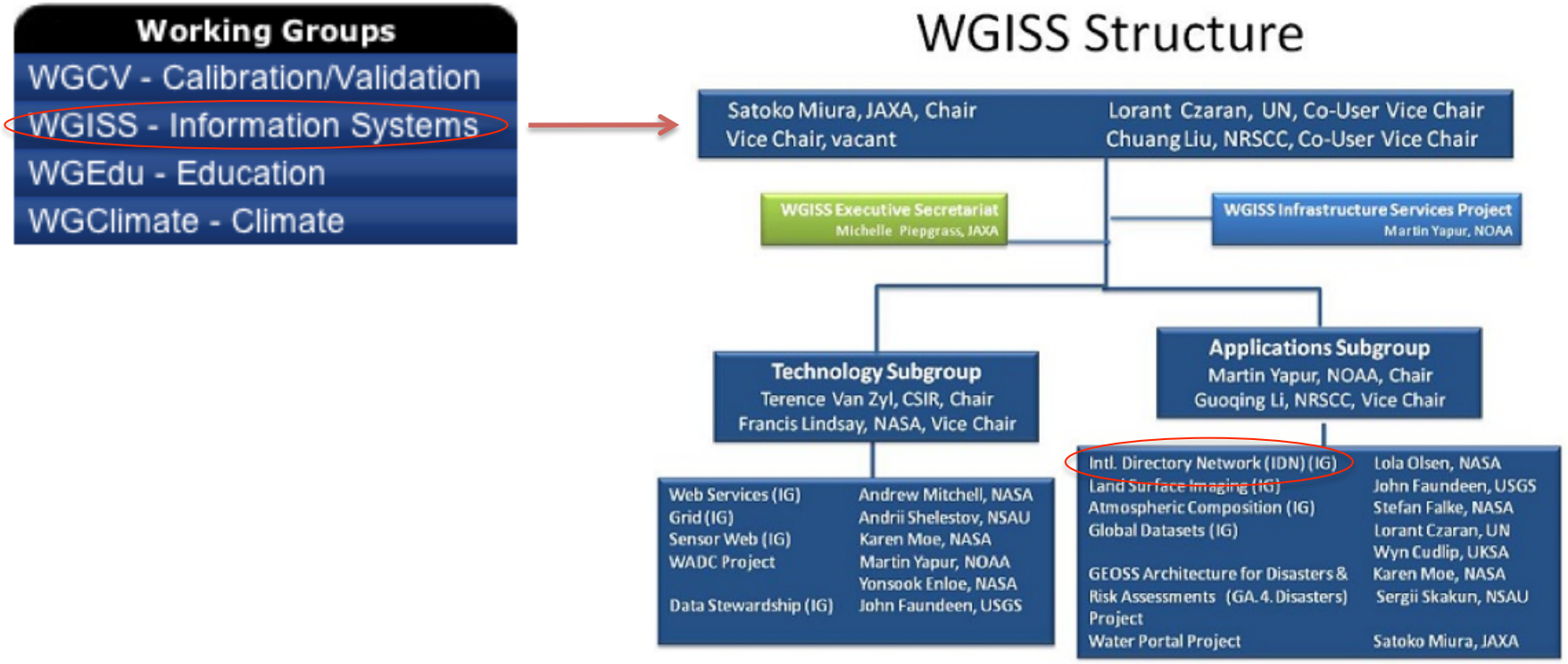
RELEVANT INTERNATIONAL PROJECTS

CEOS Committee on Earth Observation Satellites

Coordinates civil space-borne observations of the Earth. Participating agencies strive to enhance international coordination and data exchange and to optimize societal benefit

Working Groups

- WGCV - Calibration/Validation
- WGISS - Information Systems
- WGEdu - Education
- WGClimate - Climate



NASA GCMD

The Global Change Master Directory (GCMD) holds more than 25,000 Earth science data set and service descriptions, which cover subject areas within the Earth and environmental sciences

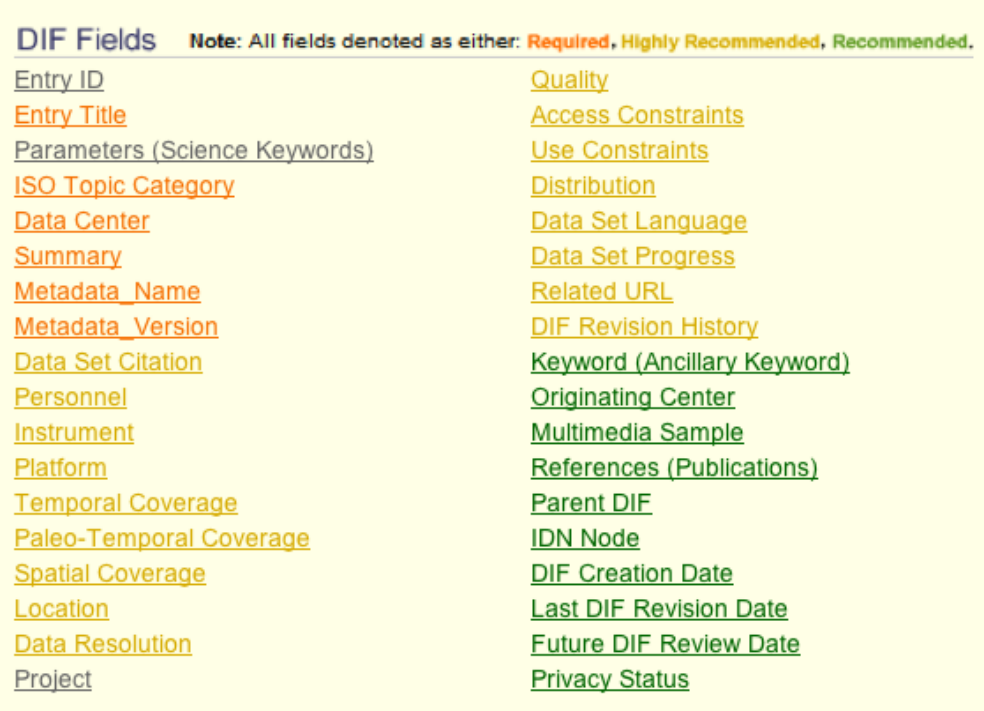
http://gcmd.nasa.gov/ (1989). Users can search through the Directory's website using controlled keywords, free-text searches, map/date searches or any combination of these. Users may also search or refine a search by data center, location, instrument, platform, project, or temporal/spatial resolution.

The project also serves as one of NASA's contributions to the international Committee on Earth Observation Satellites (CEOS), through which it is known as the CEOS International Directory Network (IDN)

CEOS IDN - DIF

The Directory Interchange Format (DIF) (1987) is the "container" for the metadata elements that are maintained in the IDN database, where validation for mandatory fields, keywords, personnel, etc. takes place.

