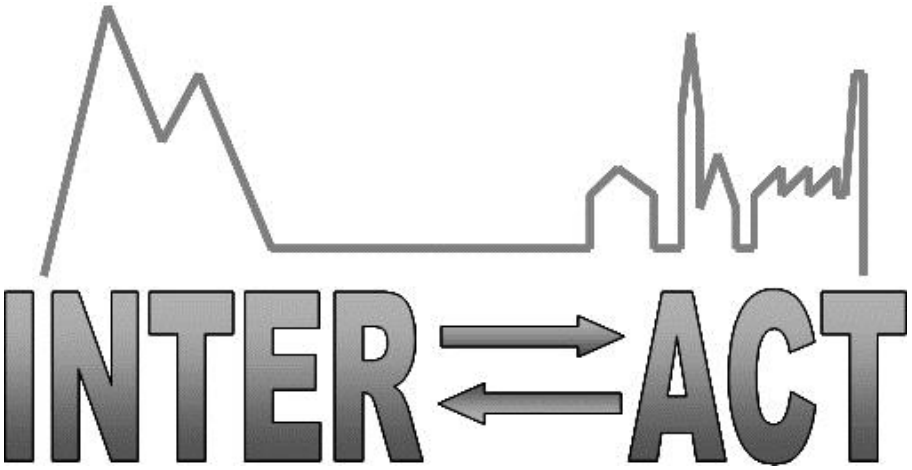
Combination of CP & CSA



D7.8 – Standard protocols document

Project No.262693– INTERACT

FP7-INFRASTRUCTURES-2010-1

Start date of project: 2011/01/01 Duration: 60 months
Due date of deliverable: 31/12/2014 (M48) Actual Submission date: 31/12/2014

Lead partner for deliverable: SLU
Author: Tomas Thierfelder, Mohsen Bashang, Sofia Bryntse

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the Consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

Table of Contents

Publishable Executive Summary

NordGIS/INTERACT GIS solely utilises OPEN source technologies. It has been programmed by using Open Layers JavaScript Libraries and Extensible Markup Language (XML), utilising PostgreSQL with PostGIS for querying the relational system Geo-database. UMN MapServer Engine is being used for rendering database map content, with the database hosted at an Apache/Linux web-server. The database content of station activities metadata are described in accordance with ISO19115 standards, indexed in accordance with the GCMD Science Syllabus, catalogued in accordance with the Open Geospatial Consortium's CS-W standard, and may be exchanged in accordance with the ISO19139 XML implementation scheme. The NordGIS/INTERACT GIS database content may be harvested in accordance with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). NordGIS/INTERACT GIS metadata standards are constantly being developed towards optimal INSPIRE compatibility.

1. Introduction

As part of the INTERACT infrastructure project (FP7-INFRASTRUCTURES-2010-1, project no.262693, www.eu-interact.org), a web-based platform for sharing data and services across field stations has been developed. This platform is, in its generic form, called NordGIS (www.nordgis.org) and hosts an information system tailored to meet the specifications of INTERACT field stations (the INTERACT GIS). These specifications are partly outlined in earlier deliverable reports of the INTERACT WP7 on Data Management, but also in the inventory of station resources performed and reported by INTERACT WP2, the INTERACT Station Manager's Forum.

The NordGIS/INTERACT GIS platform has been prototyped and developed in close collaboration with the Swedish INTERACT-stations at Abisko and Tarfala, where the single-station prototype developed at Abisko was turned into a networked multi-station system with the addition of the Tarfala station. As a result, NordGIS/INTERACT GIS can be used to communicate metadata regarding station activities across any number of connected stations, as well as to other information systems and external metadata repositories. With such communication requiring bilateral communicative standards and protocols, this document briefly describes the NordGIS standards and protocols chosen for system implementation. In addition, our bilateral partner organisations are listed, and also brief technical specifications regarding the NordGIS/INTERACT GIS software and hardware platforms.

2. NordGIS/INTERACT GIS OPEN principles

The NordGIS/INTERACT GIS information system has been developed and implemented in accordance with OPEN principles, meaning that it solely utilises freely available, community-driven software- and server platforms, as well as OPEN metadata and communication standards and protocols. What has been freely adopted should be freely returned to the OPEN society, which in the case of NordGIS/INTERACT GIS means that the information system is kept at a non-commercial basis, available for any station that wishes to connect. Connection is due to self-pricing, including necessary supportive fees. NordGIS/INTERACT GIS is an OPEN fork, meaning that the system code won't be openly returned to the general public, although some licencing procedure probably will be released with future versions.

