

Contribuciones de una IDE a la e-Ciencia: Proyecto AWARE

C. Granell[†], L. Díaz[†], M.A. Esbrí[†], M. Gould[†],
A. Lladós[‡].

Centro de Visualización Interactiva (CEVI)

[†] Departamento de Lenguajes y Sistemas Informáticos.

Universitat Jaume I

Avenida Sos Baynat s/n, 12071 Castellón

e-mail: gould@lsi.uji.es

[‡] Instituto Cartográfico de Cataluña

Parc Montjuic s/n Barcelona

e-mail: allados@icc.es

Resumen

En este capítulo analizaremos el papel que potencialmente juegan las IDEs en el ámbito de la e-Ciencia. También describimos algunas de las mejoras que aportan los componentes de una IDE a la e-Ciencia, por ejemplo, la reciente especificación de OGC para geo-procesamiento en la Web. Finalmente describimos la implementación de un WPS en un proyecto europeo en ejecución, en un contexto medioambiental marcado por GMES.

Palabras clave: e-Ciencia, Servicios Web, WPS.

1 Introducción

Vivimos en una era sin precedentes en cuanto a volumen y diversidad de datos aplicados a procesos de análisis y síntesis multidisciplinares. Los problemas medioambientales complejos como el cambio climático y el ciclo del agua trascienden barreras geográficas y disciplinarias, y requieren soluciones científicas aplicadas a sistemas integrados de información geográfica [1].

La e-Ciencia aparece en el marco de *Information Society Technologies* (IST) del programa europeo y se define como “*ciencia desarrollada a través de*

colaboraciones globales posibilitadas de forma distribuida en Internet". Una característica común a todas estas colaboraciones científicas es la necesidad de acceso a grandes colecciones de datos, recursos de computación a gran escala (Grid [2]), y rapidez en la visualización de resultados . El desarrollo de la e-Ciencia está ligado a los avances en la computación, comunicación, teledetección y modelado, para mejorar el entendimiento de nuestro entorno y nuestra habilidad de monitorizar y predecir procesos físicos. Habilitar estos recursos, datos y herramientas de forma distribuida, hace posible que los investigadores o usuarios científicos en general puedan compartirlos.[3]. En este capítulo veremos el papel que desempeñan las IDEs y no solo la tecnología Grid, en la e-Ciencia poniendo como ejemplo un proyecto europeo perteneciente a GMES actualmente en ejecución.

2 Antecedentes

Una de la iniciativas que contribuye al desarrollo de la e-Ciencia es GMES[4] (ver sección 2.1) y como parte de ella nace el proyecto AWARE (*A tool for monitoring and forecasting Available Water REsource*)[5] que tiene como objetivo el desarrollo de herramientas novedosas para la automatización de las tareas de monitorización y predicción de la disponibilidad de agua y su distribución por aquellas cuencas hidrográficas donde el deshielo es el factor más importante del balance de agua anual, una condición que tienen en común las cuencas de la región alpina.

La Universitat Jaume I de Castellón (UJI) y el Institut Cartogràfic de Catalunya (ICC) se encuentran entre los participantes del proyecto, específicamente en el papel de proporcionar geo-servicios estandarizados con funcionalidad genérica y de tal forma que estén disponibles para toda la comunidad científica con intereses hidrológicos.

Para facilitar el acceso a estos datos científicos y servicios web el caso ideal es que pertenezcan a una IDE global, una infraestructura de datos espaciales común para la comunidad científica, que sirva de plataforma para la e-Ciencia y posibilite una armonización de los datos espaciales para facilitar su geo-procesamiento.

