

Las Infraestructuras de Datos Espaciales y su aplicación en proyectos de investigación en la Universidad de Zaragoza

R. Béjar, P.R. Muro-Medrano*, J. Zarazaga, J. Nogueras-Iso, M.A. Latre, F.J. López

Depto. de Informática e Ing. de Sistemas, Instituto de Investigación en Ing. de Aragón, Universidad de Zaragoza,
C/María de Luna 1, 50018, Zaragoza, España

R. Rioja, S. Laiglesia, J. Barrera, R. Miguel, M.J. Pérez

GeoSpatiumLab S.L. , C/Carlos Marx, 6, 50015 Zaragoza, España

(24 de febrero de 2010)

Este artículo tiene por objeto proporcionar una panorámica de algunos aspectos de interés para la investigación relacionados con las infraestructuras de datos espaciales que se están llevando a cabo en la Universidad de Zaragoza. Este trabajo parte de los resultados y experiencia de nuestro grupo de investigación en los últimos 12 años en el campo de la interoperabilidad de la información geoespacial en Web, especialmente bajo el paradigma de las Infraestructuras de Datos Espaciales. Sobre esa base, se revisan las problemáticas que en este momento parecen más oportunas tanto por sus aspectos científico/técnicos, como por sus posibles repercusiones en la transferencia de conocimiento y tecnología. Después de hacer una semblanza sobre el origen y la motivación por las Infraestructuras de Datos Espaciales, se pasa a hacer una revisión de aspectos como infraestructuras de información, arquitecturas, metadatos de datos y servicios geoespaciales, recursos web, bases de conocimiento geoespacial, georreferenciación semántica, geoparsing, geocoding, crawlers e integración semántica de servicios. Aspectos que se están abordando en el marco de diversos proyectos de investigación.

Keywords: Infraestructuras de Datos Espaciales; Servicios web; Metadatos; Arquitecturas; Recuperación de Información Geoespacial; Geoparsing; Geocoding; Semántica geoespacial;

1. Introducción

Dentro del ámbito de las infraestructuras de datos espaciales, en el que la Universidad de Zaragoza ha venido trabajando en los últimos años, el presente trabajo centra su atención en algunos aspectos de interés, todavía no especialmente desarrollados que en nuestra opinión tienen un amplio margen de desarrollo, y más concretamente, en las capacidades para reconocer y aprovechar el potencial informativo gráfico, relacional y semántico de información geoespacial no explícita, relacionada con esquemas de representación espacial basados en identificadores geográficos. Es esta una tipología de información que resulta especialmente abundante en la Web y los sistemas de información corporativos. Este trabajo se focalizará en revisar los avances en el conocimiento científico y tecnológico de componentes y servicios interoperables especializados en el trabajo con esquemas de referencia espacial basados en identificadores geográficos como Geoparsing, Geocoding, Reverse Geocoding y publicación de los contenidos propios como identificadores (p.e. utilizando la aproximación de Gazetteer propuesta por OGC), para que sean capaces de aprovechar peculiaridades de la semántica de la información; de las herramientas y metodologías que permitan dar soporte a estos procesos incluyendo GKB (bases de conocimiento geoespacial) y ETL (extract, transform and load) especializados; así como el aprovechamiento de estos procesos para la orquestación/encadenamiento de servicios geoespaciales orientada al desarrollo de utilidades y sistemas finalistas. Así mismo, se abordan aspectos conceptuales y tecnológicos relacionados con metadatos, modelos, servicios y aspectos arquitecturales

*Autor al que enviar la correspondencia. Email: prmuro@unizar.es

El artículo se ha organizado de la siguiente forma, después de una breve explicación de los antecedentes del grupo de investigación en esta materia, se hace una semblanza sobre el origen y la motivación por las Infraestructuras de Datos Espaciales, para pasar después a hacer una revisión de distintos aspectos que nos parecen de interés desde el punto de vista de investigación y desarrollo, aspectos que se están abordando en el marco de proyectos de investigación financiados por la Comisión Europea, el Gobierno de España, el Gobierno de Aragón, el Instituto Geográfico Nacional y GeoSpatiumLab que es una empresa Spin Off de la Universidad de Zaragoza.

