

THE GEOPORTAL COOKBOOK TUTORIAL



Wim Hugo

SAEON/
SAEOS

SCOPE OF DISCUSSION

- Background and Additional Resources
- Context and Concepts
- The Main Components of a GeoPortal Architecture
- Implementation Options
- Freely Available Components
- Build a Live Example

OUR AIM

- It is possible, desirable, and in the public interest to:
 - Ensure that scientific data is described properly and discoverable;
 - Once discovered, its utility, quality, and scope can be understood, even if the data sets are huge;
 - Once understood; it can be accessed *freely and openly*;
 - Once accessed, it can be included into distributed processes, preferably automatically, and on large scales;
 - Once processed, the knowledge gathered can be re-used.

BACKGROUND AND ADDITIONAL RESOURCES

Why does the
opportunity
exist?

Who are the
current
Stakeholders?

BACKGROUND

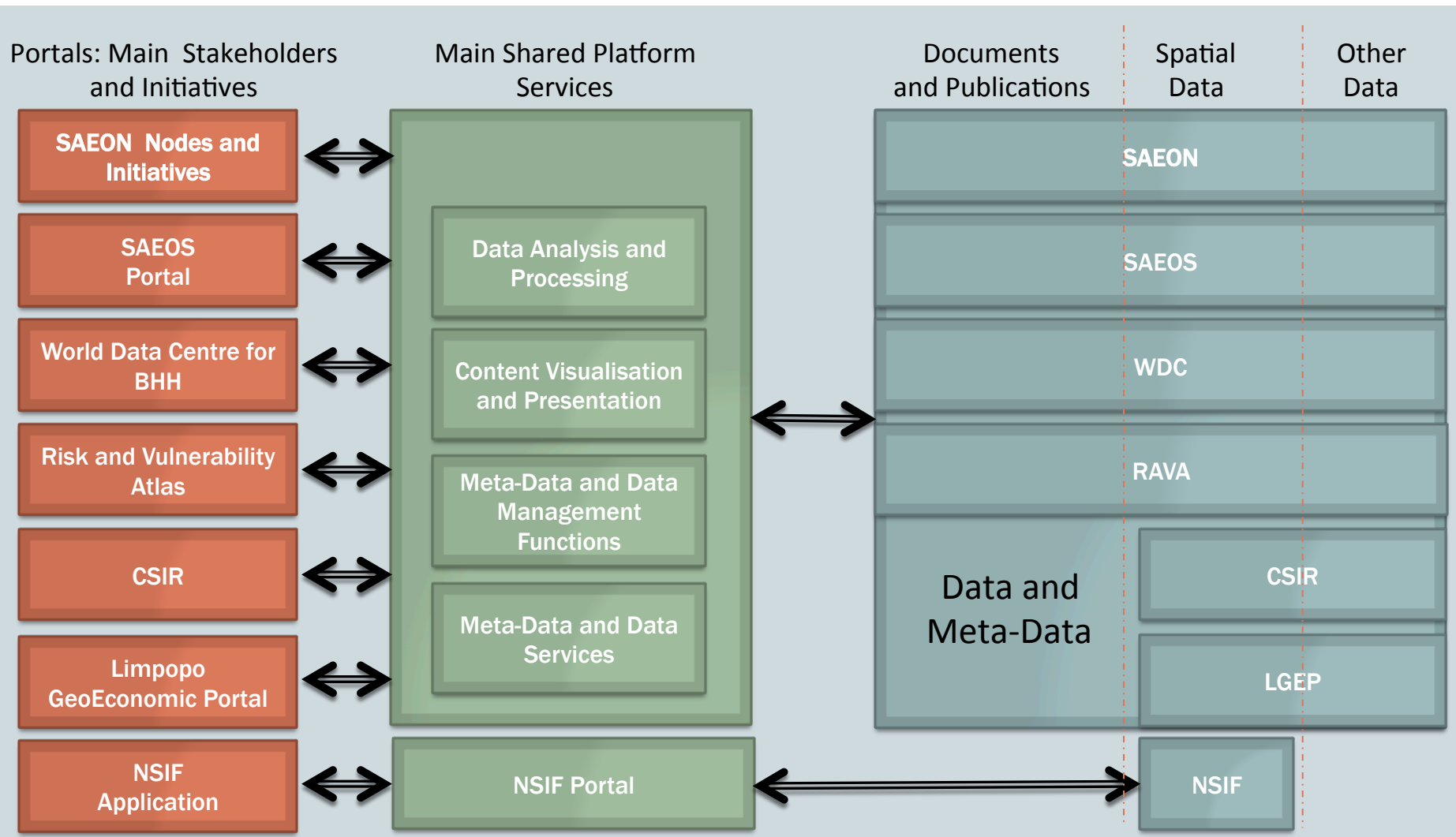
- SAEON and the CSIR has collaborated on GeoPortal infrastructure since 2004/5.
- The list of stakeholders have grown to include DST and several additional smaller initiatives.
- Recent technical developments, which is based on service-oriented architecture, has seen the emergence of loosely integrated components that are 'glued' together with open standards interfaces.
- This means that the components can be re-used In any system and do not rely on a specific implementation to work.

MAJOR CURRENT IMPLEMENTATIONS

- SAEON Data Portal
 - <http://data.saeon.ac.za>
- SAEOS (South African Earth Observation System)
 - <http://saeos.qsens.net>
- Risk and Vulnerability Atlas (SARVA)
 - <http://rava.qsens.net>
- World Data Centre for Biodiversity and Human Health in Africa
 - <http://wdc.qsens.net>
- CSIR GeoPortal
 - <http://geoportal.gsdi.csir.co.za>

The next slide provides an overview of how these initiatives are linked.

PORTALS AND PLATFORM



WHERE CAN I READ MORE?

- On the 'Shared Platform' and how it works:
 - <http://data.saeon.ac.za/documentation/it-governance/policies-and-guidelines/G342.1.0.2%20Guidance%20for%20Data%20Providers%20and%20Custodians.pdf>
- On Standards-Based GeoPortals:
 - <http://www.gsdi.org/gsdicookbookindex>
- On the Global Earth Observation System of Systems (GEOSS):
 - http://data.saeon.ac.za/documentation/it-governance/policies-and-guidelines/07_Implementation%20Guidelines%20for%20the%20GEOSS%20Data%20Sharing%20Principles%20Rev2.pdf

CONTEXT AND CONCEPTS

MetaData

Data
Standards

Automation

Mediation

MAIN CONCEPTS AND REQUIREMENTS

- Standardised meta-data repositories make it possible for users to find, apply, and visualise data. Any GeoPortal needs a way to catalogue its data holdings, and such a repository is a prerequisite for any GeoPortal.
- GeoPortals need to allow users to visualise data, preferably using a web-based map viewer. Not only that: users may also want to overlay data from different providers, manipulate the presentation of layers, and save the resulting map for future use.
- If data sources are standardised and adequately described in meta-data, they can be included into automated processes and analysis where required.
- Increasingly, data need to be linked to visualisation or processing services through additional information that is provided by the user. Structuring and storing these user interventions are referred to as 'Mediation'.

