Relation between Geospatial information projects related to GBIF

Synthesys 3.6-Synthesys 3.7-GBIF.DE-BioGeomancer

The most up to date work can always be found at:

http://www.biogeografia.com/synthesys

Javier de la Torre - Jens Fitzke - Patricia Mergen

Proposed architecture

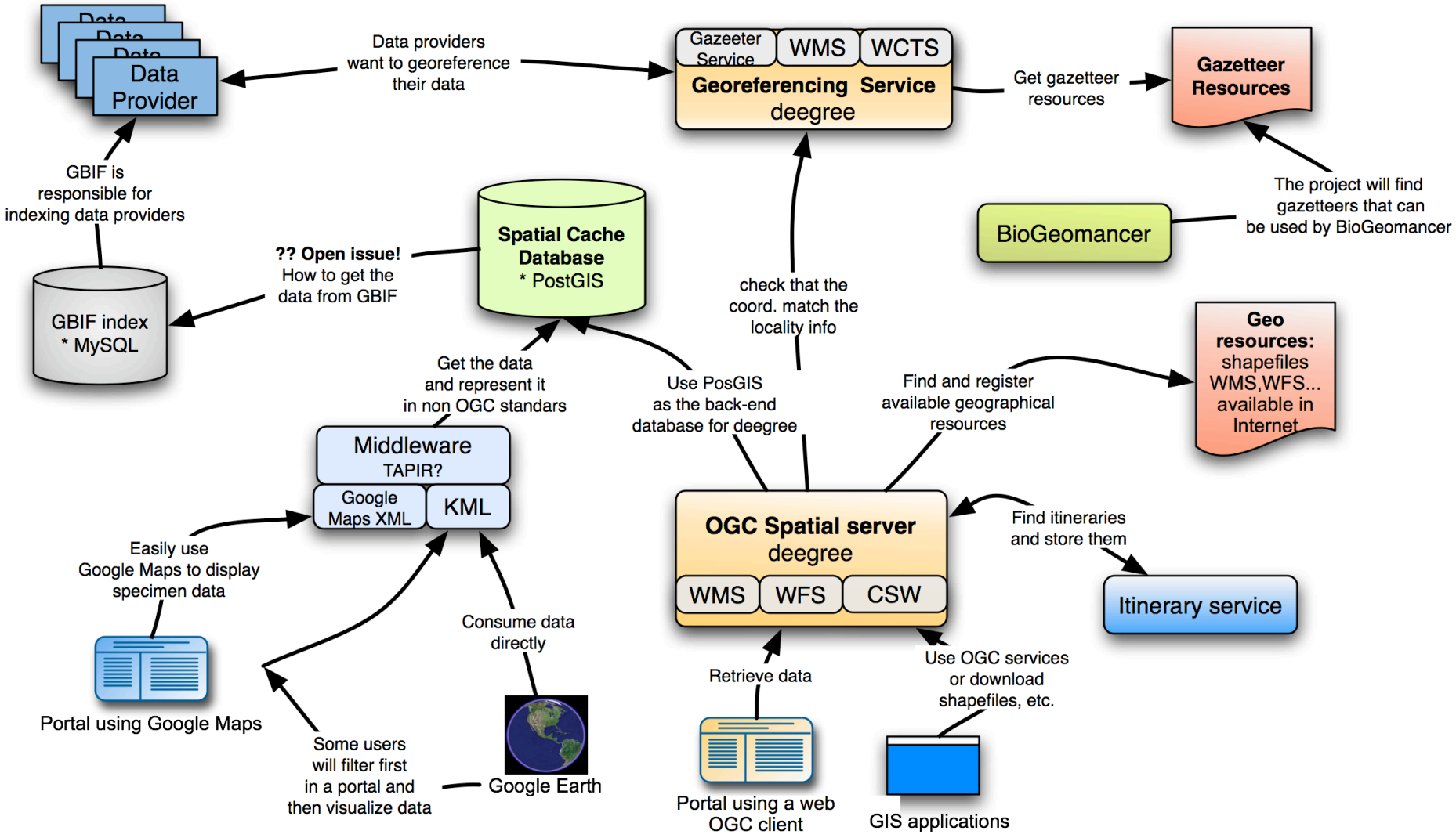


Diagram legend 1

• **Data Providers**: Specimen/observation (unit level) data providers connected to the GBIF network using a BioCASe, Digir or TAPIR wrapper. Digir 2 providers, and pywrapper probably too, will include a little WFS service to provide specimen data in GML.In fact GML for specimen databases (working group in TDWG) could be another view of TAPIR data providers.

•**Spatial cache DB**: Database populated with unit data from data providers. It will have spatial index and query capabilities to be able to do spatial searches on it. Can also include some other geographical information, like itineraries. Could maybe be synchronized with GBIF normal cache in some way.