1.1 Antecedentes

El trabajo de la Universidad de Zaragoza en esta temática comenzó con un proyecto de la CICYT en 1998 donde se abordaba la problemática de la interoperabilidad de sistemas de información geoespacial y se hacía una apuesta muy fuerte por la aproximación del OGC (Open Geospatial Consortium). Este trabajo tuvo su continuidad en el año 2000 con otro proyecto coordinado de la CICYT en colaboración con la Universidad Jaume I y la Universidad Politécnica de Madrid y el apoyo posterior del Instituto Geográfico Nacional, donde se proponía la idea de una infraestructura nacional de información geográfica y se empezó a trabajar en su arquitectura y las problemáticas de sus componentes centrales como el servicio de catálogo, el papel de los metadatos y otros servicios interoperables de Geoprocesamiento, principalmente servicios WMS. El concepto de este tipo de infraestructura pasó a denominarse Infraestructura de Datos Espaciales (IDE) y en este momento está perfectamente asentado. Este tipo de infraestructura recibió un impulso significativo en el ámbito europeo con la iniciativa INSPIRE (INfrastructure for SPatial InfoRmation in Europe) que, con el apoyo técnico del Joint Research Center de la Comisión Europea, se convirtió en Directiva Europea aprobada en marzo de 2007, y en la que nuestro grupo de investigación está profundamente implicado. Distintos aspectos científico/técnicos de la arquitectura, los componentes y aplicaciones de las IDE han sido de investigación de nuestros anteriores proyectos de la CICYT. Esto nos ha permitido desarrollar un conocimiento avanzado en aspectos relativos a metainformación, anotación semántica de recursos, modelos, arquitecturas y patrones en el ámbito de las IDE, y aspectos metodológicos. Toda esta investigación y desarrollo tecnológico se ha visto complementada con la participación como expertos en comités de estandarización (CEN, ISO, OGC, AENOR); la participación en una cantidad significativa de proyectos de I+D y transferencia de tecnología del Plan Nacional, y con instituciones y empresas entre los que destacan el desarrollo de la IDEE (nodo de IDE nacional de referencia) y SDIGER (proyecto piloto de la Directiva europea INSPIRE); así como la participación como expertos en las reglas de implementación de la Directiva INSPIRE.

2. El área multidisciplinar de las Infraestructuras de Datos Espaciales

Los últimos años están presenciando una significativa y creciente proliferación de la información geoespacial, tanto en cuanto a su disponibilidad como en sus posibilidades de utilización. Dicha proliferación se debe a aspectos como el auge de los dispositivos móviles (p.e. GPS, navegadores, teléfonos móviles, PDA, ...) y redes, la disponibilidad de cartografía más completa, precisa y barata, el auge de satélites de observación de la tierra, las posibilidades de las administraciones y empresas para la creación de información geoespacial temática, la popularización de aplicaciones y servicios gratuitos (p.e. Google maps, Google Earth, Microsoft Virtual Earth, ...), los acuerdos para la estandarización (OGC, ISO) y su implantación, el soporte para la interoperabilidad de componentes comerciales y de código libre, así como la implicación de las autoridades nacionales (Real Decreto 1545/2007 por el que se regula el Sistema Cartográfico Nacional) y europeas (Directiva Europea 2007/2/EC INSPIRE) para facilitar el descubrimiento, acceso y utilización de la información y servicios geoespaciales.

2.1 Origen y motivación de las IDE

La Información Geográfica (IG), también conocida como datos espaciales o datos geoespaciales, es la información que describe los fenómenos asociados directa o indirectamente con una localización respecto a la superficie terrestre. Actualmente, hay grandes cantidades de datos geográficos que han ido recogiendo (durante décadas) con diferentes propósitos y por diferentes instituciones y empresas. Por ejemplo, la información geográfica es vital para sistemas de toma de decisiones y sistemas de gestión de recursos (recursos naturales, redes básicas de suministro, catastro, economía, agricultura...) a diferentes niveles (local, provincial, regional, nacional o incluso global) (Buehler and McKee eds.). Además, el volumen de información crece día a día debido a importantes avances tecnológicos en sensores remotos de alta resolución (GPS), bases de datos y software de geoprocésamiento, sin olvidar el interés creciente de personas individuales e instituciones. Incluso es posible georeferenciar complejas colecciones de un amplio rango de tipos de recursos, incluyendo documentos textuales y gráficos, mapas digitales, imágenes, observaciones e tiempo real o bases de datos legadas de registros tabulares históricos.

En los últimos años, las naciones han hecho inversiones sin precedentes tanto en información como en medios para recopilar, almacenar, procesar, analizar y diseminar IG. Miles de organizaciones y agencias (distintos niveles gubernamentales, sectores privados, organizaciones sin ánimo de lucro, centros de investigación) alrededor del mundo gastan billones de Euros por año para producir y usar información geográfica (Somers 1997, Groot and McLaughlin 2000). Esto ha sido facilitado por el rápido avance de tecnologías de captura, que ha hecho que la captura de IG sea relativamente rápida y fácil. Adicionalmente, hay que mencionar el impacto de Internet en la distribución de recursos de IG. Así como otros tipos de recursos, cantidades de recursos de IG están disponibles en Internet. Y en algunos casos incluso se asume que la propia Internet es el almacén de información.