2.1 GMES

GMES (*Global Monitoring for Environment and Security*) [4] es una iniciativa colaborativa entre la Comisión Europea (EC) y la Agencia Espacial Europea (ESA). Nació bajo la responsabilidad de la Dirección General-Investigación, pasando más tarde a manos de la DG-Industria. GMES puede verse como una iniciativa I+D+I que persigue un intento de armonización política e industrial para el uso de datos pertenecientes al campo medioambiental o de seguridad. Fundamentalmente en este ámbito se trata con tres fuentes de datos espaciales: datos in situ, recogidos por los usuarios desde el mismo campo; datos procedentes de la Observación de la Tierra (OT); y datos cartográficos que pueden provenir de IDEs siguiendo la directiva de INSPIRE. GMES tiene como finalidad armonizar el acceso a estas tres clases de datos, y siguiendo su nueva óptica industrial impulsar la explotación del sector espacial europeo con especial atención en las imágenes satélite (OT).

Los usuarios de datos OT, incluyendo científicos, industria e investigadores, se enfrentan a volúmenes de datos tan variados e inmensos que es difícil extraer información relevante. GMES intenta reunir datos relevantes y servicios novedosos, fáciles de mantener y amigables con los que los responsables de la toma de decisiones sean capaces de trabajar mejor en situaciones críticas relacionadas con la gestión del entorno y la seguridad. De esta forma orienta sus esfuerzos hacia el desarrollo y establecimiento de un número inicial de servicios que deben ser trasladados desde la visión de proyecto a corto plazo a servicios mantenidos a largo plazo.

La característica principal, según GMES, para este grupo de servicios iniciales puede resumirse de la siguiente manera:

- Mapas: información relevante desde escalas locales a europeas
- Pronóstico: información sistemática de corto a largo plazo de la evolución ambiental de la tierra (aire, agua, recursos);
- Actuación en situaciones críticas: información segura y en tiempo real para la seguridad y la protección civil.

2.2 Infraestructuras de datos espaciales

Conectar los científicos de los sectores GMES, con sus datos, modelos y servicios supone una infraestructura muy similar a la que propone el sector de las infraestructuras de datos espaciales (IDE).

Una IDE es fundamentalmente la facilitación y la coordinación del intercambio y la compartición de datos y servicios entre usuarios de diferentes niveles jurisdiccionales en el ámbito de los datos espaciales. Las IDEs se han hecho muy importantes a la hora de determinar la forma en la que los datos son usados desde escalas locales a escalas internacionales. Compartir datos y servicios es muy útil desde el momento que permite al usuario economizar recursos, tiempo y esfuerzo tratando de adquirir nuevos conjuntos de datos, evitando la duplicación de costes asociados a la generación y al mantenimiento de los datos y su integración con otros conjuntos de datos.

Aunque las IDEs son principalmente marcos de colaboración institucional también definen e implementan sistemas de información heterogénea distribuida, mediante cuatro componentes software enlazados vía Internet [6]. Estos componentes tradicionalmente son:

- Editores de metadatos y servicios de catálogo asociados.
- Almacenes de datos espaciales o bases de datos
- Aplicaciones cliente para las búsquedas de usuarios y acceso a datos espaciales.
- Middleware o servicios de geo-procesamiento intermediarios que ayudan a los usuarios a encontrar y transformar datos para su uso en las aplicaciones cliente.

La figura 1 resume los componentes tecnológicos esenciales, aceptados dentro de los organizaciones de estándares geo-espaciales: - Open Geospatial Consortium (OGC) y la ISO Comité Técnico 211 - y sintetizado por el US Federal Geographic Data Committe (FGDC) [9] y la NASA. La arquitectura puede ser interpretada como una arquitectura tradicional de 3-capas, modelo cliente-middleware-servidor, donde las aplicaciones cliente buscan datos espaciales que son encontrados posteriormente, estos datos pueden ser transformados o procesados por servicios intermedios antes de su presentación en ese cliente. La arquitectura puede interpretarse también usando el modelo triangular de servicios web “publish-find-bind” [7] donde servicios de catálogo publican el contenido de los datos espaciales, estos datos son buscados y encontrados posteriormente por servicios y por último enlazados y ejecutados.