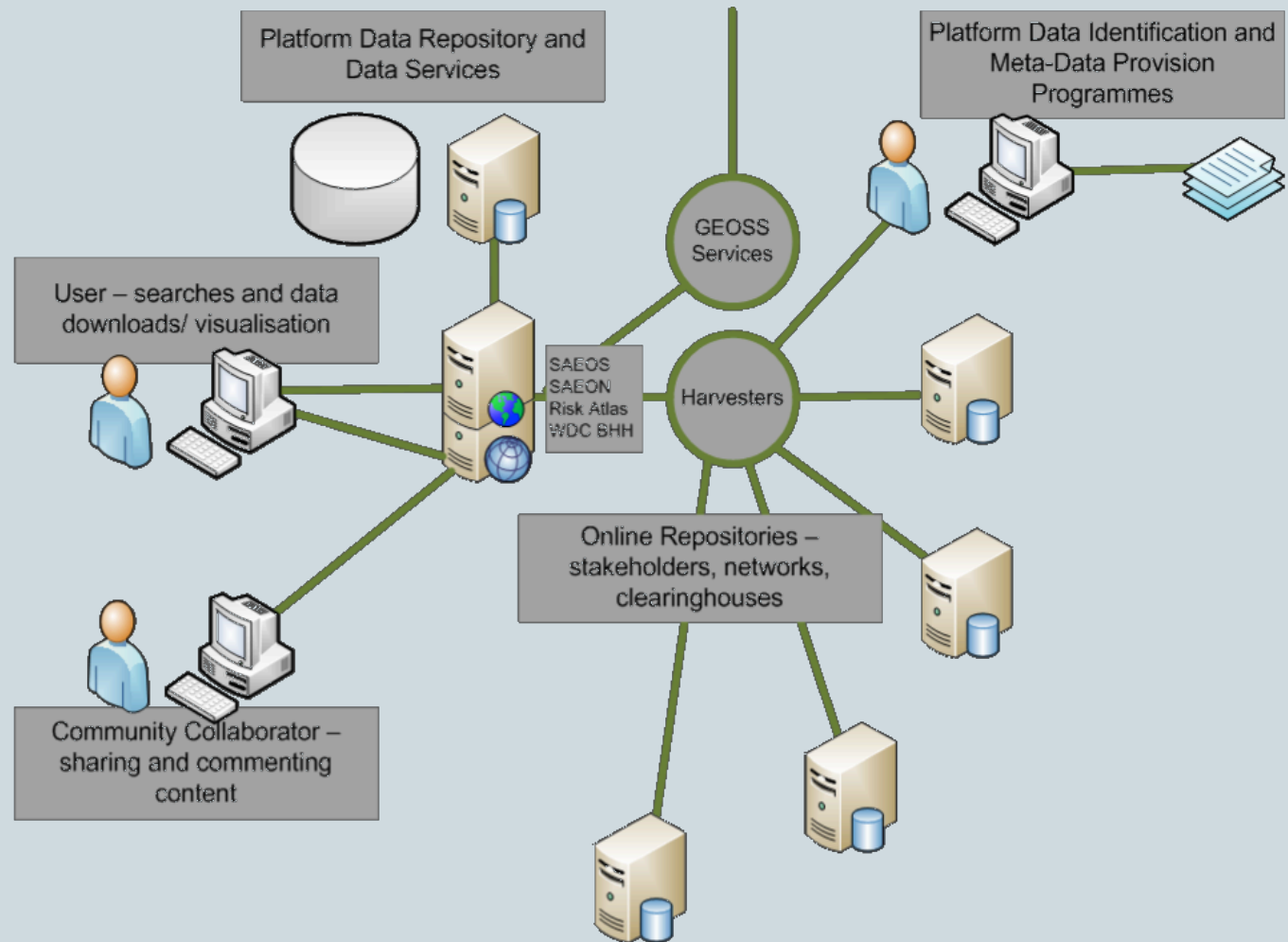
MAIN USE CASES

DISCOVERY AND VISUALISATION

The platform is based on a shared and aggregated meta-data repository, and the meta-data repository is capable of accepting and working with a range of well-established meta-data standards.

These include Dublin Core, SANS 1878, the ISO 19115 family, EML, and FGDC. The list is likely to be extended from time to time to accommodate other standards in widespread use by a user community or new provider.

Harvesters can be used to automate meta-data acquisition from standardised providers.



MAIN USE CASES

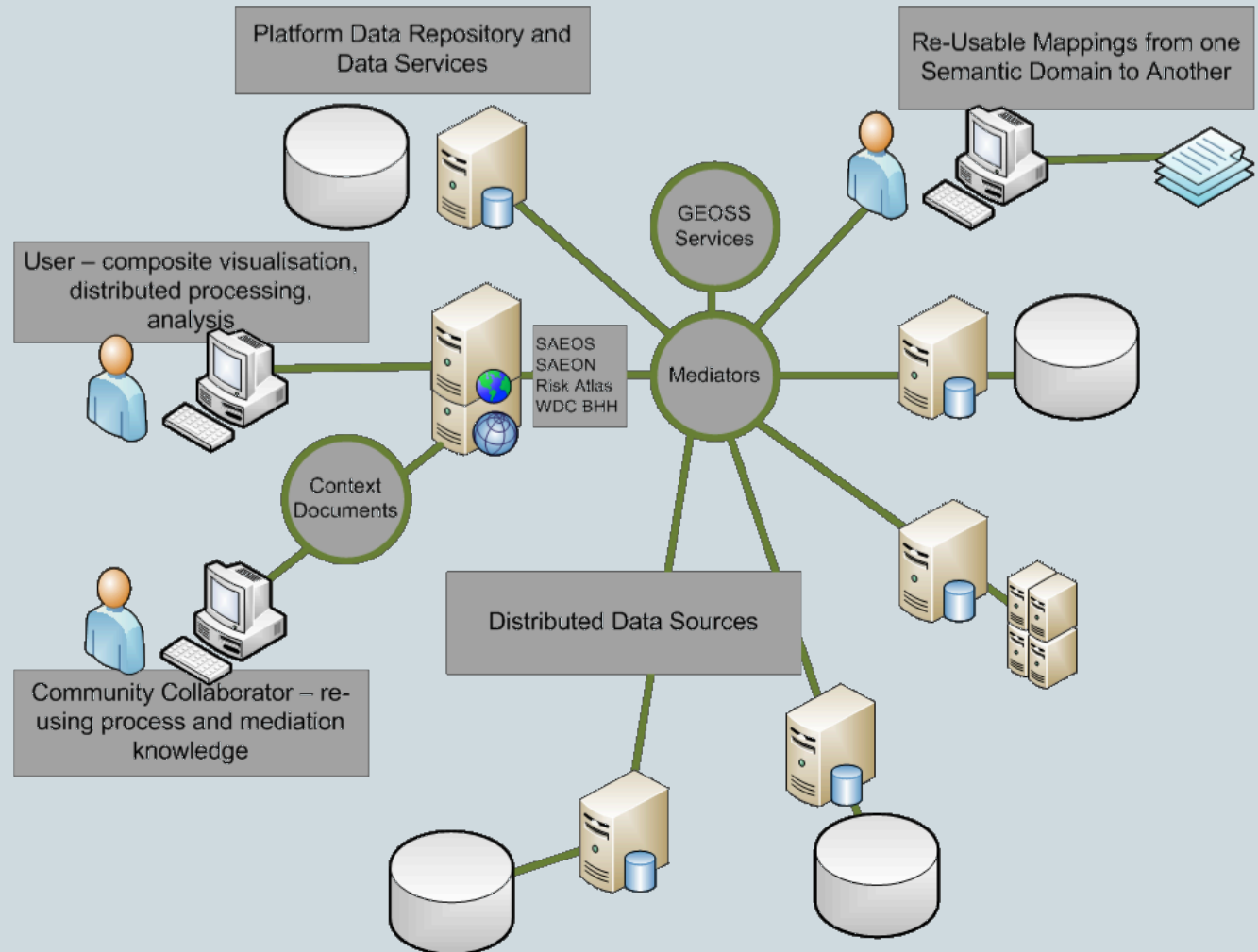
PHASE 2: MEDIATION AND COLLATION

Mediators are required to tailor (bind) a data source or sources to a process or service.