Sin embargo, es frecuente que cada nuevo proyecto o estudio que implique el uso de IG requiera la creación de nuevos recursos desde cero. Esta falta aparente de recursos reusables suele venir motivada por las siguientes circunstancias: la falta de recursos financieros provoca que se hagan esfuerzos puntuales pero sin un mantenimiento a largo plazo; algunas organizaciones, a pesar de ser públicas, son reticentes a distribuir información geográfica de alta calidad; los datos recopilados por instituciones distintas son a menudo incompatibles; la falta de conocimiento, en la mayor parte de los casos, de lo que hay actualmente disponible; la poca calidad y la poca documentación de lo que hay disponible a través del Web; y la creciente complejidad y disparidad de los sistemas de búsqueda recuperación a través de Internet.

En conclusión, a pesar de los usos potenciales de la IG y de las inversiones en su creación, la IG no se explota suficientemente (Craglia et al. 1999, European Commission 1998, Official Journal of the European Union 2003). Se dice que "la información es poder" pero con las crecientes cantidades de datos que se crean y almacenan (a menudo, no de forma organizada) hay una necesidad real de documentar los datos para su futuro uso (para que sean lo más accesibles posibles y a un público más amplio). La información se define como el conjunto de los datos más el contexto para su uso. Los datos sin el contexto no son tan valiosos como los datos documentados. Esta necesidad es de extrema importancia en el caso de la información geográfica. Una vez que se crean, los datos geográficos pueden ser usados por múltiples sistemas y para diferentes propósitos.

Como se puede deducir, hay una necesidad de crear soluciones distribuidas que faciliten la búsqueda, evaluación y acceso a los datos. Las Infraestructuras de Datos Espaciales (IDE) son las infraestructuras orientadas a la optimización de la creación, mantenimiento y distribución de información geográfica a diferentes niveles organizativos (ej., regional, nacional, global) e involucrando tanto a instituciones públicas como privadas (Nebert 2001). La primera definición formal del término IDE fue la formulada en el Registro Federal de los EE.UU. en 1994 (U.S. Federal Register 1994): "Una IDE significa la tecnología, políticas, normas, y recursos humanos para adquirir, procesar, almacenar, distribuir, y mejorar la utilización de datos geoespaciales". La definición dada por el GSDI es también muy parecida: "Una aproximación coordinada de tecnología, políticas, normas, y recursos humanos necesaria para una eficiente adquisición, gestión, almacenamiento, distribución y utilización de datos geoespaciales en el desarrollo de una comunidad global".

Dada la importancia que está adquiriendo la información geográfica, los gobiernos empiezan a considerar las infraestructuras de datos espaciales como infraestructuras básicas para el desarrollo de una nación. Se

están convirtiendo en unos elementos tan importantes como las infraestructuras básicas de suministros (electricidad, agua, gas), transporte o telecomunicaciones. Esta consideración, ya con cierta historia en países como EE.UU. (U.S. Federal Register 1994) o Canadá, ha recibido un impulso muy importante en el ámbito europeo al lanzarse desde la Comisión Europea la iniciativa INSPIRE (INfraestructure for SPatial InfoRmation in Europe) (European Commission 2007). INSPIRE tiene como objetivo el establecimiento de las bases, tanto técnicas como políticas, para poder crear un verdadero espacio común europeo de la información geográfica, en el que todo el mundo sepa cómo poder acceder a servicios Web relacionados con la información geográfica con independencia del país o la región europea en el que se encuentren. Por delegación, los países miembros y, en España, las CC. AA. tendrán el mandato de facilitar el desarrollo de la infraestructura en su área de influencia. Inicialmente se ha planteado en la temática medioambiental y posteriormente se irán desarrollando otras áreas (agricultura, transportes, ...).

Como el concepto de las IDE viene del dominio de la información geográfica, las primeras IDE se construyeron, desde el punto de vista técnico, en base a los conceptos y experiencias proporcionadas por los Sistemas de Información Geográfica (SIG) tradicionales. El término SIG se utiliza comúnmente para referirse a los paquetes de software que son capaces de integrar datos espaciales y no espaciales para obtener la información que es necesaria para la toma de decisiones. Sin embargo, estos SIG utilizados como herramientas para realizar análisis particulares en proyectos aislados se expandieron a entornos distribuidos y cooperativos, no sólo desde la perspectiva técnica sino teniendo en cuenta también políticas de cooperación entre diferentes organizaciones públicas o privadas y a distintos niveles (local, regional, nacional o global).