3. NordGIS/INTERACT GIS bilateral partners

In the academic development environment, knowledge regarding international standardisation is typically weak – academia tend to focus on exploring new frontiers, rather than adapting to existing standards. For the NordGIS/INTERACT GIS development team, it has therefore been beneficial to join with external organisations that specialise in international standardisation. If NordGIS/INTERACT GIS is being developed for compatibility with systems built by such organisations, it will automatically reach beyond the partner organisations, into the international realm. The NordGIS/INTERACT GIS development team have therefore joined with a number of partner organisations for bilateral review, where most of our partners have developed infrastructure due to international standardisation and legislation. The most important of these are the ECDS (Environmental Climate Data Sweden - https://ecds.se/) and Arctic SDI (http://arctic-sdi.org/) projects, where the NordGIS/INTERACT GIS development team have a long record as reference partners. Since ECDS and Arctic SDI may be considered as expressions of European and

international infrastructure standardisation, and by targeting their standards with NordGIS/INTERACT GIS, we benefit from their expertise while they may benefit from freely harvesting data in NordGIS/INTERACT GIS.

4. Metadata standards and protocols

4.1. Metadata profiling and indexing

4.1.1. ISO 19115

In order to communicate the metadata content of the NordGIS/INTERACT GIS database internally across connected stations, as well as to external information systems and repositories, bilaterally shared standards are required. With ISO 19115 currently constituting the most common international standard for defining the scheme required for describing geographic information and services, this is the fundamental metadata standard chosen for implementation in NordGIS/INTERACT GIS. It provides information about the identification, the extent, the quality, the spatial and temporal scheme, spatial reference, and distribution of digital geographic data.

ISO 19115 is applicable to:

- The cataloguing of datasets, clearinghouse activities, and the full description of datasets.
- Geographic datasets, dataset series, and individual geographic features and feature properties.

ISO 19115 defines:

- Mandatory and conditional metadata sections, metadata entities, and metadata elements.
- The minimum set of metadata required to serve the full range of metadata applications (data discovery, determining data fitness for use, data access, data transfer, and use of digital data).
- Optional metadata elements - to allow for a more extensive standard description of geographic data, if required.
- A method for extending metadata to fit specialised needs.

Though ISO 19115 is applicable to digital data, its principles can be extended to many other forms of geographic data such as maps, charts, and textual documents as well as non-geographic data.

4.1.2. GCMD science syllabus

The NordGIS/INTERACT GIS metadata elements are labelled with, and searchable via, keywords that constitute a subset of NASA's Global Change Master Directory (GCMD - http://gcmd.nasa.gov/learn/keywords.html), the so called GCMD Science Syllabus. Controlled keyword vocabularies like the GCMD Science Syllabus ensure that metadata are described in a consistent manner and allow for the precise searching of metadata records and subsequent retrieval of data and services. There are seven sets of controlled keywords in the directory: earth sciences, data services, data centres, geographic locations, instruments/sensors, platforms/sources, and projects. In addition, the Climate Diagnostics descriptions include two unique keyword sets: visualization type and analysis type. Most of these apply to NordGIS/INTERACT GIS, although at a rather shallow level of metadata specification.

4.2. Metadata registry service

4.2.1. OGC CSW

A service-oriented architecture like NordGIS/INTERACT GIS must support fundamental interactions like publishing resource descriptions so that they are accessible to prospective users (*publish*), discovering resources of interest according to some set of search criteria (*discover*), and then interacting with the resource provider to access the desired resources (*bind*). Within such an architecture a *registry service* plays the essential role of matchmaker by providing publication and search functionality, thereby enabling a *requester* to dynamically discover and communicate with a suitable resource *provider* without requiring the requester to have advance knowledge about the provider (*Figure 1*).