The service can be as simple as a mapping service, and the mediation as rudimentary as renaming the layers to match the context of a map.

More advanced mediations define the linkages between distributed data sets, how data should be aggregated for analysis or display purposes, which columns to match to a charting service, and so on.

Mediations are stored in 'Web Context Documents'.



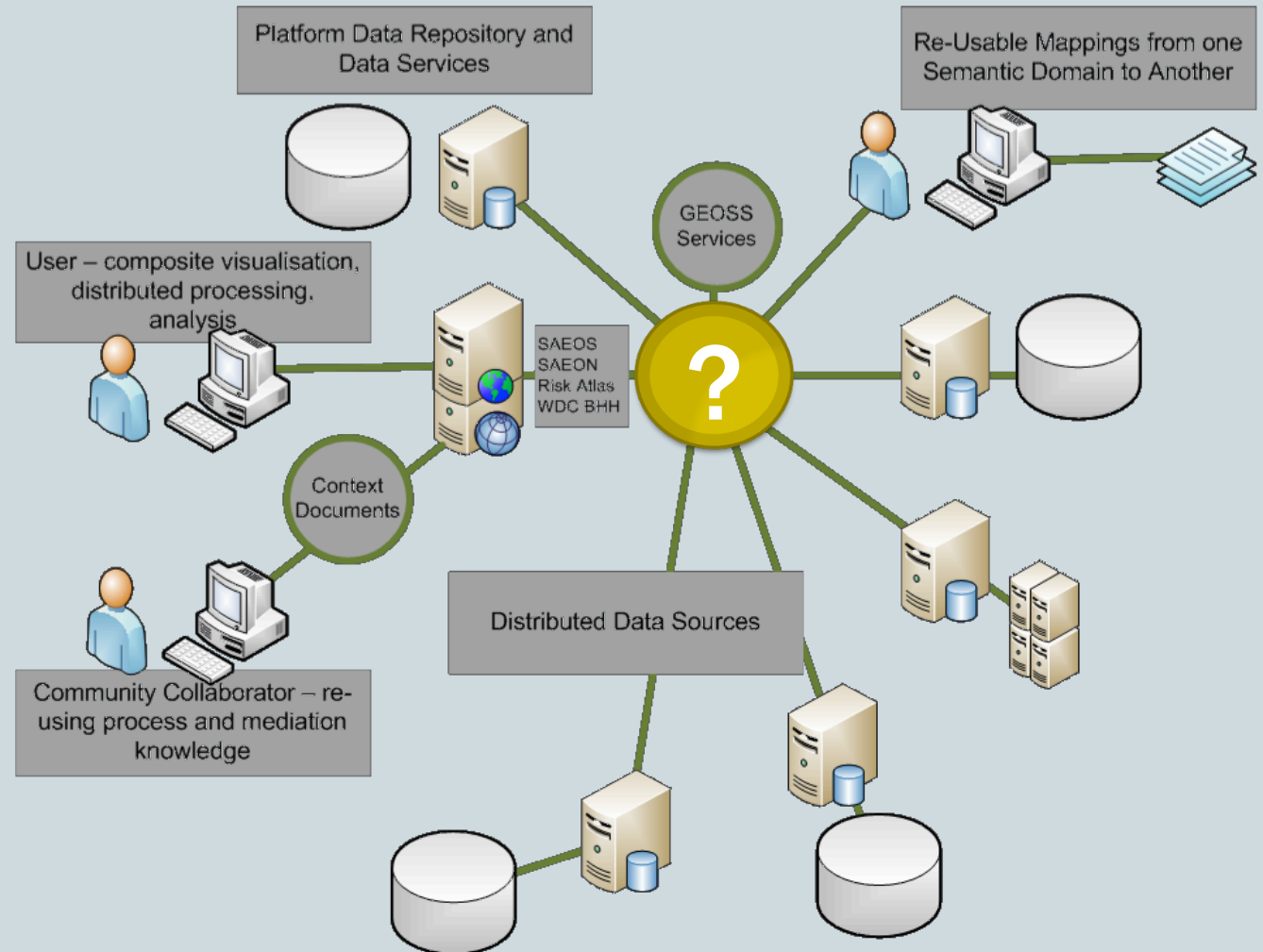
MAIN USE CASES

PHASE 3: AUTOMATED PROCESS CHAINING

Distributed data sources and processes can also be chained together: in such cases, one needs to describe the process in a standardised way for future re-use.

This happens already in e-commerce systems, and it is likely that the same standards in use for automation of web commerce will be used in scientific process automation – but the jury is still out on this.

Automation and mediation rely greatly on ontologies for their efficient functioning. These are agreed dictionaries of terms, and can automate some of the mediation.



NATIONAL PLATFORM

- “Shared Implementation”
 - a framework containing architecture guidelines, abstract specifications and user requirements, recommended interoperability standards, and reference implementations.
 - a directed open source community that contributes to, extends, and maintains the shared implementation.
 - National funding provides a baseline of operations and activity.
 - not domain-specific, but function-specific.
 - Range of engagement options.

Almost all knowledge portals share a base set of functions that are re-usable.

NATIONAL PLATFORM

- “Shared Meaning”
 - meta-data standards selection
 - crosswalks (inter-standard translations)
 - controlled vocabularies, thesauri, and ontologies associated with standards –
 - these are often domain and region-specific

NATIONAL PLATFORM

- “Shared Processes”
 - coordinated training and awareness creation on the importance of meta-data
 - data sharing and meta-data provision policies / guidelines to be implemented at research institutions and as a condition of funding
 - coordination of and assistance with contributions to national infrastructure
 - creation of forums for the validation of decisions and guidelines