DIF has full ISO compatibility, since it contains all the elements required by the ISO 19115/TC211 geospatial metadata standard



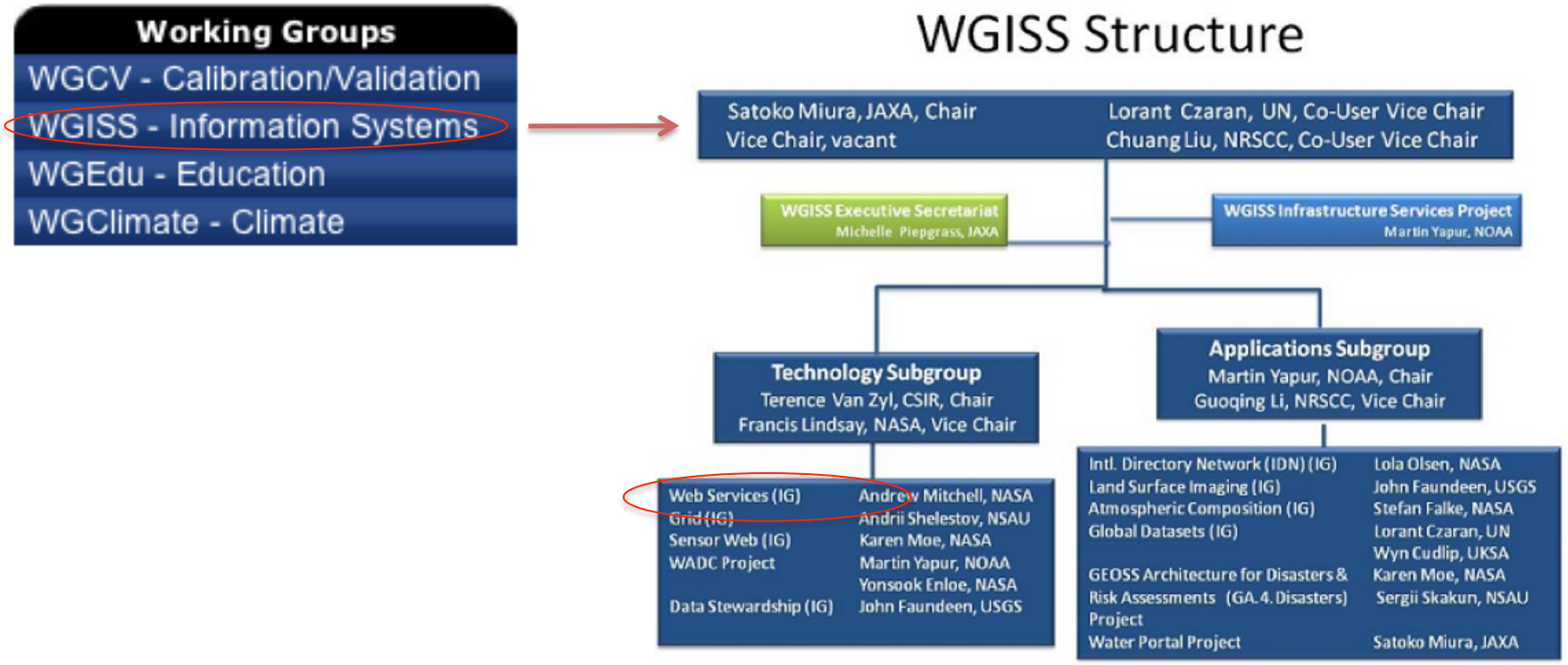
DIF Fields **Note:** All fields denoted as either: Required, Highly Recommended, Recommended.

Entry ID	Quality
Entry Title	Access Constraints
Parameters (Science Keywords)	Use Constraints
ISO Topic Category	Distribution
Data Center	Data Set Language
Summary	Data Set Progress
Metadata_Name	Related URL
Metadata_Version	DIF Revision History
Data Set Citation	Keyword (Ancillary Keyword)
Personnel	Originating Center
Instrument	Multimedia Sample
Platform	References (Publications)
Temporal Coverage	Parent DIF
Paleo-Temporal Coverage	IDN Node
Spatial Coverage	DIF Creation Date
Location	Last DIF Revision Date
Data Resolution	Future DIF Review Date
Project	Privacy Status

http://idn.ceos.org

CEOS Committee on Earth Observation Satellites

Coordinates civil space-borne observations of the Earth. Participating agencies strive to enhance international coordination and data exchange and to optimize societal benefit



CEOSS WGISS Integrated Catalog (CWIC)

CWIC provides an access point for major CEOS agency catalog systems

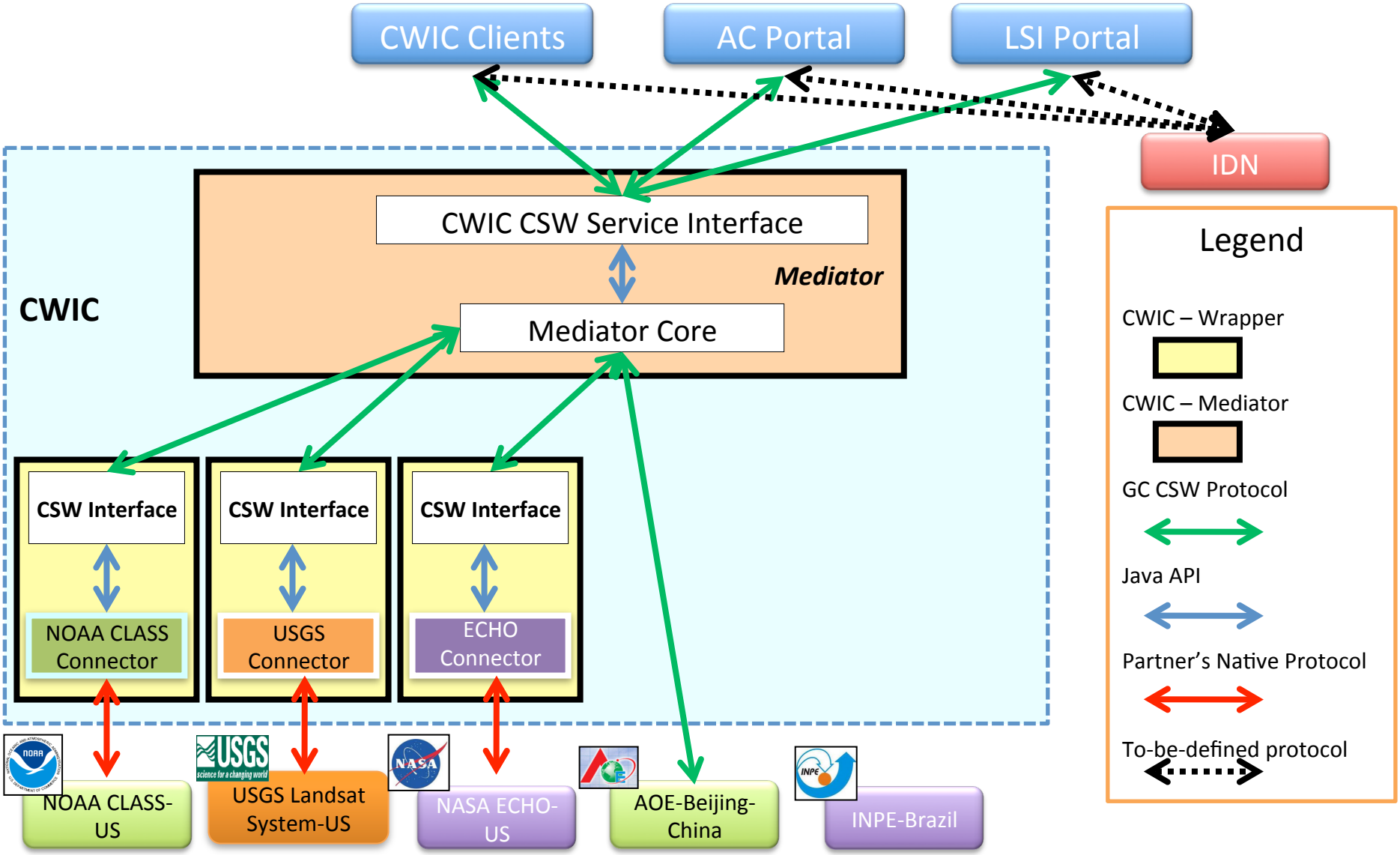
CWIC distributes inventory/product searches to the CEOS agency inventory systems using the agency systems native protocol