Actualmente, el desarrollo de infraestructuras de datos espaciales constituye un contexto aplicativo multidisciplinar que combina la experiencia y conocimientos de diferentes disciplinas. En particular, las Bibliotecas Digitales proporcionan una base de conocimiento muy importante (Zarazaga et al. 2000, Cantán et al. 2009, Béjar et al. 2009c). Allí se tiene una gran experiencia en tecnología para la distribución de recursos digitales que pueden ser usados como base para los conceptos propios de las ide, además de procesos y métodos.

3. Componentes de una Infraestructura de Datos Espaciales

Según (Coleman and Nebert 1998), entre los componentes de una IDE se deben incluir los siguientes:

- Tecnología. Las IDE se deben desarrollar sobre los componentes tecnológicos creados de la experiencia adquirida trabajando con tecnología de información genérica. Uno de los retos más importantes es la integración de toda esta experiencia, en especial la proporcionada por los SIG.
- Normas (estándares) y guías. Los estándares constituyen el enlace entre los diferentes componentes de una IDE proporcionando lenguajes y conceptos comunes que hagan posible su comunicación y coordinación. Adicionalmente, es necesario establecer guías que sean seguidas por todos los actores involucrados en las IDE. Estas guías deberían incluir distintos aspectos tales como arquitecturas, procesos, métodos y estándares.
- Recursos humanos. El desarrollo de una IDE se debe hacer sobre las necesidades de sus usuarios, tanto usuarios finales como productores de datos. Y por otro lado, el trabajo de implementar y mantener una IDE debe ser realizado por equipos cualificados de investigadores y desarrolladores. Todas estas personas integran los recursos humanos que son necesarios para el desarrollo de IDE.
- Acuerdos institucionales. Es necesario establecer decisiones políticas que permitan crear un marco institucional. Los acuerdos deben servir para establecer IDE a nivel local, y coordinarse para la creación de IDE regionales, nacionales o incluso a nivel global. Aunque los aspectos políticos no son el núcleo de las IDE, si que ejercen una gran influencia sobre el desarrollo las mismas (Nogueras-Iso et al. 2004, Zarazaga-Soria et al. 2004).
- Bases de datos espaciales y metadatos. Las IDE deben proporcionar acceso a datos geográficos, almacenados en bases de datos espaciales, y debidamente documentados a través de una serie de metadatos.
- Redes de datos. Las IDE deben ser sistemas abiertos desplegados sobre redes de datos que proporcionen el canal de acceso a los servicios que accedan desde sistemas remotos.

4. Algunos aspectos de interés para investigación en IDE

En los siguientes apartados pasamos a repasar distintos aspectos de las Infraestructuras de Datos Espaciales que han recibido nuestra atención, aspectos que serán la base de los trabajos de los proyectos de investigación.

4.1 IDE como Infraestructuras de Información

Las Infraestructuras de Información (II) son grandes sistemas de información distribuidos que no son propiedad de, ni están controlados por, una sola organización. Están compuestos de componentes sociales y técnicos, interconectados en redes, distribuidos en grandes áreas, son compartidos por una comunidad de usuarios, proporcionan soporte a ciertas tareas y aplicaciones y han ido evolucionando sobre una base instalada Star (1996), Coleman (1997), Hanseth (1998). Estas características hacen posible considerar las II como un tipo de Sistema de Sistemas (SoS por sus siglas en inglés), centrado en compartir información. Un SoS es un sistema distribuido compuesto por varios sistemas cuya gestión es independiente, al menos hasta cierto punto. Además de esto, un SoS tendrá que mostrar comportamientos emergentes, complejos y no fácilmente predecibles a partir de los sistemas que lo componen, y un desarrollo evolutivo Maier (2008), DeLaurentis (2004), Boardman (2006). Los SoS son difíciles de abordar con las técnicas de ingeniería tradicionales. Éstas se adaptan bien a sistemas con un control centralizado, donde todos sus componentes son visibles globalmente, y donde hay estructuras jerárquicas y actividades coordinadas, sin embargo en los SoS el control está distribuido, y podemos encontrar cooperación, influencias, efectos en cascada, orquestación y otros comportamientos emergentes difíciles de abordar en la actualidad Fisher (2006).