MAIN COMPONENTS OF A GEOPORTAL

Meta-Data
Repository

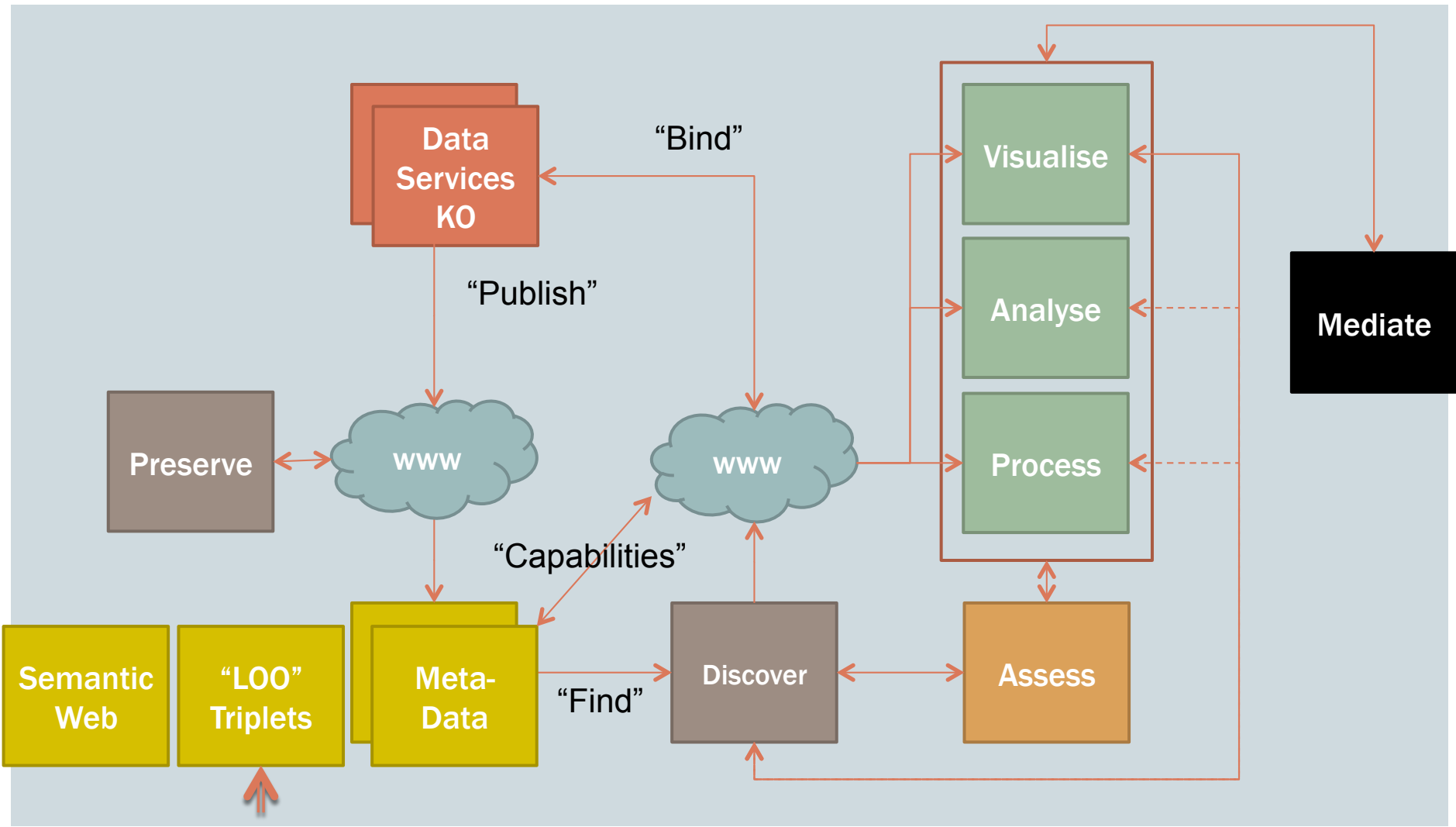
Search
Interfaces and
Services

Meta-Data
Visualisation

Data
Visualisation
Components

Map
Compositions

MAIN COMPONENTS



META-DATA REPOSITORY

- A meta-data repository is required to describe and reference data sets (spatial data layers, mostly, in a GeoPortal).
- Open Source Options:
 - GeoNetwork
 - Suitable for ISO 19115, FGDC and SANS 1878.
 - Requires an Apache/ Tomcat web server.
 - MetaCAT
 - Designed for EML (Ecological Markup Language)
 - Can be altered to accommodate other standards.
 - Requires an Apache/ Tomcat web server.
- SAEON/ CSIR Repository
 - Available for SAEON, CSIR, SAEOS, WDC, and Risk Atlas meta-data.
 - Supports simple meta-data management for a variety of standards.

DATA

- **Links to Data can be any of the following:**
 - A standardised, interoperable 'end point'
 - A link to a downloadable data set
 - A link to a sub-setting and query page
 - A link to a third-party web page
- **Interoperable Data Sets:**
 - **OGC WxS Data Sources**
 - Web Map Service (WMS)
 - Web Feature Service (WFS/ WFS-T)
 - Web Coverage Service (WCS)
 - KML
 - GeoRSS
 - **Sensor Observation Services**
 - **NetCDF Data Sources**
- **Data Services are provided by**
 - ArcGIS and Services
 - GeoServer
 - MapServer
 - Etc,
- **Data Storage can be**
 - File-based shape and raster collections
 - PostGres with Spatial Extensions
 - Oracle Spatial (\$\$\$)
 - SQL Server 2008 R2
- **We can also accommodate data as**
 - NetCDF (used extensively for multidimensional EO data)
 - SensorWeb (time-series type observations, for example from weather stations).

SPECIALISED DATA MODELS

- **‘Observations and Measurements’**
 - Very generalised, not only applicable to Earth and Environmental Observation
 - Advantage of time dimension coupled to spatial dimension
 - Used globally for water-related observation and weather stations
- Can be exposed as **‘SensorWebs’** – collating data from multiple observation networks.

DISCOVERY

- Searches are executed against a Meta-Data Repository.
- The Shared Platform provides
 - Automatable search services based on CS/W
 - A search interface that executes searches against such repositories.
 - JavaScript-based interfaces to render search results.
- GeoNetwork provides standards-compliant CS/W Search facilities out-of-the-box
- We will shortly complete work on
 - Search aggregations and visualisations of search results.
 - Search result analyses.
 - OAI=PMH Interface