Hasta ahora las definiciones de IDEs han puesto poco énfasis en el geo-procesamiento que ocurre posteriormente al descubrimiento y transformación de los datos. El geo-procesamiento se lleva a cabo de 3 maneras posibles:

1. Sistema de información geográfica. (SIG)
2. Cliente pesado, por ejemplo gvSIG (<http://www.gvsig.gva.es/>)

3. Servicios Web.

El proyecto AWARE tiene su enfoque en esas dos últimas formas de geo-procesamiento.

3 OGC Web Processing Services

Open Geospatial Consortium (OGC) define una arquitectura abierta e interoperable - basada en la reutilización de los componentes estandarizados [8] – permitiendo la creación de nuevas aplicaciones de forma flexible y escalable, facilitando además el cambio de componentes por otros similares, ya sea por motivos económicos, de funcionalidad o de rendimiento. Estas aplicaciones ofrecen al usuario la funcionalidad para solucionar un problema en concreto a partir de la unión y/o concatenación de operaciones atómicas ad hoc, en contraste con las aplicaciones GIS monolíticas, donde toda la funcionalidad está implementada en una caja negra a la que no tiene acceso el usuario.

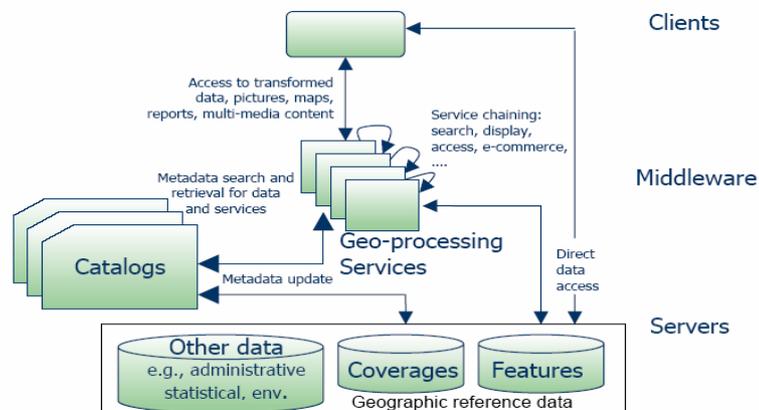


Figura 1. Modelo de referencia de la arquitectura de las IDEs (fuentes: FGDC[9] , e INSPIRE[10])

Hasta el momento, se ha estado haciendo hincapié mayoritariamente en la implementación de servicios OGC de acceso y visualización de información geográfica (WMS, WFS, WCS, servicio de catálogo), sin embargo, con la nueva especificación de OGC (WPS) para acceder a de servicios de geo-procesamiento es

posible realizar operaciones más complejas (ej. análisis espacial, camino mínimo en red, buffer, etc.).

Así nace el nuevo concepto de geo-servicios, servicios web definidos por W3C como "una aplicación software identificado por una URI, cuyo interfaz y enlaces pueden describirse y ser descubiertos como artefactos XML" [11] pero que en este caso encapsulan funcionalidad SIG (es decir, geo-procesamiento).

4 Geo-servicios en AWARE

El propósito del proyecto AWARE es el de proveer mediante el uso integrado de los recursos procedentes de la Observación de la Tierra, herramientas novedosas para la monitorización y predicción de la disponibilidad y distribución del agua proveniente del deshielo en las cuencas alpinas.

El proyecto está dirigido a cubrir la brecha existente entre los proveedores de información ambiental (temperatura, precipitación, elevación...) de una cuenca hidrográfica y los técnicos/científicos que manejan esta información (trabajando para compañías hidroeléctricas, administración local, etc.).

El proyecto aborda esta problemática proporcionando servicios de catálogo para dar acceso a repositorios de datos y servicios geográficos, e implementando servicios de geo-procesamiento sobre la información que manejan.