CWIC employs a mediator-wrapper architecture to fulfill distributed searches

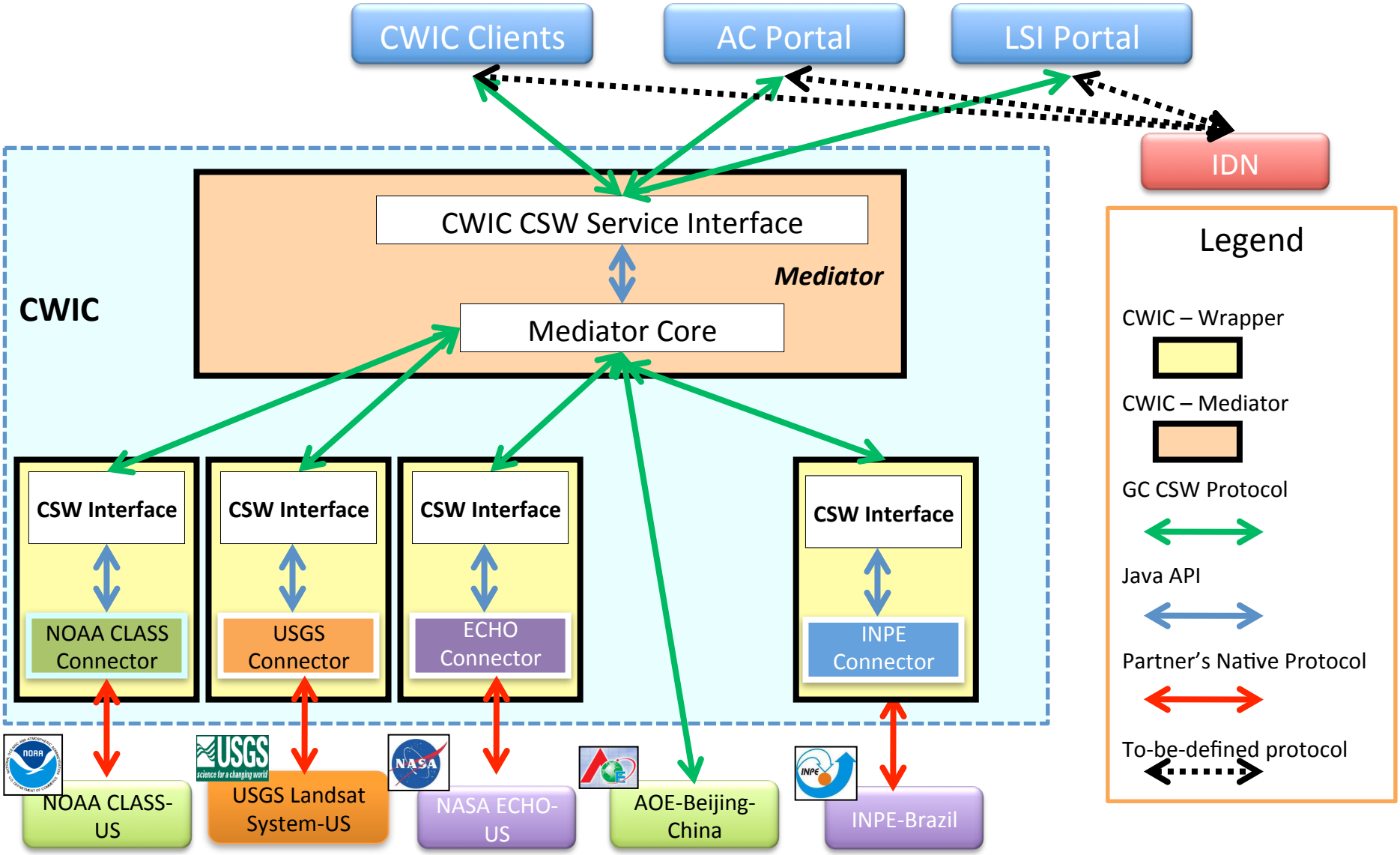
CWIC exposes an OGC CSW 2.0.2 Core profile/ ISO profile-compliant interfaces on the front-end

Extensions to the OGC CSW 2.0.2 are designed and utilized in the CWIC front-end interface