• **Itinerary finder service**: The Synthesys 3.7 project tries to identify itineraries in already existing collection datasets. It will retrieve this data from the spatial cache DB, and other sources, and will insert into the spatial cache DB the info about the found itineraries. This will also be related to quality mechanism not described in this document and potentially to BioGeomancer project. The way to describe these itineraries could be GML. This GML format is still not defined.

• **OGC spatial server**: This software makes possible to open geographical data trough OGC standards. From the collection of Open Source packages implementing OGC services (Geoserver, UMN, deegree, etc.) deegree was chosen because of the expertise among the partners with it and the more completeness set of available services trough it.

- **WMS**: Web Map Service. Maps with distribution of specimens or any other map that can be created using the data available trough deegree.
- **WFS**: Web Feature Service. Information about specimens in GML format.

Diagram legend 2

• **NON OGC Services**: We think that we have to provide a way to display the data in non standard compliant software like Google services.

- **Google Map Client**: A simple interface using Google Maps taking the data from the spatial cache, and optionally layers available.
- **Easy XML**: Maybe some other simple formats that could be used in other software (we can consider a TAPIR wrapper here).
- **KML**: Keyhole Mark-up Language. To see the data in Google Earth.

• **Portals using an OGC web Client**: A web client able to retrieve data from OGC services. Ideally should also be a way for GIS users to filter data on the database, by refining their search and finally be able to see the data in a map and then be able to download it to their GIS application. There is some available software in deegree that Portals can use to implement this. Examples on how to do it will be provided.

• **Georeferencing service:** Web, or not web? Interface to help data providers to georeference their specimen data. Somebody like the curator goes to a web page, introduce his wrapper URL, or select it from a list of registered, select a dataset of his database to georeference and runs the service. The Georeferencing Server uses various Gazetteer service instances to georeference the records provided by the user.

- **Gazetteer service**: The OGC Gazetteer Service allows a client to retrieve georeferenced well-known place-name vocabulary from a gazetteer database by using the operations of a Web Feature Service.

- **WCTS**: In several point of the architecture here described there will be needs of harmonizing geographical coordinates, for example UTM Spanish grid to decimal coordinates in WGS-84. There is an upcoming standard in OGC for doing this type of transformations.

Diagram legend 3

•**Gazetteer resources**: There are gazetteer resources available at different levels, national, European, worldwide, etc. All partners will look for resources and make them available. A Wiki page will be created where people can inform about discovers.

• **GML developed by the Spatial Data Standard TDWG subgroup**: GBIF specimen information can be described using GML. There is a subgroup in TDWG working on this. This GML schema should be agreed between all partners, Digir 2 developers and any other project related to GIS use of GBIF data.

• **BioGeomancer**: Worldwide collaboration effort to provide a georeferencing framework for natural collections. The scope of this project is much bigger than the one described here in this diagram. Therefore we would like to provide to this project all the resources here available and if possible integrate BioGeomancer services into this infrastructure. Meeting needed to define interfaces.

• **CSW:** The OGC Catalog Service 2.0 specification defines a common interface that enables diverse but conformant applications to perform discovery, browse and query operations against distributed and potentially heterogeneous catalog servers. Most commonly a catalog service is used to provide access to metadata about geospatial data and/or geospatial services by exposing an XML/HTTP interface. The underlying database may implement Dublin Core or ISO 19115/19119 metadata elements. Various implementations of the Catalog Service - Web Profile (CSW) exist, among which the deegree catalog service is currently the only open source package available.

• **Geographic resources:** As explained this is necessary to do analysis of the data. All partners should try to identify resources freely available. If necessary there is some money in Synthesys 3.6 to buy some good ones for Europe.

Diagram legend 4

• **Google Earth:** A pretty impressive software by Google to visualize the world. It has its own geographical laguage called KML, with similar goals as GML, but simpler. XLTs can be considered to transform GML into KML and viceversa…

• **GIS applications**: There are many. Some of them can consume directly data from OGC compliant services. We may consider three applications as example: Arcview (with Plugin, from ESRI), uDig (OpenSource), Diva-GIS (OpenSource), JUMP (Open Source).