Partiendo de unos modelos hidrológicos existentes, el proyecto AWARE pretende automatizar su uso desarrollando un conjunto de geo-servicios que implementan los pasos necesarios para calibrar el modelo y realizar las predicciones de esorrentía. Este conjunto de geo-servicios junto con otros componentes, tales como asistentes al usuario e información estática, están disponibles a través de una interfaz web que permite a los usuarios (científicos, en primera instancia) ejecutar de forma sencilla el modelo escogido con sus propios datos a una escala local, obteniendo los resultados concernientes al área de interés indicada.

En la figura 2 se pueden ver los servicios de geo-procesamiento que en conjunción con otros módulos componen el modelo hidrológico SRM [12]. Uno de estos geo-servicios es el que realiza el cálculo de temperaturas por zonas de elevación.

Uno de los pasos intermedios a realizar antes de poder ejecutar estos modelos es el cálculo de la temperatura media que hay en cada una de las zonas de elevación en

que ha sido dividida la cuenca hidrográfica. En el cálculo de este parámetro intervienen cuatro factores tomados a partir de los datos provenientes de imágenes satélite y de los datos in-situ. Estos factores representan la localización de cada una de las estaciones meteorológicas distribuidas por toda la cuenca hidrográfica, las mediciones de temperatura tomadas en estas, un modelo digital de elevación del área de estudio y las zonas de elevación en que ha sido dividida la cuenca hidrográfica.

El número de estaciones meteorológicas de que se dispone en la cuenca es reducido, por lo que hemos de recurrir a cálculos geo-estadísticos (kriging) para obtener la temperatura en los puntos para los que no se dispone de esta información.

Hay varias fuentes de información a las que se necesita acceder para realizar estos cálculos:

- **Datos meteorológicos:** Temperatura media mínima de las estaciones meteorológicas (junto con coordenadas y elevación) obtenidas a través de un Web Feature Service (WFS).
- **Modelo de elevación digital** de del área de estudio en formato GeoTiff ofrecido a través de un Web Coverage Service (WCS).