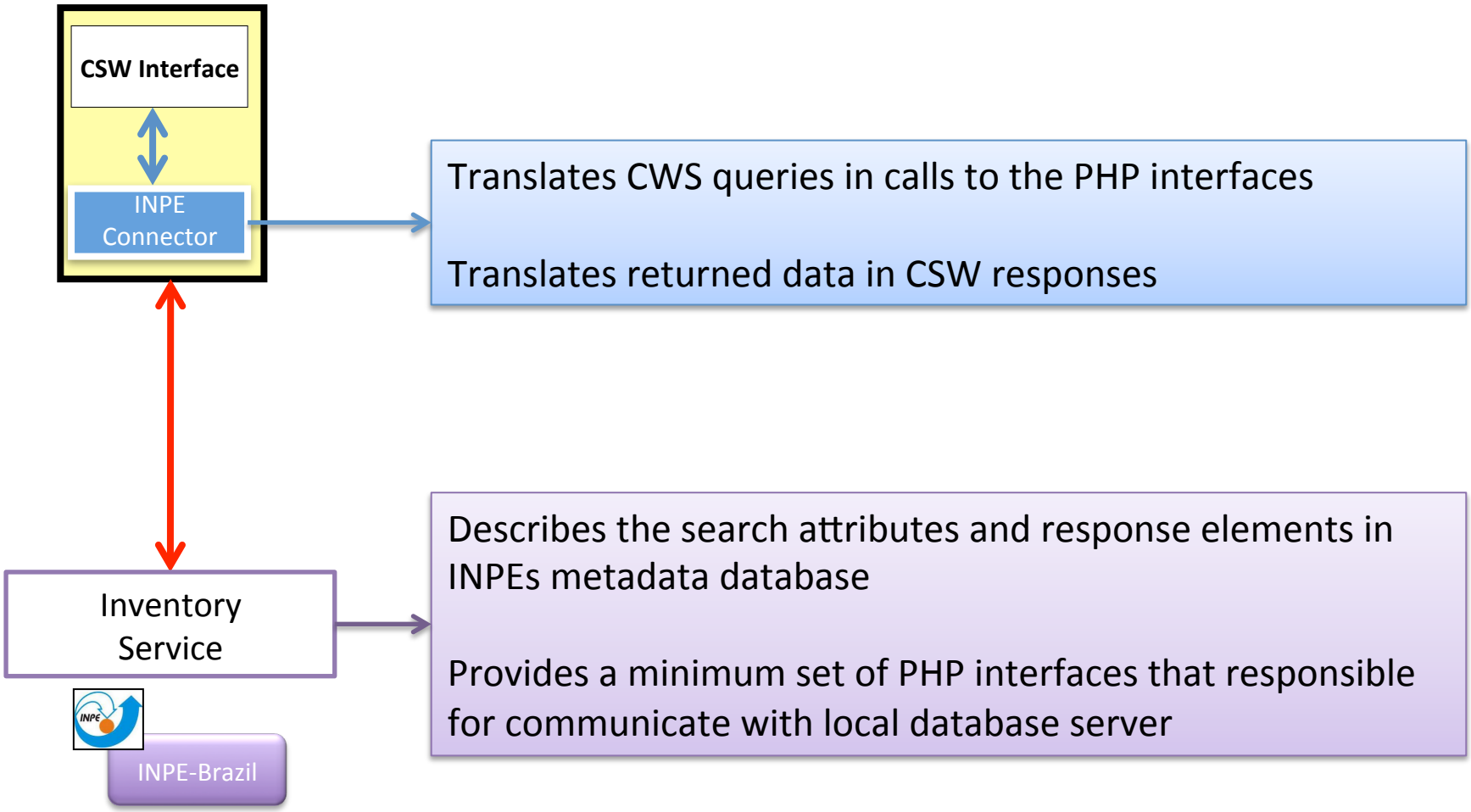
CWIC Context Diagram



CWIC Context Diagram



INPE Connector



CWIC Demo

http://cwic.csiss.gmu.edu:9002/testpage.htm

CWIC will provide a listing of all dataset ids and all data sources reachable from its Mediator in its Capabilities document. Scenario:

1. Keyword search against GCMD using PROJECT=CWIC.
2. Response is a collection of matching datasets, identified by DIF Entry_ID and dataset name.
3. Retrieve CWIC Capabilities document from the CWIC server. Human-readable titles and identifiers will be available in the Capabilities document for each CWIC-accessible dataset.
4. Submit GetRecords request to CWIC to find matching data granules in brief records.
5. Select desired granule by ID from GetRecordsResponse.
6. Submit GetRecordById request for full record, using granule ID.
7. Get full record for granule of interest in GetRecordByIdResponse.
8. Retrieve data granule from host system using the URL in the full record response.

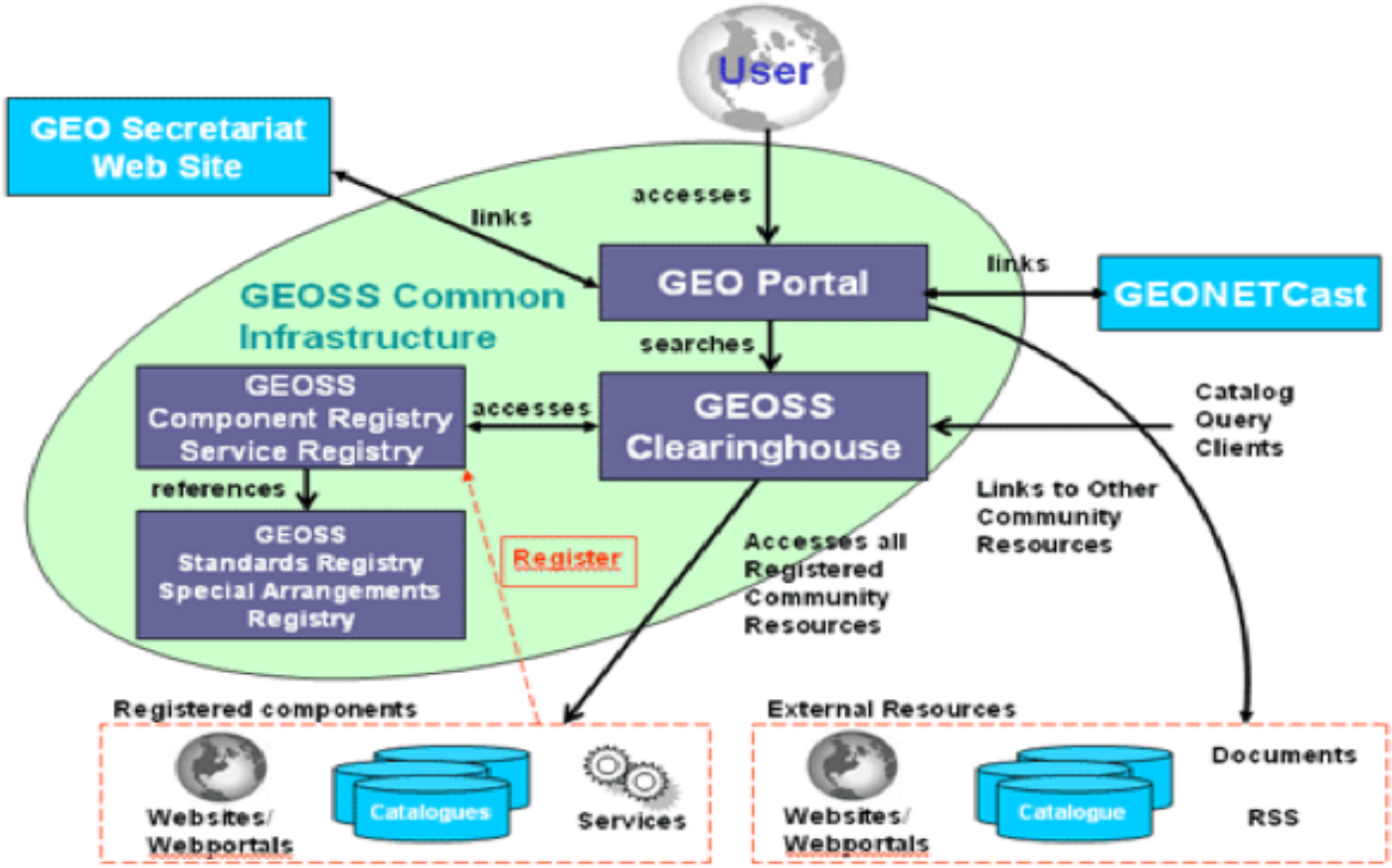
GROUP ON EARTH OBSERVATIONS

The Group on Earth Observations is coordinating efforts to build a Global Earth Observation System of Systems, or GEOSS

The Global Earth Observation System of Systems will provide decision-support tools to a wide variety of users.