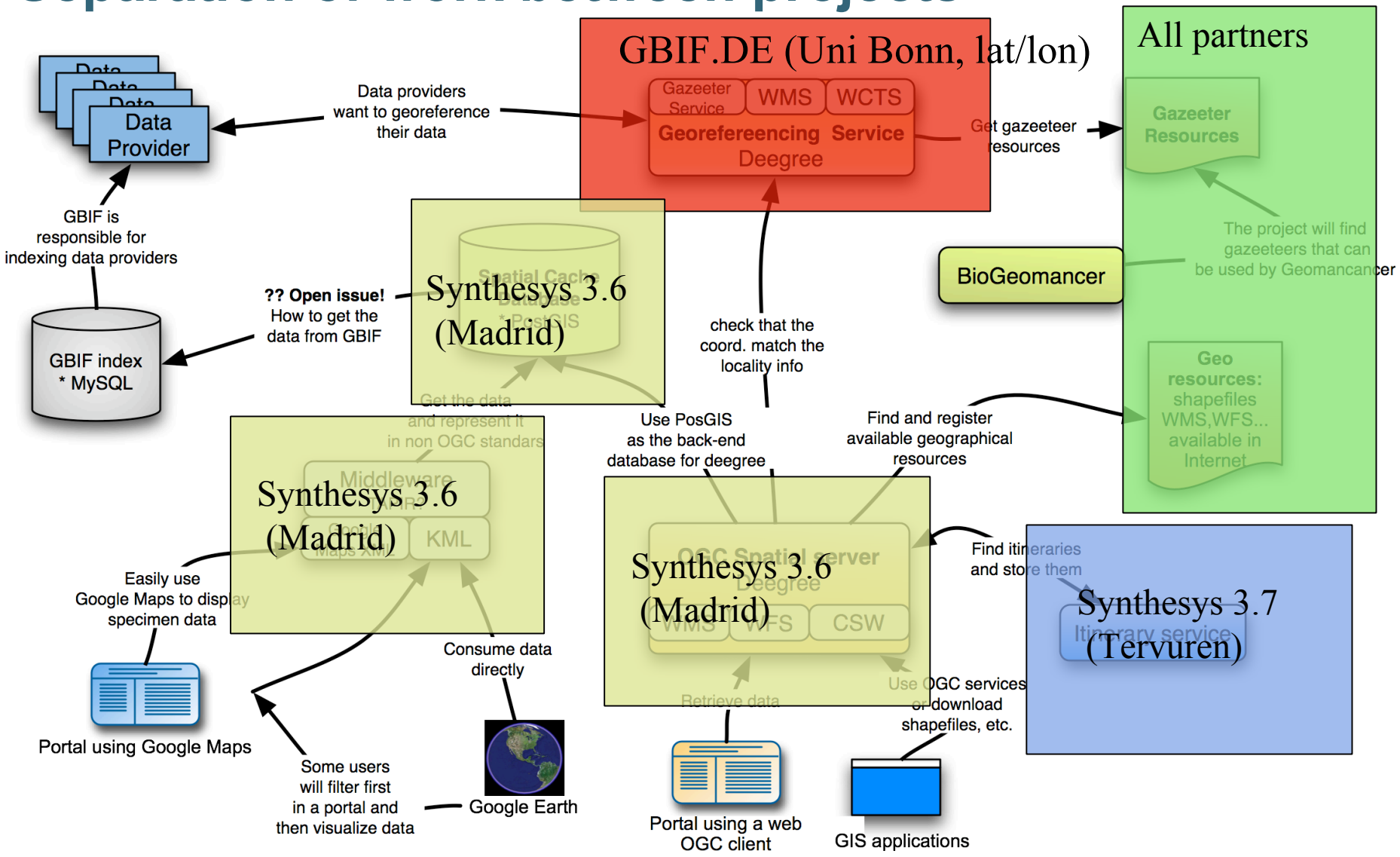
• **Portals**: Biodiversity informatics portals like GBIF.NET, BioCASE, GBIF.DE, GBIF Spain, etc. They present the unit level data in different ways to different consumers. They may integrate some of the services here described or they will link to them.

Open issues

• **Relation between the spatial cache database(s) and the GBIF index:** Some kind of synchronization should be considered to avoid data incongruence between them. Also the GBIF index have more powerful capabilities that will make the data more harmonize. The best possibility could be that the spatial database only retrieves data from the GBIF index, not directly from the data providers.
The GBIF index database is implemented in MySQL so some software that do incremental updates of the spatial cache database will have to be considered.
For project Synthesys 3.6 this is not a high priority and the work will be done in an example dataset of the GBIF index. This relation between the index an the spatial cache database is a wider issue that has to be discussed inside the GBIF community.

• **GML application schema for specimen data**: In several services data will be transferred using GML. Therefore we need a GML application schema. There is a subgroup in TDWG called Spatial Data Standards that is working on this.