Las Infraestructuras de Datos Espaciales (IDE) son un tipo especializado de II Georgiadou (2006), Nedovic-Budic (2006), cuyo objetivo principal es compartir información espacial. Las IDE suelen estar apoyadas en diversos esfuerzos provenientes de las administraciones públicas, como marcos legales, fondos públicos y cierta relación con otras iniciativas de gobierno electrónico Parliament and Council (2007), Annoni (2002), Zarazaga-Soria et al. (2004). Por ello pueden considerarse cercanas a las Infraestructuras de Información como una política pública tal y como son descritas en Borgman (2003). Esta visión, que ya ha sido ya puesta de manifiesto por nuestro grupo de investigación Béjar et al. (2009b), está atrayendo también nuestra atención sobre aspectos arquitecturales de las IDE Béjar et al. (2009a).

4.2 Metadatos

Los metadatos se definen comúnmente como “datos estructurados acerca de los datos” o “datos que describen los atributos de un recurso” o más simplemente “información acerca de los datos”. Describen el contenido, calidad, condición y otras características de un recurso, constituyendo el mecanismo para caracterizar datos y servicios de forma que usuarios (y aplicaciones) puedan localizar y hacer uso de esos datos y servicios. Los registros de metadatos se suelen publicar a través de sistemas de catálogos, en ocasiones también denominados como directorios o registros. Los catálogos electrónicos no difieren demasiado de los catálogos tradicionales de una biblioteca excepto por el hecho de que ofrecen una interfaz estandarizada de servicios de búsqueda. Así pues, estos catálogos son la herramienta que ponen en contacto a los consumidores con los productores de información.

Para extender el uso y entendimiento de los metadatos por diferentes comunidades de usuarios, por ejemplo para permitir las búsquedas distribuidas a través de una red de catálogos, es preciso utilizar contenidos bien definidos, y por lo tanto es preciso un estándar (norma) de metadatos. Sin embargo, se debe tener en cuenta que estas descripciones varían según el tipo de recurso a describir. Autoridades de reconocido prestigio en cada dominio de aplicación establecen una estructura específica de metadatos, esto es, un esquema o un estándar de metadatos.

Aparte del estándar elegido, cabe mencionar que los sistemas de catalogación de metadatos deben soportar (reconocer) tres formatos de metadatos (Nebert 2001): el formato de implementación (dentro de una base de datos o sistema de almacenamiento), el formato de exportación o codificación (diseñado para la transferencia de metadatos entre distintos sistemas y computadores), y el formato de presentación (un

formato apropiado para ser leído por las personas). Para los dos últimos formatos, hay un consenso general respecto al uso de XML (eXtensible Markup Language) (Bray et al. 2000) dado que es un lenguaje de marcado con reglas estructurales forzadas a través de un fichero de control (XML-Schema) que permite validar la estructura del documento, es decir, comprobar la conformidad respecto a un estándar de metadatos. Además, a través de una especificación complementaria (eXtensible Stylesheet Language ó XSL)(W3C 2004a), un documento XML puede ser usado junto a una hoja de estilo (expresada en XSL) para crear presentaciones o informes según los requerimientos del usuario.

A continuación se detallan diferentes esquemas de metadatos en el dominio de las IDE: esquemas para la descripción de IG; estándares de propósito general que aunque no describen directamente la IG, se usan también en este contexto; y esquemas para la descripción de servicios.

4.2.1 Metadatos geográficos. Tal como se menciona en (Nebert 2001), los metadatos geográficos ayudan a las personas involucradas en el uso de información geográfica a encontrar los datos que necesitan y a determinar la mejor manera de cómo usarlos. En (Federal Geographic Data Committee (FGDC) 2000) se reconoce que la creación de metadatos geográficos persigue tres objetivos (y a su vez beneficios) principales:

- Organizar y mantener la inversión en datos hecha por una organización. Es decir, los metadatos buscan fomentar la reusabilidad de metadatos sin tener que recurrir al equipo humano que se encargó de su creación inicial. Aunque la creación de metadatos parezca un coste adicional, el valor de los datos a la larga es dependiente de su documentación.
- Publicitar la existencia de la información geográfica a través de sistemas de catálogo. Publicando los recursos de información geográfica a través de un catálogo, las organizaciones pueden encontrar: datos a usar; otras organizaciones con las que compartir datos y esfuerzos de mantenimiento; y clientes para esos datos.
- Proporcionar información que ayude a la transferencia de los datos. De hecho, los metadatos deberían acompañar siempre a los propios datos. Los metadatos ayudan al usuario u organización que los recibe en el procesamiento, interpretación, y almacenamiento de los datos en repositorios internos.