As with the Internet, GEOSS will be a global and flexible network of content providers allowing decision makers to access an extraordinary information at their desk

The GEOSS Common Infrastructure



http://www.geoportal.org/web/guest/geo_home

CEOS/IDN Issue 32 September 2011

The IDN Adapts Earth Science Data Set Contributions for US GEO's GEOSS Data Core for Global Discovery

By Scott Ritz, GCMD Science Team Leader

The IDN plays an important role in making Earth science data sets discoverable by future scientists (students)—not just scientists throughout world. Through the IDN's Catalog Services for the Web (CSW), 20,985 metadata records in ISO-19115 format are available to the GEOSS portal. A subset of these records (contributed to the GCMD by US GEO partners) has been "tagged" to assure users that the particular data set is accessible free-of-charge or at-cost, according to the GEOSS Data-CORE guidelines. GEOSS Data CORE translates to "Data **C**ollection of **O**pen **R**esources for **E**veryone". A Data-CORE compliant data set according to the GEOSS Data Sharing Plan is one that is "contributed by the GEO community on the basis of full and open exchange (at no more than the cost of reproduction and distribution) and unrestricted access". When searching the GEOSS portal, users may perform specific searches for Data CORE-compliant data sets using these "tags". The "tags", as established by the GEOSS Data Sharing Task Force (DSTF), are "GEOSS Data Core" and "GEOSS No Monetary Charge", respectively. There are currently 11,075 US GEO records in the IDN CSW collection that have been identified and tagged as Data CORE-compliant. The US GEO records, characterized by the Data CORE "tags" within, are currently available in the GEOSS Clearinghouse through the IDN.

Source: CEOS IDN Newsletter, September 2011

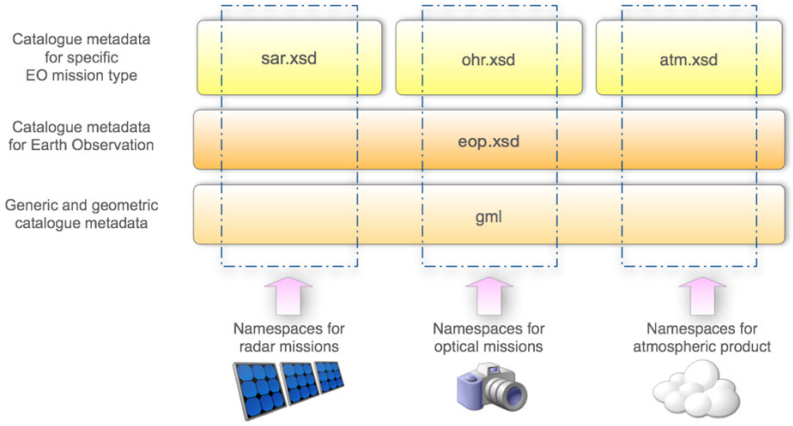
The objective of HMA - Heterogeneous Mission Accessibility is to establish harmonized access to heterogeneous EO missions' data from multiple mission G/S's, including national missions and ESA Sentinel missions

The HMA project focused on the definition of 5 interfaces:

1. **Catalogue service**, to browse and retrieve metadata on collections and products available across the HMA collaborating catalogues
2. **Ordering service**, to order products identified in a catalogue
3. **Programming Service**, to place request for new acquisitions onto the HMA partner missions' ground segments
4. **Mission Planning Service**, to facilitate programming requests preparation
5. **Online Data Access Services**, to retrieve products from the online access archives offered through the HMA

HMA

HeterogeneousMissionsAccessibility



EO Profile Schema

Standards

OGC 06-080 GML 3.1.1 Application schema for Earth Observation products

OGC 06-131 OGC Catalogue Services Specification 2.0 Extension Package for ebRIM (ISO/TS 15000-3) Application Profile

OGC 06-141 Ordering Services for Earth Observation Products

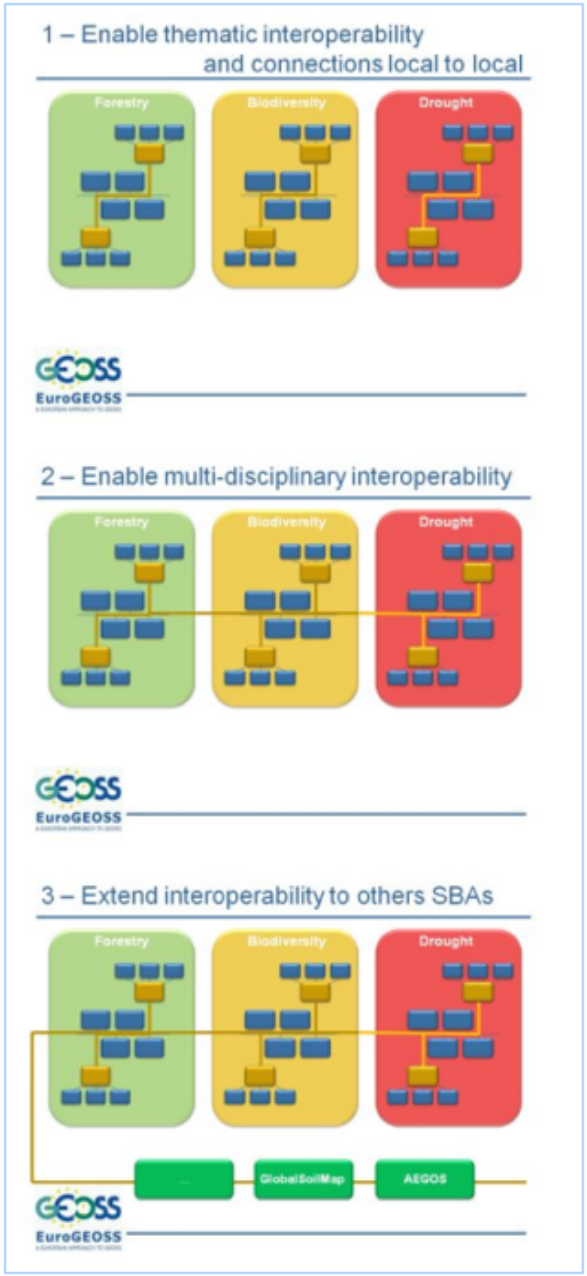
OGC 07-018 Opengis Sensor Planning Service Application Profile for EO Sensors

http://www.geoportal.org/web/guest/geo_clearinghouse

EuroGEOSS is a large scale integrated project in the Seventh Framework programme of the European Commission. It is part of the thematic area: "ENV. 2008.4.1.1.1: European environment Earth observation system supporting INSPIRE and compatible with GEOSS"

The EuroGEOSS project adopted a Brokering Approach to implement multi-disciplinary interoperability and lower entry barriers for both Users and Data Providers.