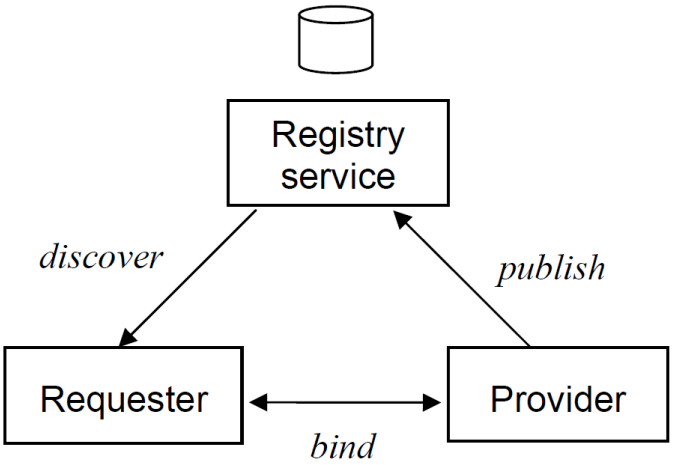


Figure 1 – Essential interactions in a service-oriented architecture

NordGIS/INTERACT GIS utilises the Open Geospatial Consortium's (OGC) CSW-ebRIM Registry Service for enabling the above interactions, which is a "profile" based on the HTTP protocol binding (the CSW part) documented in Clause 10 of the OGC Catalogue Services Specification (version 2.0.2, OGC 07-006r1). The profile imposes some constraints on the use of the base specifications and introduces additional search, retrieval, and registry management capabilities. It provides facilities for advertising and discovering a wide variety of information resources, in particular when metadata are profiled and indexed as described above.

4.3. Metadata exchange

4.3.1. ISO 19139 XML

The importance of metadata describing digital geographic data is explained in detail in the text of the ISO 19115 referred to above. ISO 19115 provides a structure for describing digital geographic data by defining metadata elements and establishing a common set of metadata terminology, definitions and extension procedures in NordGIS/INTERACT GIS. It is abstract in that it provides a worldwide view of metadata relative to geographic information, but no encoding. Therefore, the actual implementation of geographic information metadata could vary based on the interpretation of metadata producers. In an attempt to facilitate a standardisation of implementations, the comprehensive ISO 19139 metadata implementation specification provides a definitive, rule-based encoding for implementing ISO 19115 in NordGIS/INTERACT GIS. The ISO 19139 technical specification provides Extensible Markup Language (XML) schemas that are meant to enhance

interoperability by providing a common specification for describing, validating and exchanging metadata regarding geographic datasets, dataset series, individual geographic features, feature attributes, feature types, feature properties, etc.

ISO 19115 defines general-purpose metadata in the field of geographic information. More detailed metadata for geographic data types and geographic services are defined in other ISO 19100 series standards and user extensions (ISO 19115). The ISO 19139 technical specification is also used to define implementation guidelines for general-purpose metadata. Where necessary, interpretations of some other ISO 19100 series standards are incorporated in ISO 19139.

ISO 19118 describes the requirements for creating encoding rules based on UML schemes and the XML-based encoding rules as well as providing an introduction to XML. The ISO 19139 technical specification utilizes the encoding rules defined in ISO 19118 and provides the specific details of their application with regard to deriving XML schemas for the UML models in ISO 19115.

4.3.2. OAI-PMH

The NordGIS/INTERACT GIS metadata catalogue is available for external harvesters via the Open Archives Initiative protocol for Metadata Harvesting (OAI-PMH). The essence of the open archives approach is to enable access to Web-accessible material through interoperable repositories for metadata sharing, publishing and archiving. It defines a mechanism for harvesting records containing metadata from repositories. The OAI-PMH gives a simple technical option for potential data providers like NordGIS/INTERACT GIS to make their metadata available to services, based on the open standards HTTP (Hypertext Transport Protocol) and XML (Extensible Markup Language). The metadata that is harvested may be in any format that is agreed by a community (or by any discrete set of data and service providers), although unqualified Dublin Core is specified to provide a basic level of interoperability. Thus, metadata from many sources can be gathered together in one database, and services can be provided based on this centrally harvested, or "aggregated" data. The link between this metadata and the related content is not defined by the OAI protocol. It is important to realise that OAI-PMH does not provide a search across this data, it simply makes it possible to bring the data together in one place. In this perspective, NordGIS/INTERACT GIS constitutes a potential provider, with its own search engines as declared in this document. These search engines are, hence, not exported with eventual external data harvest.