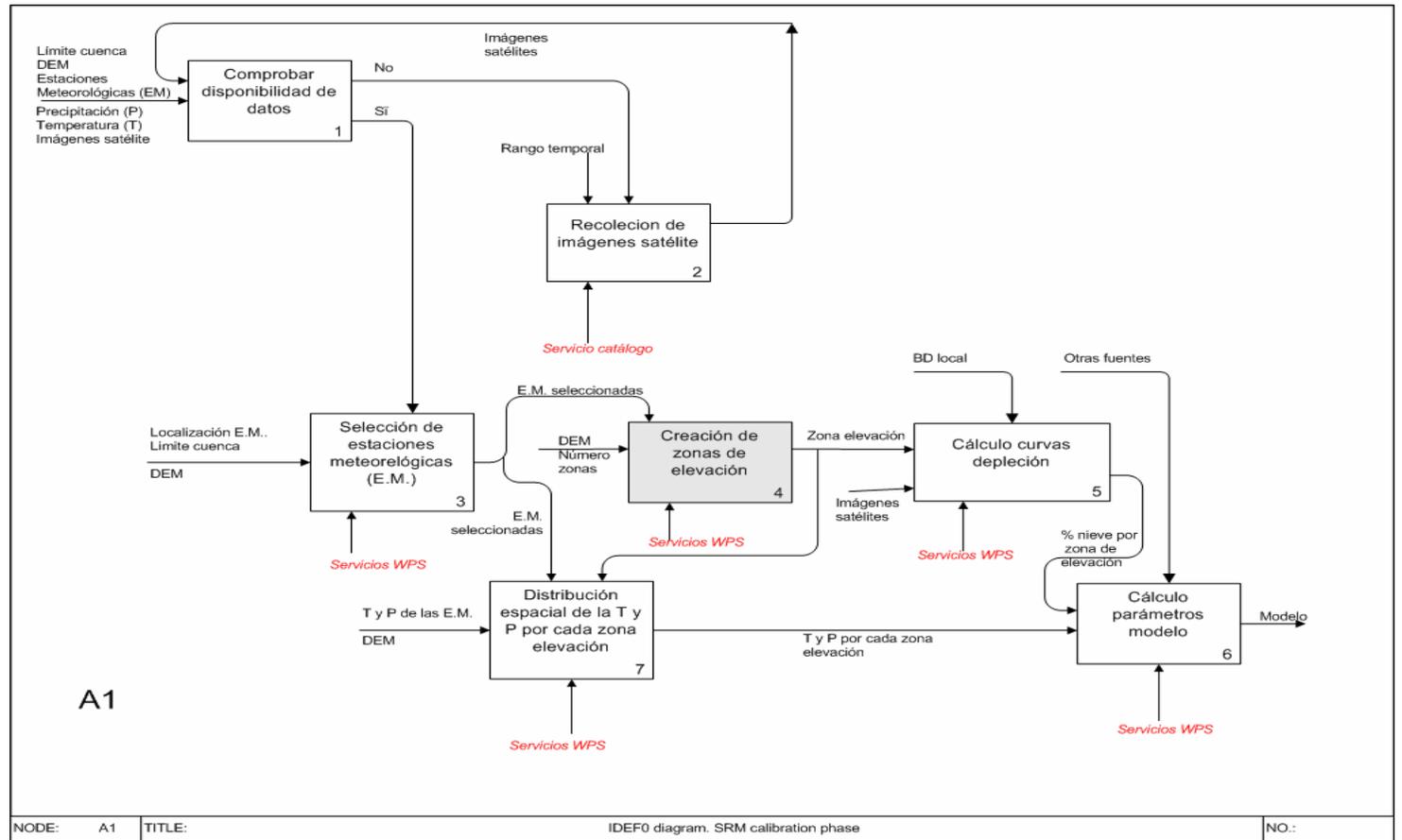


Figura 2 - IDEFO del Modelo SRM. Ilustra el flujo de los datos a través del modelo hidrológico SRM.

4.1. Prototipo inicial

El prototipo WPS creado se ejecuta en un medio distribuido e interoperable, ejemplificando la escalabilidad, flexibilidad y reusabilidad de los componentes en una IDE.

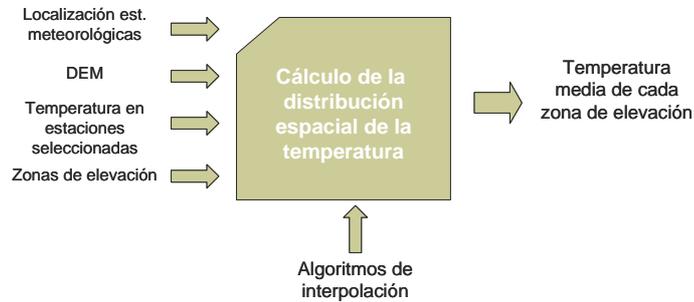


Figura 3 – Modelo conceptual del geo-servicio creado

La implementación del geo-servicio está basada en la especificación para WPS definida por el OGC [13] (versión 0.4.0), de forma que el acceso a los cálculos estadísticos que este ofrece sea realizado de una manera interoperable y estándar. En la figura 3 se puede ver el modelo conceptual del geo-servicio propuesto, a modo de versión simplificada en la que no se muestran aspectos de la implementación.

El OGC Web Processing Service (WPS) proporciona acceso a cálculos y modelos geo-espaciales, del mismo modo que podríamos hacer a través de un sistema GIS tradicional (ej. cálculos estadísticos, análisis de buffers, etc.). A través de sus interfaces podemos describir y anunciar las capacidades ofrecidas por el servicio (GetCapabilities) a través de sus metadatos, identificar los inputs requeridos para realizar los cálculos (DescribeProcess) e iniciar los cálculos, informar del estado de los mismos y devolver los outputs producidos (Execute).