According to such an approach, Users and Data Providers are not asked to implement any specific interoperability technology but to continue using their tools and publishing their resources according their standards as much as possible.



The EuroGEOSS Broker

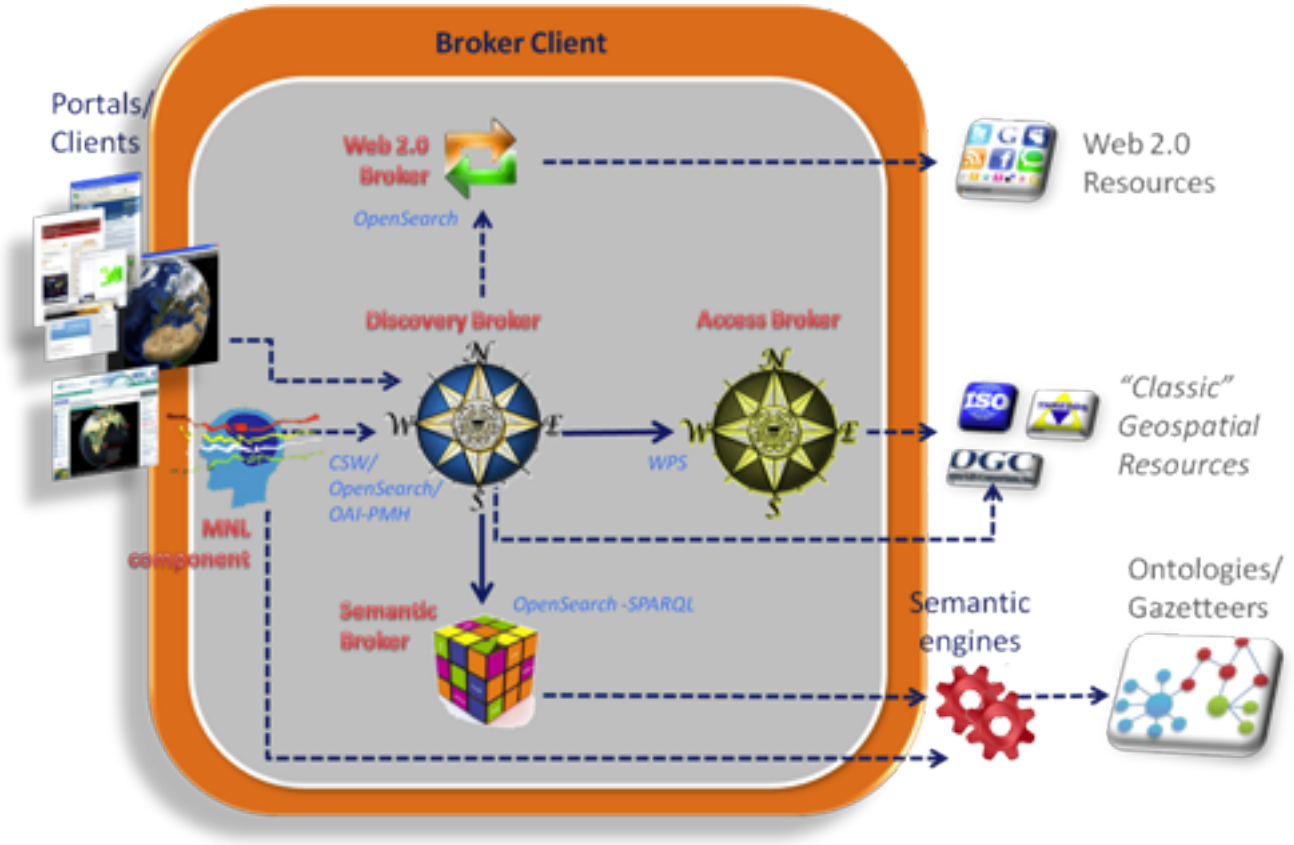
The EuroGEOSS project adopted a Brokering Approach to implement multi-disciplinary interoperability and lower entry barriers for both Users and Data Providers.

According to such an approach, Users and Data Providers are not asked to implement any specific interoperability technology but to continue using their tools and publishing their resources according their standards -as much as possible.

In keeping with the Brokering Approach, EuroGEOSS has developed a Brokering framework (i.e. a set of Brokers) to bind the heterogeneous resources published by the Data Providers, adapting them to the tools commonly utilized by the Users.

EuroGEOSS developed a Broker for each main interoperability functionality, namely: Discovery Broker, Access Broker, Semantic Broker