VISUALISATION

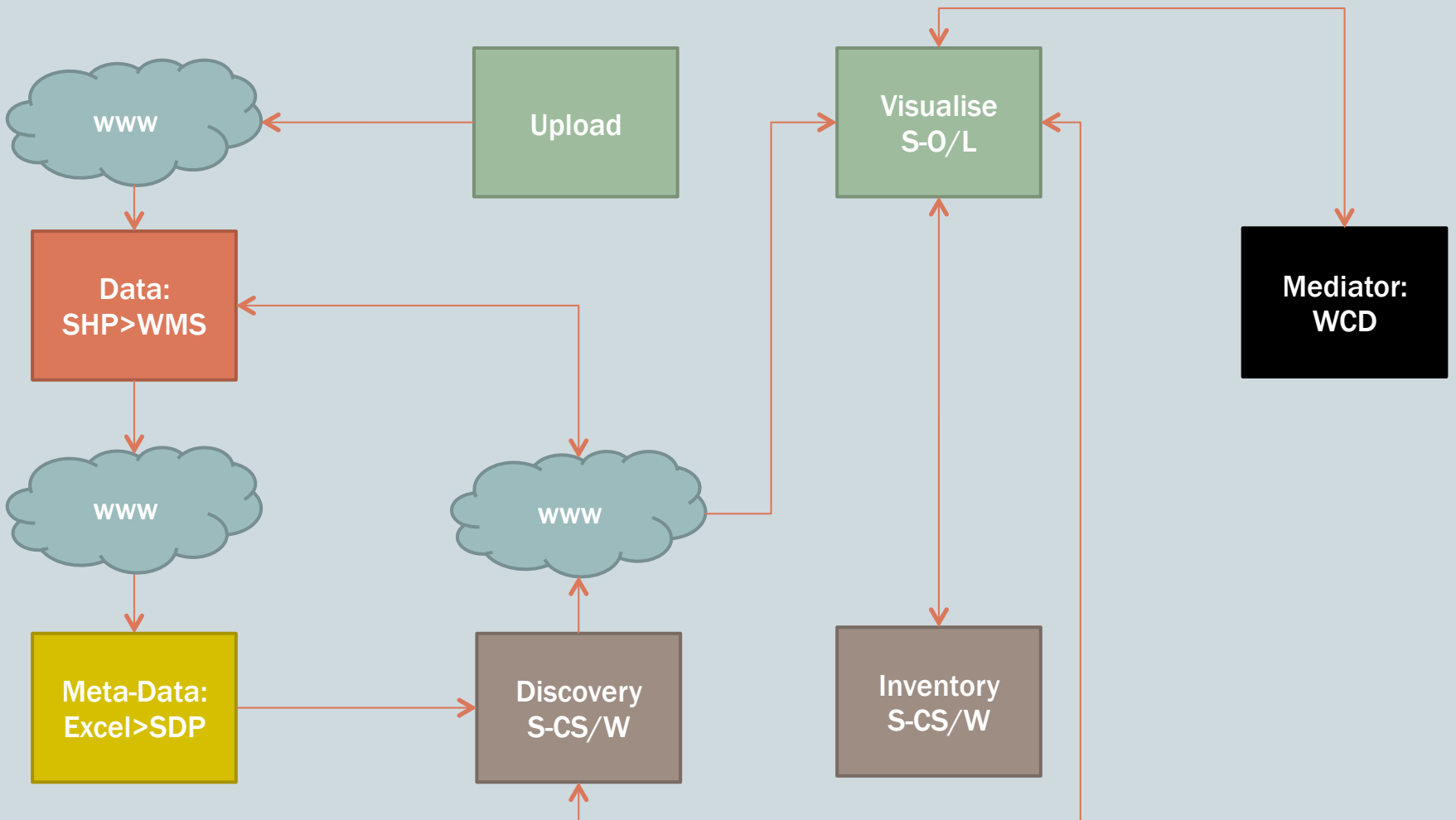
■ Spatial Data

- Most common libraries: Geo-EXT, GeoDjango, and OpenLayers
- SAEON uses Geo-EXT and OpenLayers
- Supports
 - WMS
 - WFS
 - WCS
 - KML
 - GeoRSS
 - GML
- SAEON/ SAEOS is Adding
 - Sensor Observation Services
 - NetCDF
- Integrated with Search Facility

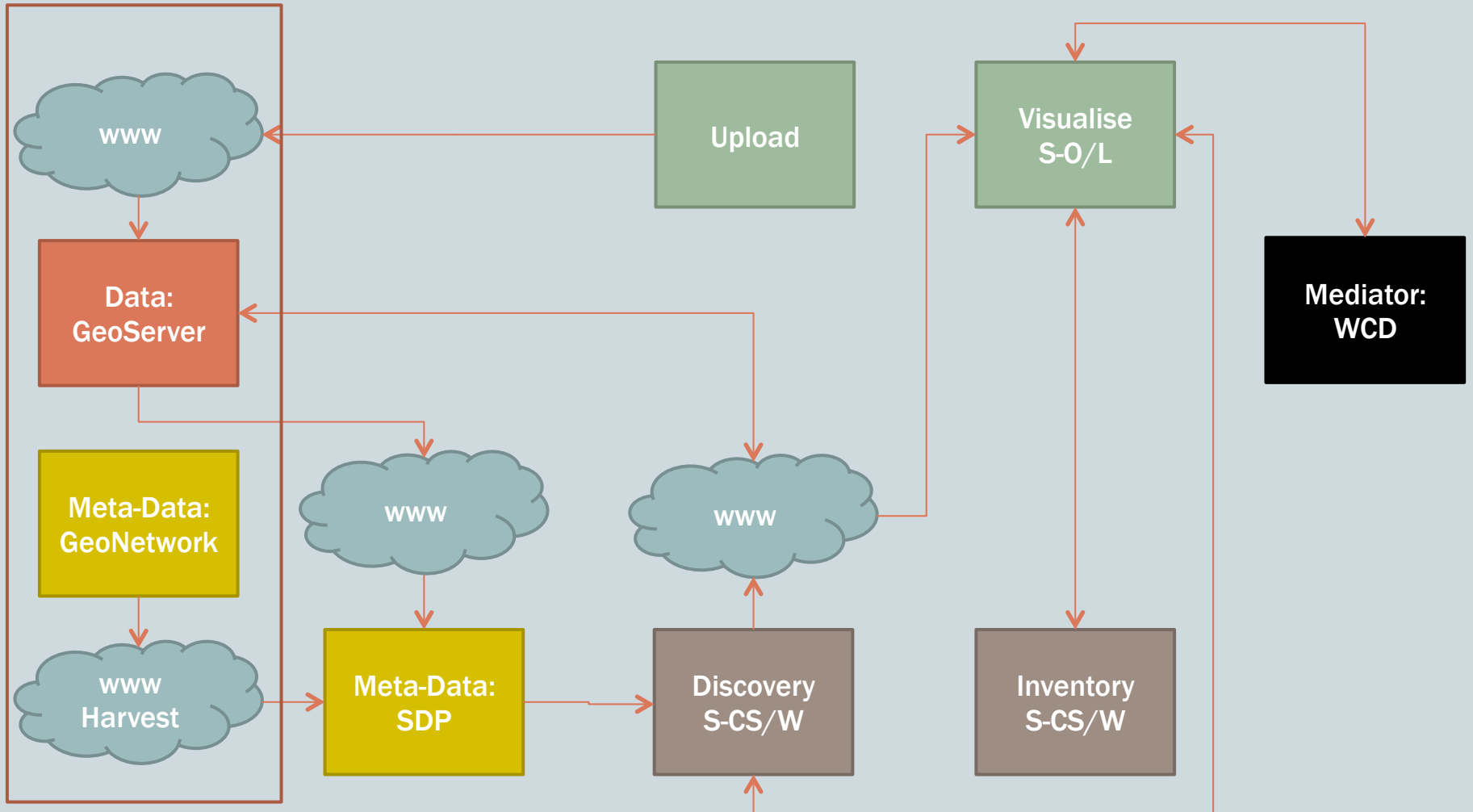
IMPLEMENTATION OPTIONS

'Minimum'
'Medium'
'Clone'

MINIMUM IMPLEMENTATION



MEDIUM IMPLEMENTATION



**FREELY AVAILABLE
COMPONENTS**

PLATFORM COMPONENTS

- Lightweight, JavaScript-based
- Implementations of
 - OpenLayers:
http://196.213.187.51/PLATFORM_2/MAP/index.htm
 - Search Component:
http://196.213.187.51/PLATFORM_2/MAP/search_portal.htm
 - CS/W Component:
http://196.213.187.51/PLATFORM_2/MAP/csw_query.asp