Todas las operaciones del servicio implementado soportan peticiones realizadas a través de los métodos Get (parámetros codificados como pares clave-valor) y Post (documento XML que incluye todos parámetros en etiquetas XML).

Los cálculos geo-estadísticos proporcionados por este servicio son realizados de forma transparente para el usuario a través del programa estadístico R [14] y el

paquete geo-estadístico para R, *gstat* [15], accesibles a través de una pasarela web (*Rserve*, <http://stats.math.uni-augsburg.de/Rserve>).

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Execute service="WPS" version="0.4.0" store="true" status="false" xmlns="http://www.openeospatial.net/wps"
xmlns:ows="http://www.openeospatial.net/ows" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openeospatial.net/wps
..\wpsExecute.xsd">
  <ows:Identifier>CalculateTemperatureMap</ows:Identifier>
  <DataInputs>
    <Input>
      <ows:Identifier>StationTemperature</ows:Identifier>
      <ows:Title>Station Temperature Measures</ows:Title>
      <ComplexValueReference ows:reference="http://geoinfo.dlsi.uji.es/cgi-bin/wps-
test?service=WFS&version=1.0.0&request=GetFeature&typename=TempJan"
schema="http://geoinfo.dlsi.uji.es/cgi-bin/wps-
test?service=WFS&version=1.0.0&request=DescribeFeaturetype&typename=TempJan" />
    </Input>
    <Input>
      <ows:Identifier>ElevationMap</ows:Identifier>
      <ows:Title>ElevationMap</ows:Title>
      <ComplexValueReference ows:reference="http://geoinfo.dlsi.uji.es/cgi-bin/wps-
test?service=WCS&version=1.0.0&request=GetCoverage&bands=1&COVERAGE=dem&am
p;CRS=EPSG:4326&FORMAT=GEOTIFF&HEIGHT=128&WIDTH=146&BBOX=562.5713
55905512,4903.05270866142,660.287891338583,4988.722" />
    </Input>
    <Input>
      <ows:Identifier>NumElevationZones</ows:Identifier>
      <ows:Title>Number of Elevation Zones</ows:Title>
      <LiteralValue>5</LiteralValue>
    </Input>
  </DataInputs>
  <OutputDefinitions>
    <Output>
      <ows:Identifier>InterpolatedTemperatureMap</ows:Identifier>
      <ows:Title>Interpolated temperature in each zone.</ows:Title>
    </Output>
  </OutputDefinitions>
</Execute>
```

Como puede verse en el ejemplo de una petición *Execute* utilizando XML como forma de envío, el geo-servicio recibe los inputs – en este caso las mediciones de la temperatura en las estaciones meteorológicas y su localización (*StationTemperature*) como una consulta al WFS, la capa de elevación en formato *GeoTiff* como una consulta al WCS (*ElevationMap*) y el número de zonas de elevación (*NumElevationZones*) como un número entero – y el output a generar. El geo-servicio devuelve una respuesta XML, indicando si este ha finalizado satisfactoriamente o no, y en caso afirmativo, una URL desde la cual poder acceder

a los resultados. En este caso, un archivo GML en el cual se presentan las temperaturas medias de cada una de las zonas de elevación indicadas y de los polígonos que las definen.

5 Conclusiones

En un ámbito donde la comunidad científica tiene la necesidad de compartir datos y funcionalidades para colaborar y economizar recursos se considera necesario la puesta en marcha de una IDE que resuelva estas necesidades, aportando un entorno abierto e interoperable - basado en la reutilización de los componentes estandarizados proporcionados en ellas – facilitando la creación de nuevas aplicaciones de una manera más sencilla, flexible y escalable. Estas aplicaciones ofrecen al usuario la funcionalidad para solucionar un problema en concreto a partir de la unión y/o concatenación de operaciones simples ya existentes, en contraste con las aplicaciones GIS estándar, donde normalmente solo un pequeño porcentaje de las funcionales ofrecidas es utilizado. Aunque no pretende resolver el problema de composición de servicios [16].