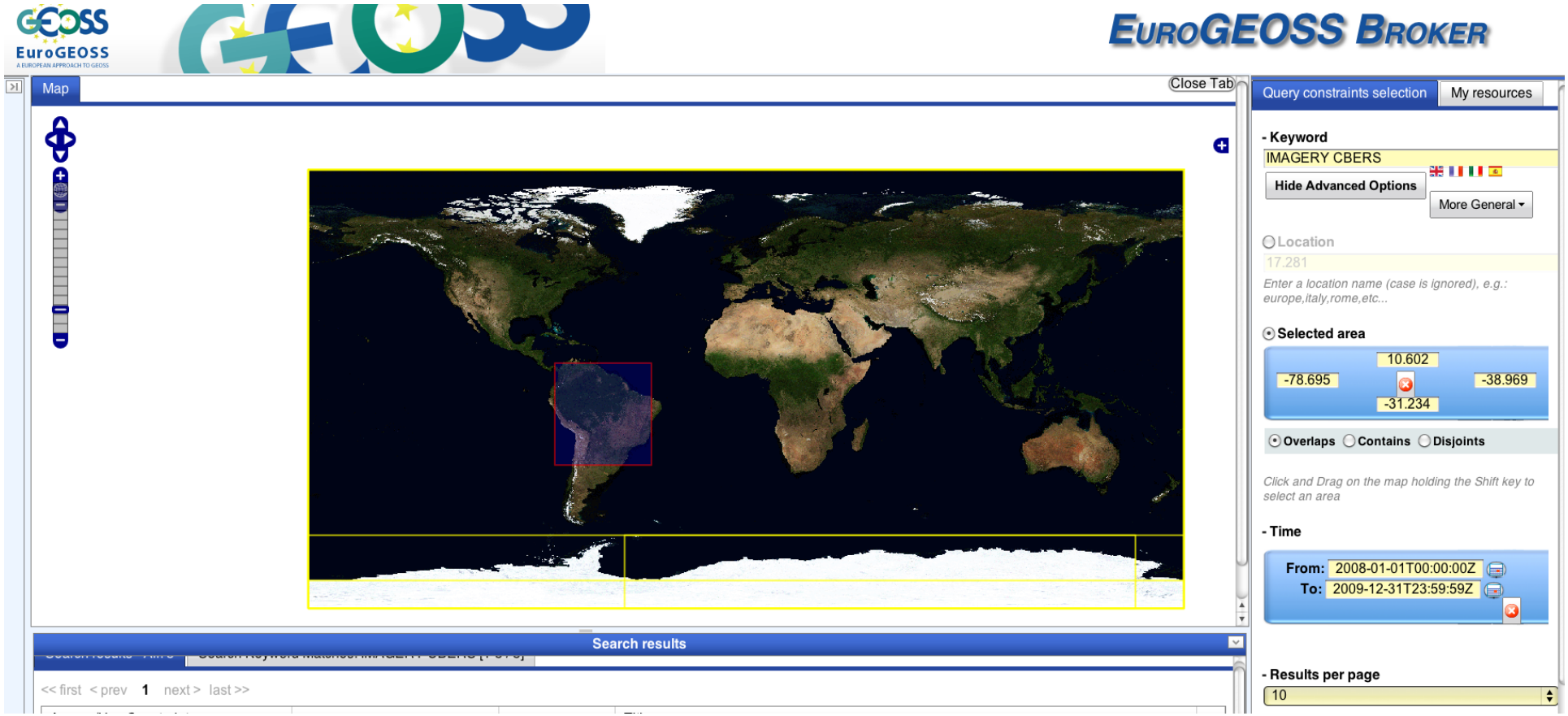
The EuroGEOSS Broker



The EuroGEOSS Brokering framework is empowered by the GI-cat technology

Demo EuroGEOSS Broker

http://www.eurogeoss-broker.eu/

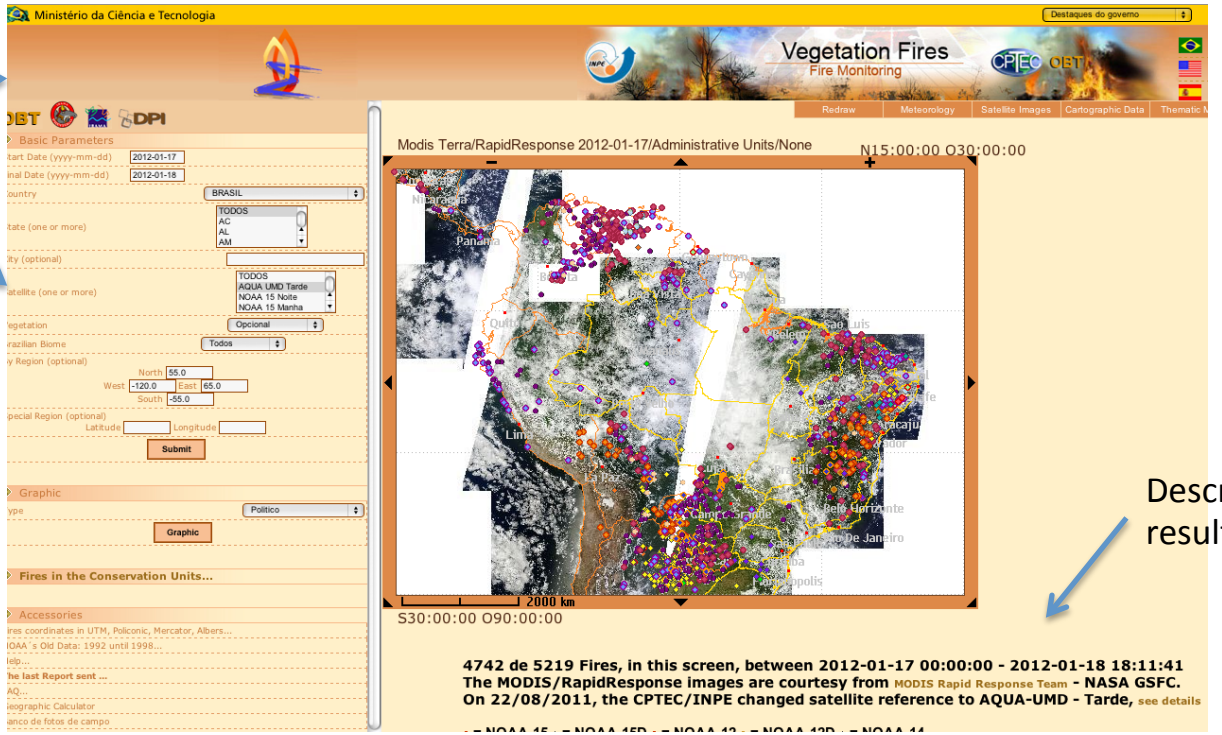


EuroGEOSS Broker and INPE data

In order to initiate a collaboration for use of the EuroGEOSS broker with the INPE data and archives, a little experiment with the fire monitoring is being done

Our web interface

Description of the allowed search parameters



Description of results

EuroGEOSS Broker and INPE data

According to the mapping of the INPE information model to GI-cat one, based on ISO 19115 is described in:

http://zeus.pin.unifi.it/cgi-bin/twiki/view/GIcat/INPEMapping

A GI-cat connector was developed, so that a preliminary prototype of the EuroGEOSS broker accessing INPE Bdqueimadas web interface is now available for testing at

http://giida.cnr.it/gi-cat-inpe/gi-portal/index.jsp

(Demo)