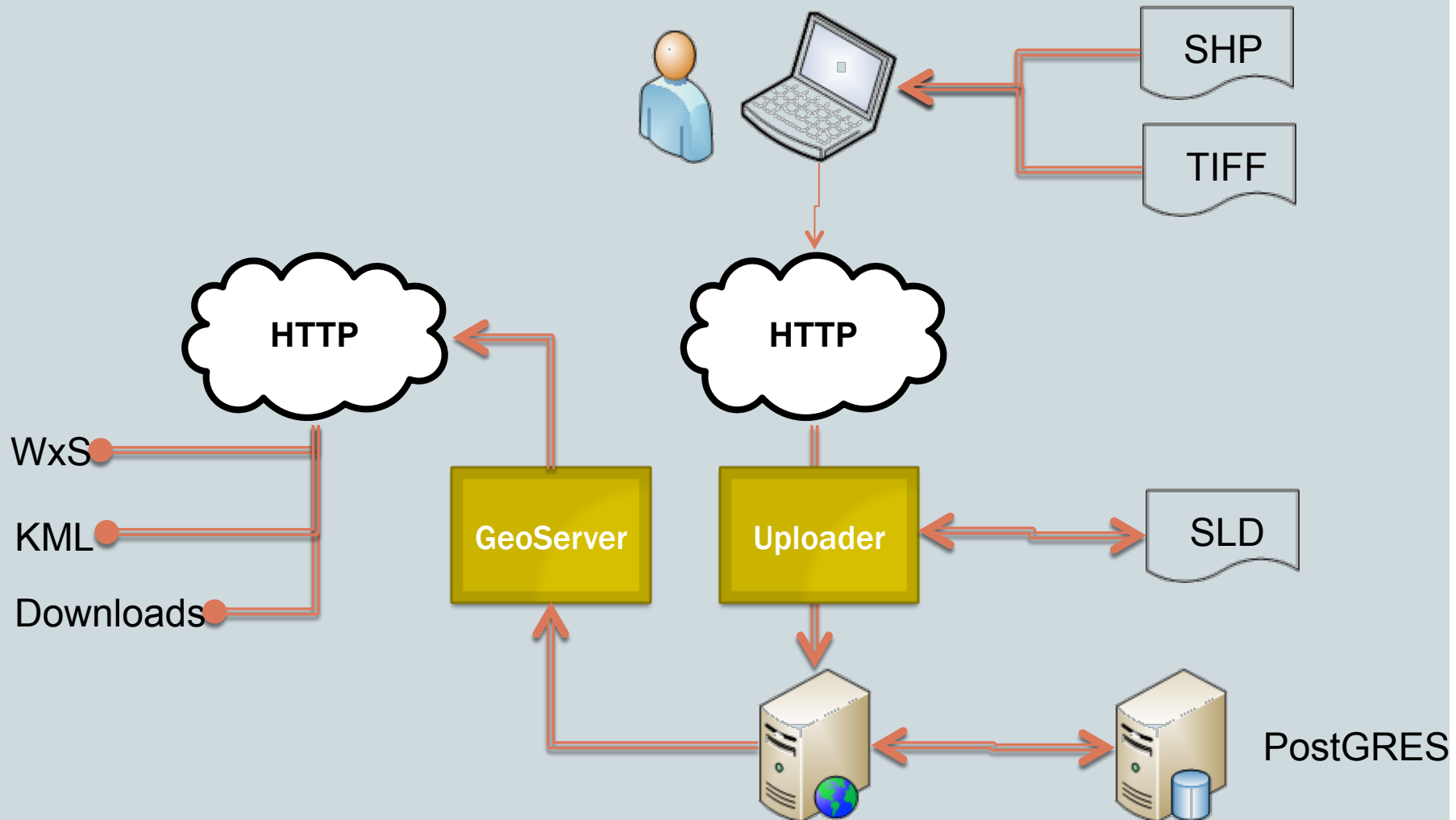
OPEN SOURCE COMPONENTS

- Apache and TomCat (we are using Apache Tomcat/6.0.16)
- Java (we are using Sun Microsystems Inc.: 1.6.0_21 (Java HotSpot(TM) Client VM))
- Postgres (we are using v 8.4) with spatial profiles/ extensions – see <http://postgis.refractions.net/documentation/>
- GeoServer (we are using GeoServer 2.0.1). Downloads and documentation at <http://docs.geoserver.org/>
- GeoNetwork (we are using v 2.3) – download from <http://sourceforge.net/projects/geonetwork/>

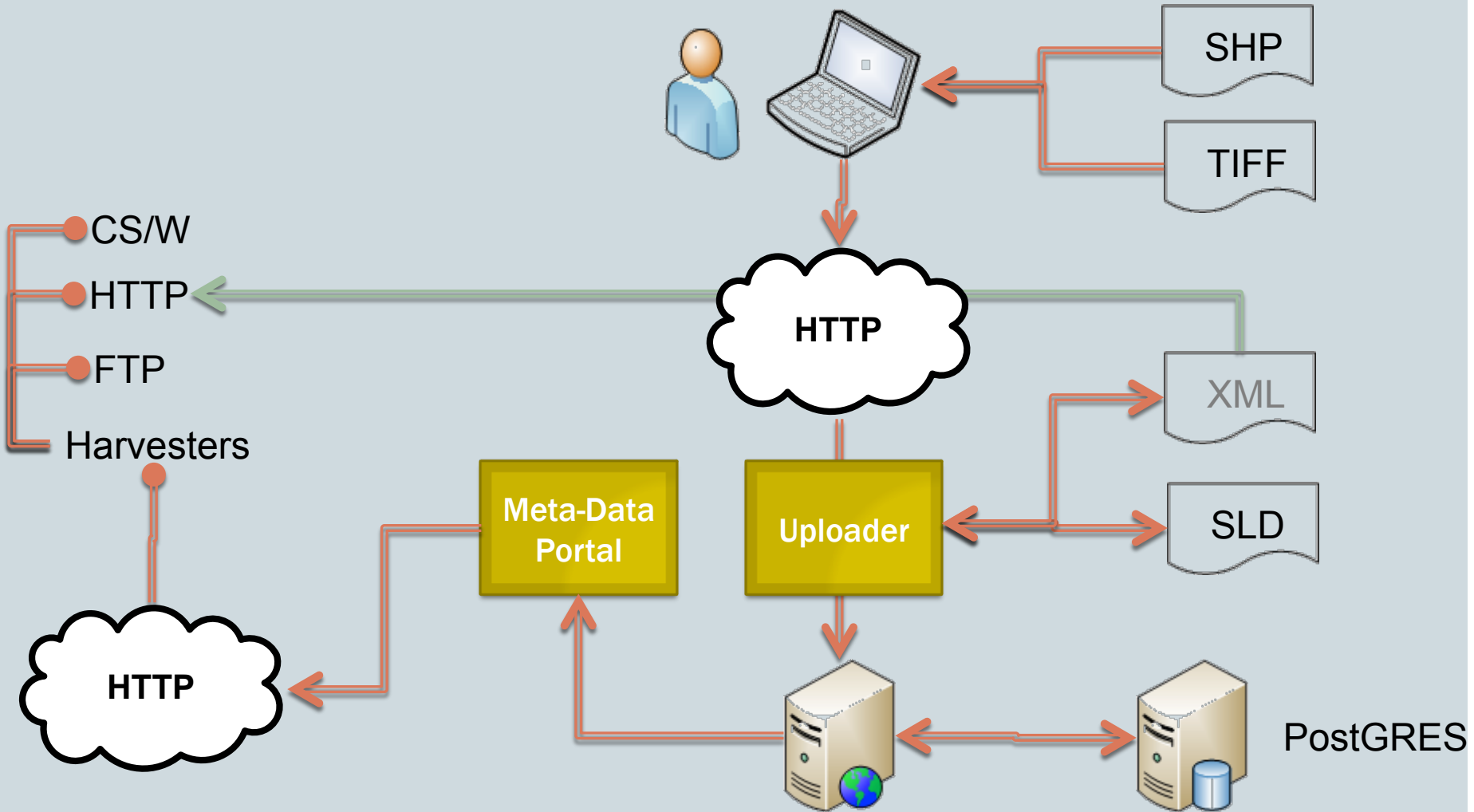
**SIMPLE, ALMOST
IMMEDIATE
IMPLEMENTATION**