Respecto a los estándares de metadatos geográficos existentes en la actualidad, hay que destacar dos propuestas principales a nivel internacional:

- El estándar Content Standard for Digital Geospatial Metadata(CSDGM) (Federal Geographic Data Committee (FGDC) 1998, 2000). Fue creado en 1994 por el Comité Federal de Datos Geográficos (Federal Geographic Data Committee -FGDC) de EE.UU. para dar soporte a la construcción de una Infraestructura Nacional de Datos Espaciales. Aunque es un estándar a nivel nacional, fue el primero en aparecer y se ha difundido a nivel internacional dada su integración en diversas herramientas SIG o su utilización en redes de distribuidas de catálogo a nivel internacional. De hecho, este estándar se adoptó en otros países como Sudáfrica o Canadá.
- El estándar internacional ISO 19115 (International Organization for Standardization (ISO) 2003a, Maganto et al. 2008) para metadatos de información geográfica. La organización internacional ISO (International Organization for Standardization) creó en 1992 el comité 211 (ISO/TC 211) con responsabilidades en Información Geográfica y Geomática. Este comité se ha encargado de preparar una familia de estándares en este contexto. El estándar ISO19115 se aprobó en Mayo de 2003 y define elementos que permiten describir, entre otros, la identificación, extensión, calidad, el esquema de representación espacial, los sistemas de referencia utilizados, y la forma de distribución de los datos. Y aunque este estándar está principalmente orientado a la catalogación de conjuntos de datos geográficos (incluyendo también series o features/entidades individuales) en formato digital, también se puede extender a otras formas de datos geográficos como mapas, documentos textuales o datos no geográficos.

Aparte de estas propuestas principales, ha habido otras muchas iniciativas surgidas a nivel nacional ¹, europeo ² o dentro de un dominio específico. Sin embargo, todas estas iniciativas han ido derogándose en busca de la armonización con el estándar internacional ISO19115. Incluso, la nueva versión del estándar americano CSDGM convergerá con el estándar internacional.

Independientemente del estándar de metadatos utilizado, es usual clasificar los elementos de metadatos respecto al rol que cumplen dentro del paradigma “descubrimiento, evaluación y acceso.” establecido en (Nebert 2001):

- Los metadatos de descubrimiento son aquellos elementos que permiten describir mínimamente la naturaleza y contenido de un recurso. Estos elementos suelen responder a las preguntas “Qué, Por qué, Cuando, Quien, Dónde y Cómo”. Los elementos típicos en esta categoría serían el título, la descripción del conjunto de datos o su extensión geográfica.
- Los metadatos de exploración proporcionan la información que permiten verificar que unos datos que los datos se ajustan al propósito deseado, permiten evaluar sus propiedades, o permiten contactar con la entidad que facilitará más información.
- Los metadatos de explotación incluyen aquellas descripciones necesarias para acceder, transferir, cargar, interpretar y utilizar los datos en la aplicación final que los explote.

Otro aspecto importante relativo a los esquemas de metadatos es su nivel de detalle, que viene definido mediante la elección del propio estándar y la creación de extensiones especiales y perfiles. En primer lugar el estándar elegido define un conjunto más o menos grande de elementos con diferente condicionalidad: obligatorios, obligatorios si aplicable y opcionales. Una extensión del estándar consiste habitualmente en la adición de nuevas restricciones (ej., conversión de elementos opcionales en obligatorios), ampliación de listas de códigos y la creación de nuevos elementos y entidades. ISO 19115 y CSDGM proporcionan métodos dentro del propio estándar para la extensión de los metadatos. Y si esas características adicionales son muy amplias (involucran la creación de un número considerable de elementos), ISO19115 recomienda la petición formal de creación de un perfil de aplicación específico para esa comunidad de usuarios que lo requiera.

Sin embargo, aunque los perfiles específicos y la condicionalidad de los elementos facilitan cierta flexibilidad de los metadatos geográficos, hay que reconocer que resultan todavía muy detallados y complejos de manejar. CSDGM e ISO19115 definen más de 350 elementos cada uno distribuidos en múltiples secciones jerárquicas. Esta complejidad implica que para completar los metadatos geográficos, haya que dedicar gran cantidad de tiempo y recursos humanos altamente cualificados. Este problema está provocando que muchas organizaciones se planteen el uso de otros estándares de metadatos de propósito general (ver a continuación las propuestas de Dublin Core), focalizando los esfuerzos disponibles en mantener al día al menos los metadatos de descubrimiento que se mencionaban con anterioridad.