4.3.3. INSPIRE

With OGC having a long record of international standardisation, the European Community has entered the process relatively late, although quite forceful. A major European development has been the entering in force of the INSPIRE Directive in May 2007, establishing an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment. INSPIRE is based on the infrastructures for spatial information established and operated by the 28 Member States of the European Union, although much has been copied from OGC's international work of standardisation. The Directive addresses 34 spatial data themes needed for environmental applications, with key components specified through technical implementing rules. INSPIRE is based on a number of common principles:

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.

- Geographic information needed for good governance at all levels should be readily and transparently available.
- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used

The INSPIRE directive came into force on 15 May 2007 and will be implemented in various stages, with full implementation required by 2019. It regulates in detail how metadata should be indexed for internal cataloguing as well as for internal and external communication. Adopting an infrastructure architecture like NordGIS/INTERACT GIS to the INSPIRE standard is an industrious process since it includes all connected users, but should be finished by 2019.

5. Software platform

5.1. OpenLayers JavaScript Libraries

The source code of the NordGIS/INTERACT GIS architecture is utilising OpenLayers JavaScript Libraries to facilitate dynamic user interaction with web-based on-screen maps. OpenLayers is a pure JavaScript library for displaying map data in most modern web browsers, with no server-side dependencies. It implements a JavaScript application programming interface (API) for building web-based geographic applications, similar to the Google Maps and MSN Virtual Earth APIs, with one important difference; OpenLayers is Free Software, developed for and by the Open Source software community. One of the guiding principles of OpenLayers development has been to maintain a set of small examples of most functionalities, allowing the library to demonstrate most of what it can do by example, see http://docs.openlayers.org.

5.2. Extensible Markup Language

The source code of NordGIS/INTERACT GIS utilises Extensible Markup Language (XML) to define internal as well as external communication of the metadata that constitute the NordGIS/INTERACT GIS database. As described above, we consider XML to be an ISO 19139 set of rules for system implementation of the ISO 19115 metadata standard. XML is a markup language that defines a set of rules for encoding documents in a format which is both human- and machine-readable. It is defined by the W3C's XML 1.0 Specification, as well as by several other related specifications, all of which are free open standards. The design goals of XML emphasize simplicity, generality and usability across the Internet. It is a textual data format with strong support via Unicode for different human languages. Although the design of XML focuses on documents, it is widely used for the representation of arbitrary data structures such as the web services provided by NordGIS/INTERACT GIS.

5.3. PostgreSQL/PostGIS

The source code of NordGIS/INTERACT GIS utilises PostgreSQL with PostGIS to define user interactions with the NordGIS/INTERACT GIS database content. Such interactions range from PI entering of new data, via administrative editing of data, to public query. PostgreSQL is an object-relational database management system (ORDBMS) with an emphasis on extensibility and compliance with standards. As a database server, its primary function is to store data securely, supporting best practices, and to allow for retrieval at the request of other software applications.

PostGIS is an open source software program that adds support for geographic objects to the PostgreSQL object-relational database. PostGIS follows the Simple Features for SQL specification from the Open Geospatial Consortium.

6. Hardware platform

6.1. UMN MapServer Engine

NordGIS/INTERACT GIS utilises the UMN MapServer Engine platform to dynamically deliver the map content called for by users as they query the system database. Since this deliverance is performed internally within NordGIS/INTERACT GIS, rather than exported to some external client, the map content is “received” by internal OpenLayers/PostGIS functionalities for user interactions as described above (WMS/WFS would have been required for deliverance to external clients). UMN MapServer is an open source development environment for building server-hosted spatially enabled internet applications. It can run as a Common Gateway Interface (CGI) program or via MapScript which supports several programming languages. MapServer was originally developed by the University of Minnesota (UMN) ForNet project in cooperation with NASA, and the Minnesota Department of Natural Resources (MNDNR). Later it was hosted by the TerraSIP project, a NASA sponsored project between the UMN and a consortium of land management interests.

6.2. Apache/Linux

NordGIS/INTERACT GIS is physically situated at a webserver operated with Linux OS, which uses Apache software to communicate with the web. The Apache HTTP Server Project is an effort to develop and maintain an open-source HTTP server for modern operating systems including UNIX and Windows. The goal of the project is to provide a secure, efficient and extensible server that provides HTTP services in sync with the current HTTP standards.

Linux is a Unix-like and mostly POSIX-compliant computer operating system (OS) assembled under the model of free and open-source software development and distribution. The defining component of Linux is the Linux kernel, an operating system kernel first released on 5 October 1991 by Linus Torvalds.