Option 1

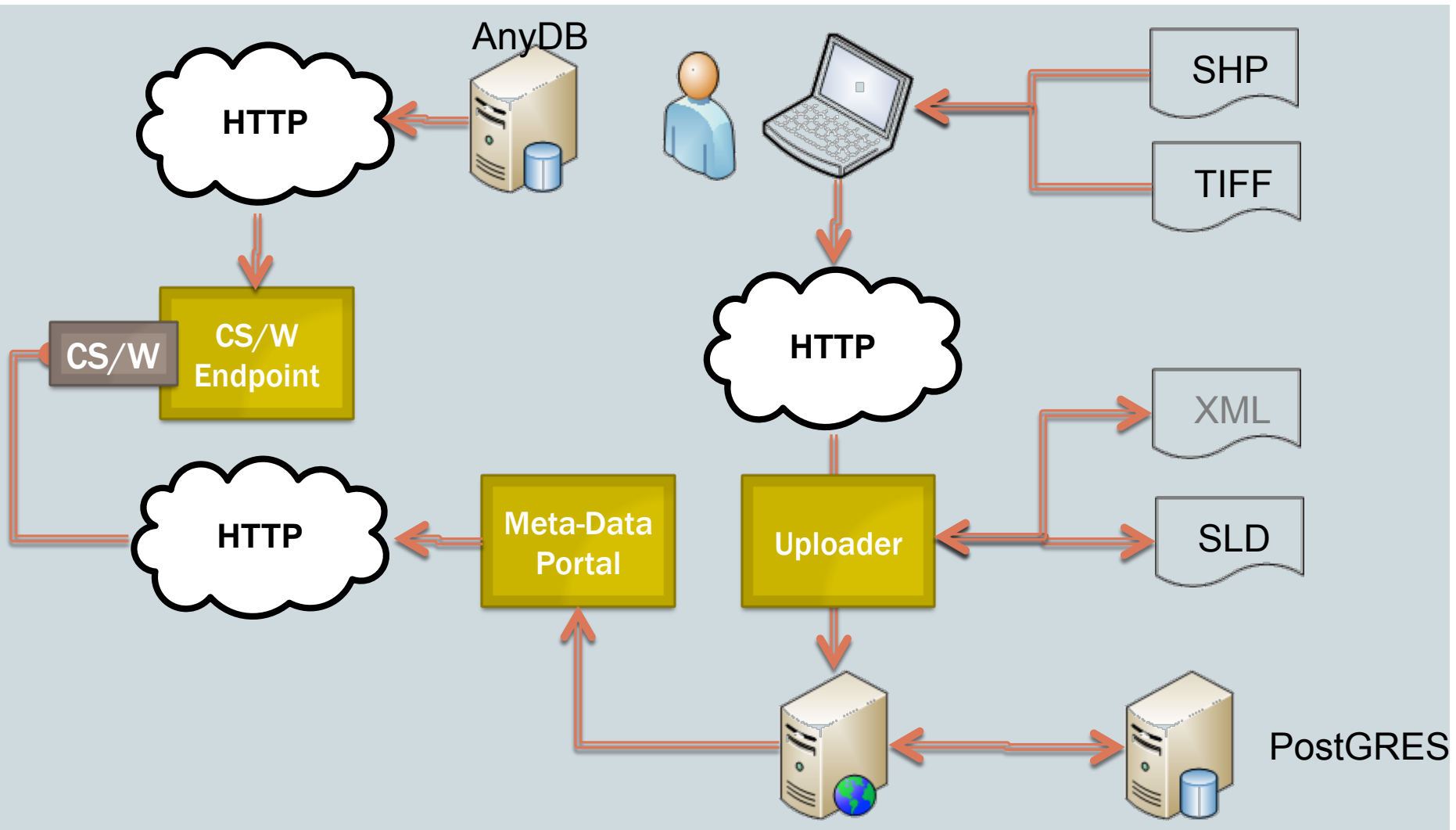
OPTION 1.1: DATA STORAGE



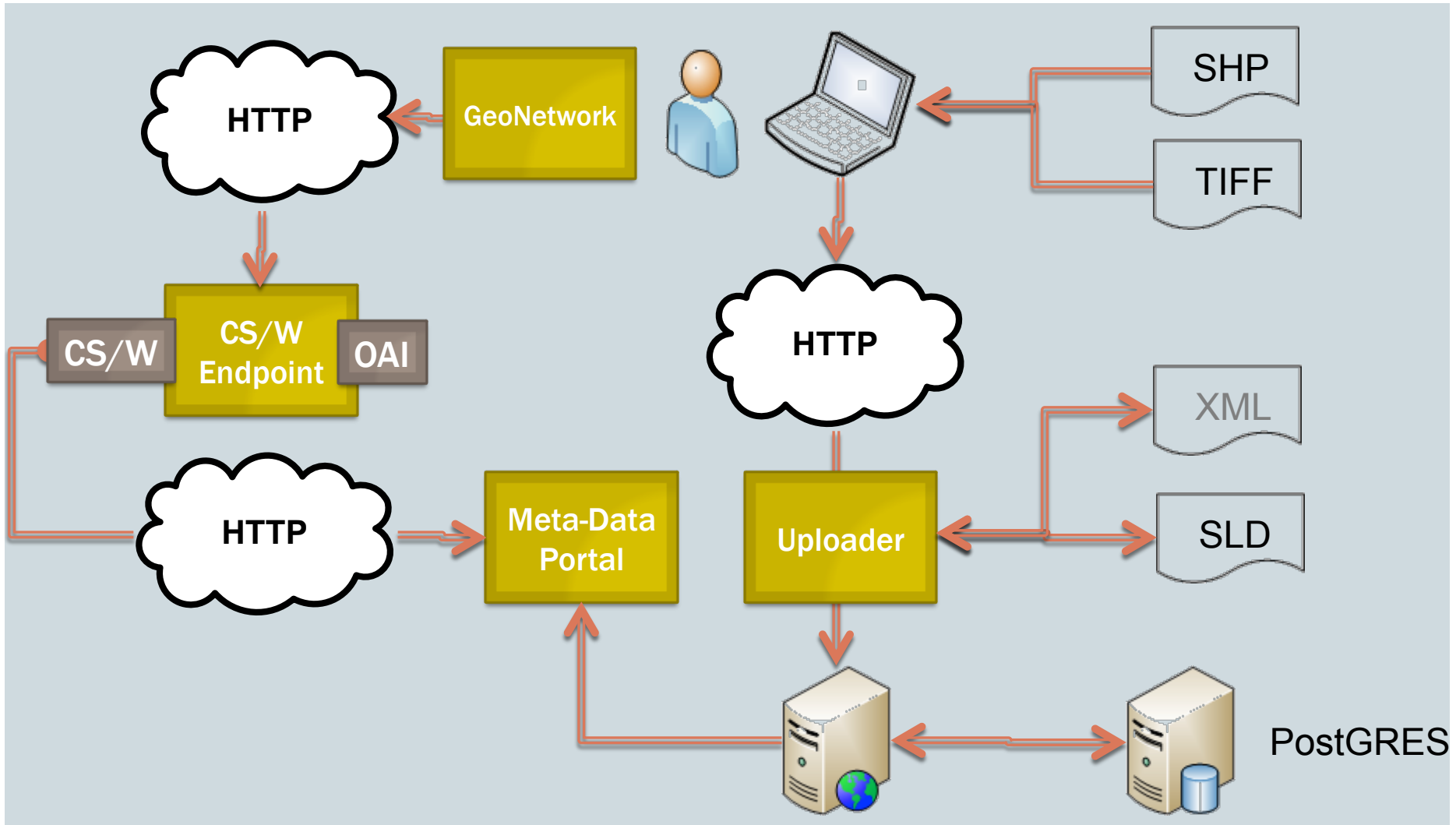
OPTION 1.2.1: META-DATA



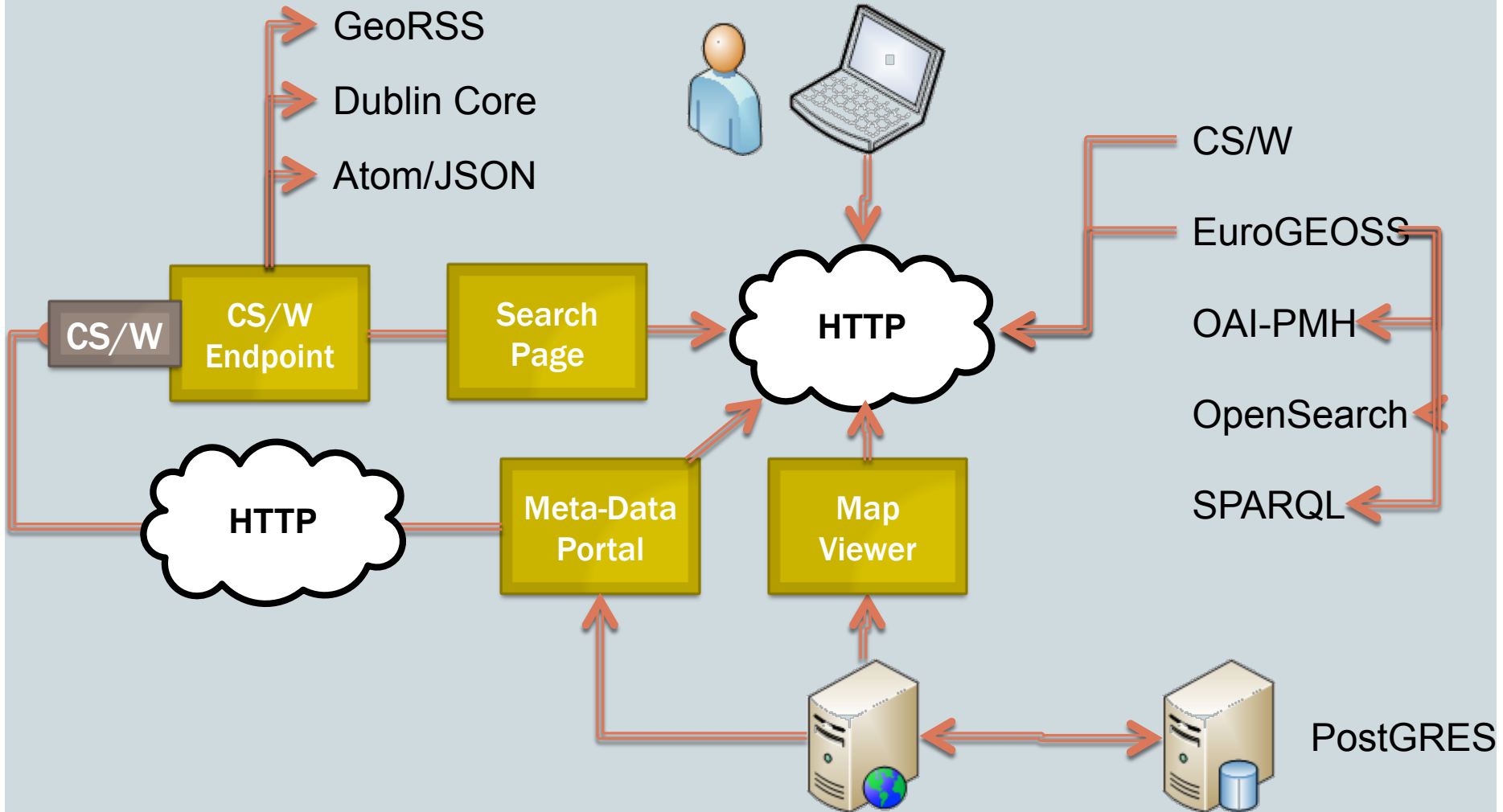
OPTION 1.2.2: META-DATA