Separation of work between projects



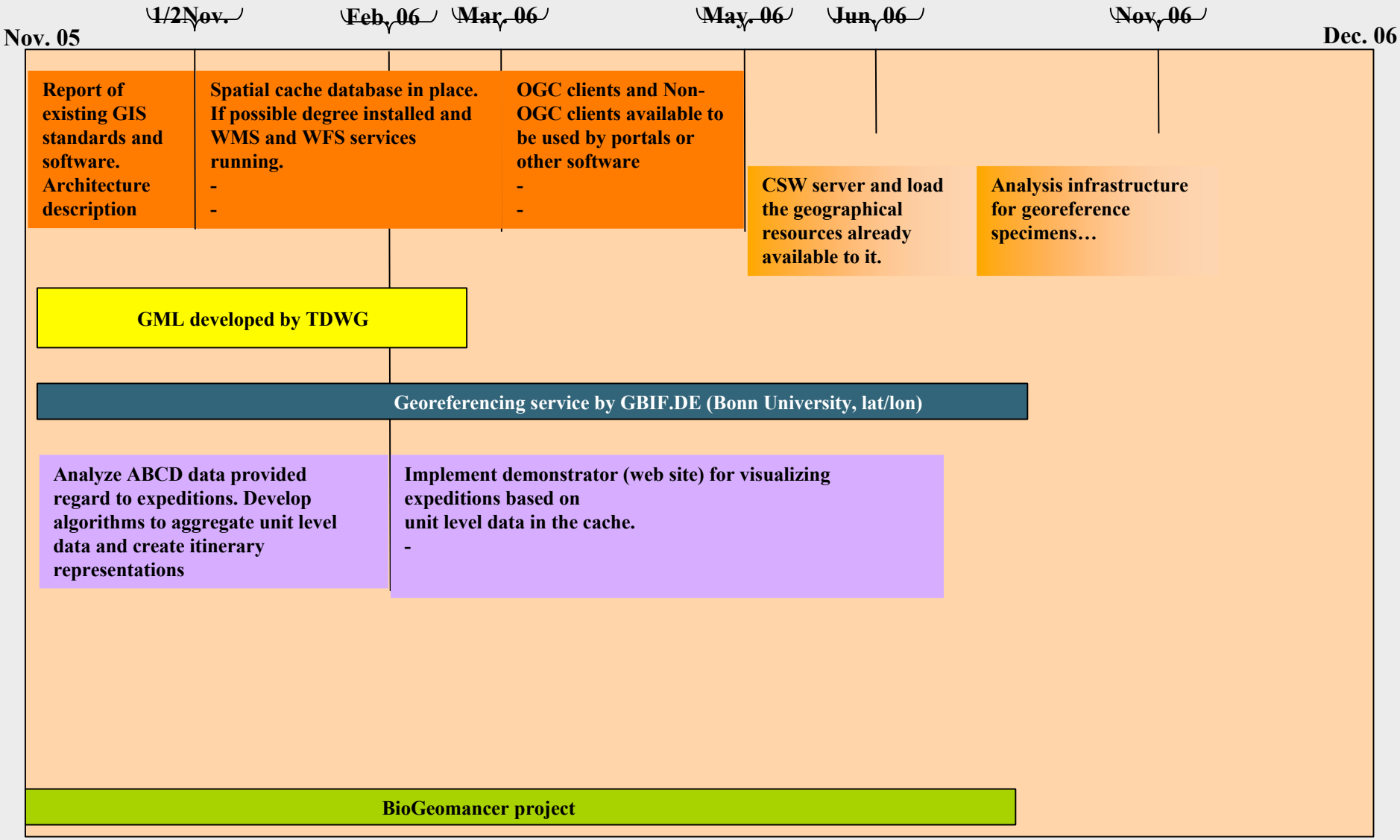
Time planning…

Synthesys 3.6 (Madrid)

Synthesys 3. 7 (Tervuren)

GBIF.DE (Uni Bonn, lat/lon)

BioGeomancer



Links

With no special order…

- **Deegree:** http://deegree.sourceforge.net/
- **Geoserver:** http://geoserver.sourceforge.net/
- **Open Geospatial Consortium:** http://www.opengeospatial.org/
- **Google Maps:** http://maps.google.com
- **Google Earth:** http://earth.google.com/
- **PostGIS:** http://postgis.refractions.net/
- **TAPIR:** http://ww3.bgbm.org/protocolwiki/
- **BioGeomancer:** http://www.biogeomancer.org
- **Synthesys:** http://www.synthesys.info/
- **Lat/Lon:** http://www.latlon.de/
- **GBIF.DE:** http://www.gbif.de/
- **GBIF:** http://www.gbif.org
- **GBIF Spain:** http://www.gbif.es
- **ArcView:** http://www.esri.com/software/arcview/
- **uDig:** http://udig.refractions.net/confluence/display/UDIG/Home
- **Diva-GIS:** http://www.diva-gis.org/
- **BioCASE:** http://ww.biocase.org
- **TDWG:** http://www.tdwg.org/