4.2.2 Metadatos de propósito general. Un buen ejemplo de normas de metadatos de propósito general es la propuesta de la iniciativa “Dublin Core Metadata Initiative”(DCMI) (Dublin Core Metadata Initiative (DCMI) 2004). Esta iniciativa, creada en 1995, promueve la difusión de estándares/normas de metadatos interoperables y el desarrollo de vocabularios de metadatos especializados que permitan la construcción de sistemas de búsqueda de información más inteligentes. Dublin Core es una norma para la descripción de recursos de información en dominios cruzados, es decir, descripción de todo tipo de recursos independientemente de su formato, área de especialización u origen cultural. Esta norma consiste en quince descriptores básicos que son el resultado de un consenso internacional e interdisciplinario. Actualmente, Dublin Core se ha convertido en una parte importante de la infraestructura emergente de Internet.

Dublin Core ha atraído un apoyo internacional y multidisciplinar ya que muchas comunidades estaban ansiosas por adoptar un núcleo común semántico para la descripción de recursos. La norma Dublin Core

¹Un ejemplo de norma surgida a nivel nacional es la propuesta española MIGRA (Mecanismo de Intercambio de Información Geográfica Relacional formado por Agregación) (Comité Técnico de Normalización 148 de AENOR(AEN/CTN 148) 1998).

²Norma voluntaria europea prEnv 12657 elaborada en 1998 por el Comité Europeo de Normalización (CEN) (European Committee for Standardization(CEN) - CEN/TC 287 1998).

se ha traducido a más de 20 idiomas, y tiene un carácter oficial ya que se ha aprobado como norma americana (ANSI/NISO Z39.85) (American National Standards Institute (ANSI) 2001), se ha adoptado dentro del comité técnico europeo CEN/ISSS (European Committee for Standardization / Information Society Standardization System), y desde Abril de 2003 también tiene carácter de norma ISO internacional (ISO 15836) (International Organization for Standardization (ISO) 2003b).

La simplicidad de Dublin Core, solo quince descriptores básicos, permite un fácil emparejamiento con otros esquemas de metadatos más específicos (Tolosana-Calasanz et al. 2006). Por tanto, muchas organizaciones en el dominio de la IG consideran la adopción de Dublin Core en determinadas situaciones: para servir de formato de intercambio entre sistemas que utilizan distintos estándares de metadatos (por ejemplo, exponer metadatos geográficos ISO19115 en otras comunidades interesadas en la utilización puntual de información geográfica); para hacer recopilación (harvesting) de metadatos provenientes de distintas fuentes; o para facilitar la rápida creación de contenidos de catálogo.

Otra ventaja de Dublin Core es la flexibilidad ofrecida para la creación de perfiles de aplicación para dominios específicos. El concepto de perfil de aplicación nació dentro de Dublin Core como medio de declarar qué elementos se deberían usar en una aplicación, proyecto o dominio particular. Los perfiles de aplicación se adaptan al dominio específico mediante la posibilidad de definir cualificadores (refinamientos y esquemas de codificación) o la inclusión, si fuese necesario, de nuevos elementos definidos en el dominio específico. Por ejemplo, el CEN/ISSS workshop ha desarrollado un perfil de aplicación geoespacial de Dublin Core donde ha participado el candidato (Zarazaga-Soria et al. 2003). Este perfil geoespacial define: los elementos tomados del modelo general de Dublin Core; el dominio de valores de cada elemento (especificando el uso de esquemas de codificación específicos); los cualificadores y elementos adicionales que se toman del dominio particular de la IG; y la condicionalidad y ocurrencia de los elementos.

Adicionalmente, también cabe destacar el uso combinado de Dublin Core con la tecnología RDF (Resource Description Framework): RDF (Manola and Miller 2004) y RDF Vocabulary Description Language (RDFS) (Brickley and Guha 2004). RDF está adquiriendo gran importancia porque es una de las tecnologías básicas en la nueva concepción de la Web: la Web Semántica. Según (W3C 2004b, Berners-Lee et al. 2001), “la Web Semántica es la extensión de la Web actual dentro de la cual la información recibe un significado bien definido, permitiendo que computadores y personas puedan trabajar en cooperación”.

RDF es una recomendación W3C para el modelado e intercambio de metadatos, que se expresa en formato XML. La mayor ventaja de RDF es su flexibilidad. RDF no es realmente un estándar de metadatos al uso definiendo una serie de elementos. Por el contrario, se puede considerar como un metamodelo que permite la definición de esquemas de metadatos o la combinación de ellos. El núcleo básico de RDF únicamente ofrece un modelo simple para describir relaciones entre recursos en términos de propiedades con un nombre asociado y una serie de valores. Para la declaración y la interpretación de esas propiedades RDFS proporciona un amplio conjunto de constructores que permite definir y restringir la interpretación de los vocabularios en una comunidad de información particular.