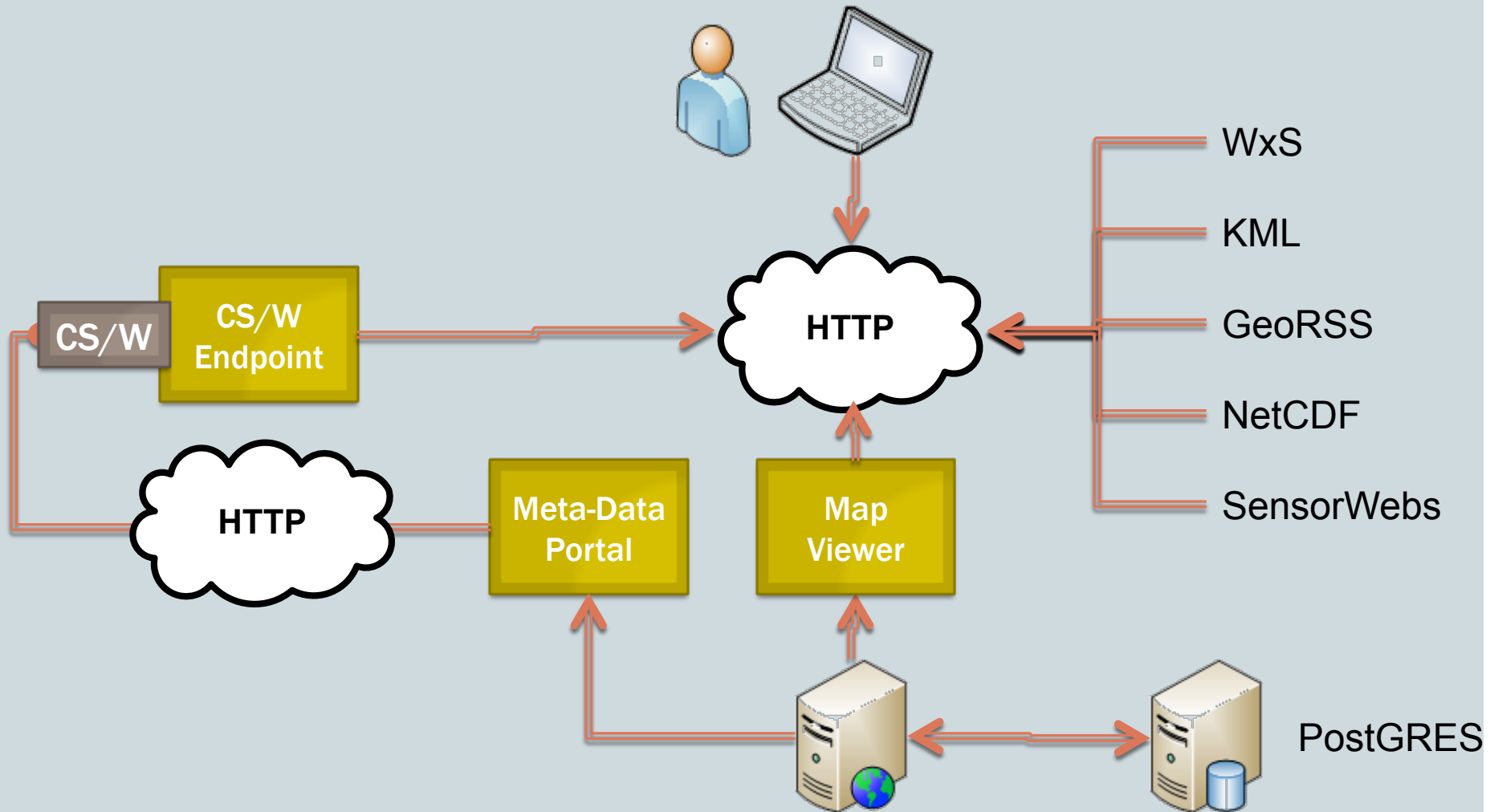
OPTION 1.2.3: META-DATA



OPTION 1.3: DISCOVERY



OPTION 1.4: VISUALISATION



**PORTAL INSTANCE
HOSTED AT CSIR/ CHPC**

Option 2

**PORTAL INSTANCE
HOSTED ON PROVIDER
SITE**

Option 3