Para que una IDE de estas características funcione de manera global se requieren unas directivas con políticas de colaboración, de accesibilidad, y estándares que posibiliten la conexión semántica y sintáctica de cada elemento. Con este propósito en Europa se adopta INSPIRE como la directiva para la creación de IDEs, utilizando para la interconexión los interfaces definidos por las normas y los estándares del OGC. GMES se apoya en esta directiva para el desarrollo de unos servicios, que formando parte de una IDE, aportarán funcionalidades y datos requeridos en el mundo científico en el ámbito del medioambiente y la seguridad. El proyecto AWARE aporta uno de estos servicios impulsados por GMES, desarrollado como diferentes OGC Web Processing Services, ofreciendo un conjunto de funcionalidades estándar escalables y reutilizables.

Referencias

[1] M. K. Ramamurthy, A new generation of cyberinfrastructure and data services for earth system science education and research. *Advances in Geosciences*, 8: 69 – 78. Junio 2006

[2] I. Foster, C. Kesselman, S. Tuecke, The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *Intl J. Supercomputer Applications* Vol. 15, No. 3. 2001

[3] J. Graff, "E-Science Potencial". International Ocean Systems, 6 – 7. Noviembre/Diciembre 2004.

[4] The "Global Monitoring for Environment and Security"
<http://www.gmes.info/>

[5] AWARE Project: A tool for monitoring and forecasting Available WAter REsource
<http://www.aware-eu.info/es/home.htm>

[6] C. Granell, M. Gould, M.A. Bernabé y M.A. Manso. Spatial Data Infrastructures. In H. Kasimi (Ed.) Encyclopedia of Geoinformatics, Idea Group Press, en prensa.

[7] K. Gottschalk, S. Graham, S. Krueger, J. Snell. Introduction to web services architecture. IBM Systems Journal 41(2), 2002.

<http://researchweb.watson.ibm.com/journal/sj/412/gottschalk.html> (accessed 4 April 2002)

[8] C. Granell. Reutilización de Servicios Web mediante components integrados. Tesis Doctoral. Departamento de Lenguajes y Sistemas Informaticos. Universitat Jaume I. 2006

[9] Federal Geographic Data Committee (FGDC): <http://www.fgdc.gov>
Geospatial Interoperability Reference Model (GIRM):
<http://gai.fgdc.gov/girm>

[10] P. Smits, et al (2002): INSPIRE Architecture and Standards Position Paper. Architecture and Standards Working Group.
http://inspire.jrc.it/documents/inspire_ast_pp_v4_3_en.pdf

[11] Web services definition from W3C Web Services Architecture Working Group, Web Services Architecture Requirements, W3C Working Draft (Aug. 19, 2002);
<http://www.w3.org/TR/2002/WD-wsa-reqs-20020819>

[12] Martinec, J., Rango, A., and Major, E. The Snowmelt Runoff Model (SRM) User's Manual. NASA RP-1100. Greenbelt, Maryland, p. 110. 1983

[13] Open Geospatial Consortium Inc. Web Processing Service. Discussion paper OGC 05-007r4.

[14] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 2006.
URL <http://www.R-project.org>

[15] E. J. Pebesma and C. G. Wesseling, Gstat: a program for geostatistical modelling, prediction and simulation. *Computers & Geosciences* Vol. 24, No. 1, pp. 17-31. 1998.
URL <http://www.gstat.org>

[16] R.Lemmens, C. Granell, *et al.* Integrating Semantic and Syntactic Descriptions for Chaining Geographic Services. *IEEE Internet Computing*. September-October 2006.