GEOSS interoperability for Weather, Ocean and Water

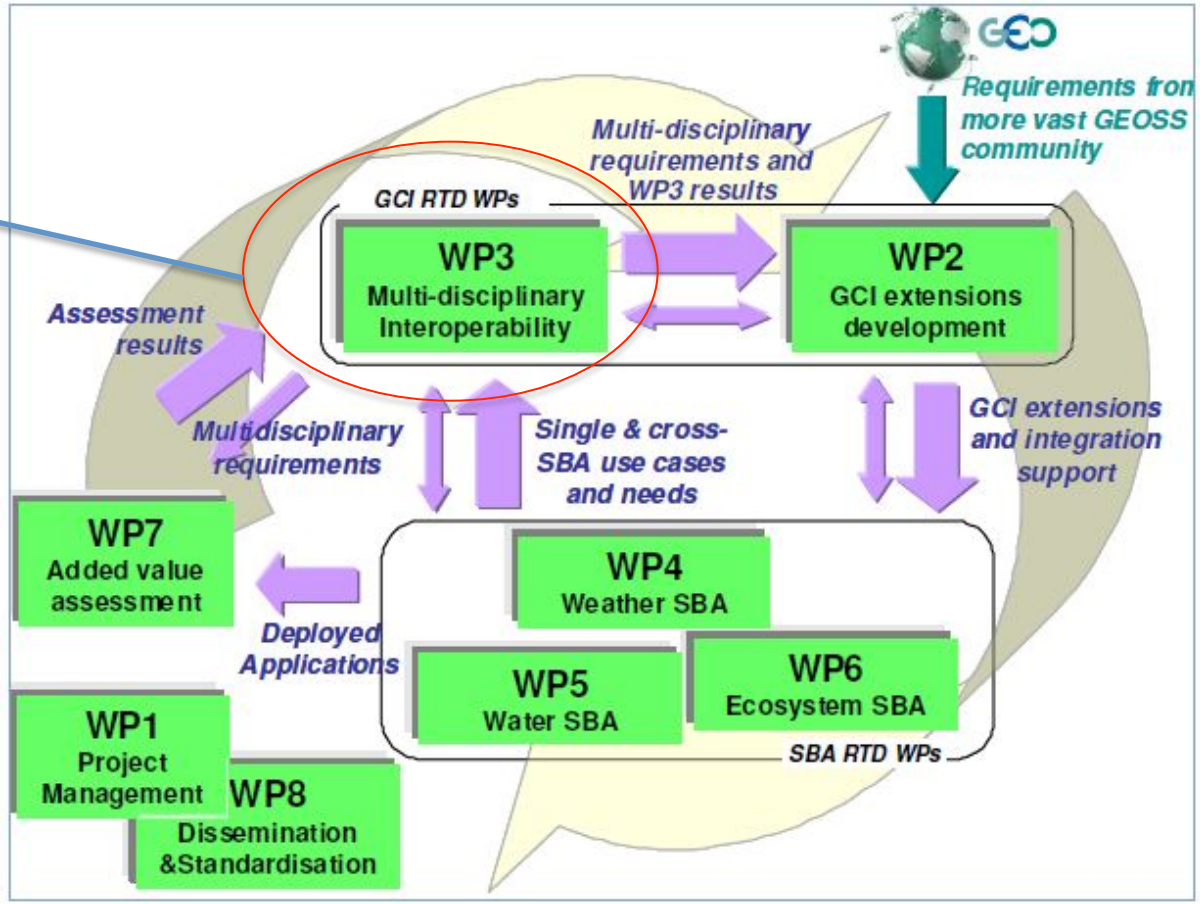
GEOWOW is a project, co-funded under the European Community's Seventh Framework Programme FP7/2007-2013 in response to call ENV.2011.4.1.3-1 “Interoperable integration of Shared Earth Observations in the Global Context”.

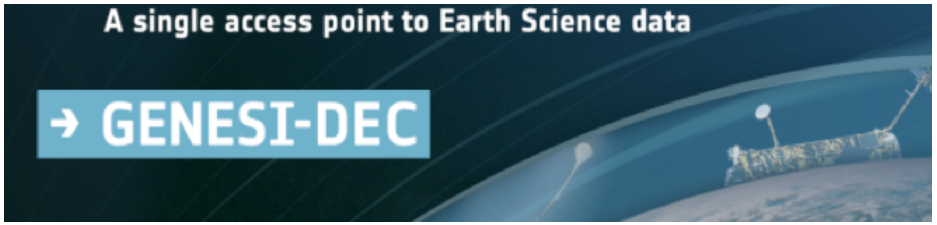
It is implemented by a consortium of 15 partners from Europe, Brazil and Japan and is coordinated by the Italian establishment of the European Space Agency (ESA-ESRIN) in Frascati.

The GEOWOW Portal is foreseen to give access to a vast amount of Geo-information and related resources. Enabling users to query, discover and access data in a simple/ efficient way, it will evolve GEOSS and in particular the GEOSS Common Infrastructure (GCI).

GEOSS interoperability for Weather, Ocean and Water

INPE is suppose to collaborate with 50 person/month (document)





A project funded by the EC to establish open data and services access, allowing European and worldwide Digital Earth Communities to seamlessly access, produce and share data, information, products and knowledge. This will create a multi-dimensional, multi-temporal, and multi-layer information facility of huge value in addressing global challenges such as biodiversity, climate change, pollution and economic development.

http://portal.genesi-dec.eu/search/

OpenSearch was created by Amazon's A9.com and is a simple means to interface to a search service by declaring a URL template and returning a common syndicated format

There is support for search operation control parameters, but no constraints are placed on the query string which is regarded as opaque

OpenSearch is thus a means to interface to arbitrary search APIs (both standard and proprietary) and to retrieve results using a common list-based format

OpenSearch is built as an extension on top of the wildly popular RSS 2.0 XML schema

OpenSearch is a plug-and-play technology

OpenSearch

In the context of GEOWOW: services to transform OGC Catalogues in OpenSearch services so that they can be more easily integrated to browsers and mashup applications:

An OpenSearch service:

http://cwic.genesi-dec.eu/search?series=INPE:A1_MODIS&count=10&startindex=&startdate=2011-01-01&stopdate=2011-01-10&bbox=-83,-20,-54,0

Embedding search in a mashup application:

http://maps.google.com/maps?q=http:%2F%2Fcwic.genesi-dec.eu%2Fsearch%3Fseries%3DINPE:A1_MODIS%26count%3D10%26startindex%3D%26startdate%3D2011-01-01%26stopdate%3D2011-01-10%26bbox%3D-83,-20,-54,0&hl=en&sll=-5.02475,-60.54175

CONCLUSIONS AND DISCUSSION

CWIC

HMA
HeterogeneousMissionsAccessibility

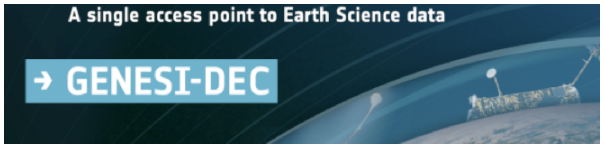
A single access point to Earth Science data
GENESI-DEC

GEOSS
EuroGEOSS
A EUROPEAN APPROACH TO GEOSS

GEOWOW

Aims at being contributions to GEOSS
Recognize the social benefit areas and aims at interoperability between them
Mostly based on catalogues, federation of catalogues and metadata standards
Recognize the community portals
There is some overlapping between goals and methodology

CWIC



EO products, mostly satellite imagery

Semantics
Simpler interfaces

CWIC

A single access point to Earth Science data

→ GENESI-DEC

EuroGEOSS

A EUROPEAN APPROACH TO GEOSS

GEOWOW

INPE as data provider

INPE as research partner to discuss new strategies, methodologies, projects, chairing committees and organizing meetings.

Conclusions

INPE has being doing a great job in several aspects:

- Useful community portals : PRODES, BDQUEIMADAS, CPTEC, etc.
- Maintenance of organized metadata and data repositories
- Definition of clear access policies: FREEDON
- Research and technology development capacity: Terra*, LAF, etc.

However…

- We have little commitment to the standards generally accepted
  - Other people are developing connectors to our databases

- We lack internal coordination in order to enhance our interoperability capacities
- We are not participating actively of the top of the moment discussions like semantics, remote processing, etc.

A small step: lets create our catalogues servers.

References

Bai, Y. & Di, L. Providing access to satellite imagery through OGC catalog service interfaces in support of the Global Earth Observation System of Systems. *Computers & Geosciences* **37**, 435-443 (2011).

Govedarica, M.; Bošković, D., Petrovački, D.; Ninkov, T.; Ristić, A.

Metadata Catalogues in Spatial Information Systems, *Geodetski list* v. **64**.

(87), No.4, 313-33 (2010).

THANK YOU!