4.2.3 Metadatos para la descripción de servicios. En el contexto específico de la Información Geográfica, el mecanismo propuesto por el Comité Internacional ISO/TC211 para la descripción de servicios es la utilización de una extensión de la norma ISO 19115:2003 (ya descrita anteriormente para la documentación de datos geográficos) para la descripción de las características propias de los servicios. Esta extensión de la norma ISO 19115:2003 viene definida dentro de la norma ISO 19119:2005 (ISO 2003), la cual establece las bases generales de los servicios de información geográfica, incluyendo los modelos de metadatos para su descripción.

Siguiendo la misma propuesta que ISO/TC211, la Comisión Europea aprobó recientemente unas normas de ejecución de metadatos en conformidad con la directiva INSPIRE para la descripción de datos y servicios (European Commission 2008). Estas normas de ejecución definen a un nivel abstracto aquellos descriptores que resultan esenciales para el descubrimiento de datos y servicios. Además, estas normas van acompañadas de guías que establecen la correspondencia entre estos descriptores y las normas de metadatos internacionales más reconocidas como son ISO 19115:2003 o Dublin Core. En particular, la guía que establece la correspondencia entre las normas de ejecución e ISO 19115:2003 (European Commission, 2008b) también contempla la correspondencia de los descriptores específicos de servicios con la norma ISO

19119:2005 (Nogueras-Iso et al. 2009).

Sin embargo, estas propuestas de INSPIRE e ISO/TC211 parecen resultar insuficientes para las necesidades actuales de construcción de aplicaciones siguiendo Arquitecturas Orientadas a Servicios (Lieberman 2003), donde se potencia la descripción sintáctica y semántica de Servicios Web.

Actualmente, el desarrollo de los servicios ofrecidos por una IDE (Infraestructura de Datos Espacial), y en general el desarrollo de servicios en cualquier tipo de infraestructura distribuida, viene guiado por la Arquitectura de Servicios Web (Web Services Architecture) (Booth et al. 2004) propuesta por W3C (World Wide Web Consortium). Esta arquitectura pretende proporcionar los medios estandarizados para interoperar entre diferentes aplicaciones software (Servicios Web), ejecutándose sobre una amplia variedad de plataformas. Según el W3C Web Services Glossary (Haas and Brown 2004), un Servicio Web (Web service) es un sistema software diseñado para soportar interacciones máquina a máquina sobre una red. Además, se menciona específicamente que su interfaz debe ser descrita en un formato procesable automáticamente con el fin de que otros sistemas puedan interactuar con el Servicio Web en la forma prescrita por su descripción utilizando mensajes SOAP, transportados normalmente sobre HTTP y serializados en XML (junto con la utilización de otros estándares Web relacionados).

Por tanto, el gran impacto de los Servicios Web ha incrementado la importancia de los metadatos que describen las capacidades de procesamiento de los servicios. Los detalles de un Servicio Web se pueden publicar en un catálogo para que una petición de un cliente a ese catálogo pueda desembocar en una invocación al servicio descrito por el catálogo.

El estándar más relevante y aceptado para metadatos que describen la sintaxis de los servicios Web es WSDL (Web Services Description Language) (Christensen et al. 2001). Especificaciones de Directorios (Registros) de Servicios como UDDI (OASIS 2004) utilizan WSDL para describir las conexiones a los servicios registrados. La principal desventaja de WSDL es que no tienen la habilidad de caracterizar el contenido semántico proporcionado por esos servicios.

OWL (Web Ontology Language)¹ es una de las opciones para formalizar estas descripciones semánticas, y más concretamente OWL-S², que es una ontología de nivel superior basada en OWL que modela características de servicios Web, y que se utiliza para anotar las descripciones de los servicios con información semántica. Otras opciones son la utilización de WSMO (Web Services Modeling Ontology)³ o WSDL-S (Web Service Semantics)⁴. Por un lado, WSMO es un modelo conceptual para describir semánticamente todos los aspectos relevantes de los servicios Web con el objetivo de facilitar la automatización de su descubrimiento, invocación y composición. Para definir este modelo conceptual se utiliza como lenguaje el WSML (Web Services Modeling Language). Por otro lado, WSDL-S define un mecanismo para asociar anotaciones semánticas a servicios Web previamente descritos mediante WSDL.

4.3 *Arquitectura típica de una IDE y el papel de los metadatos*

Como se ha mencionado ya anteriormente, las IDE son una solución para gestionar eficientemente la información geográfica (Nebert 2001, Béjar et al. 2009a). El principal objetivo de este tipo de infraestructuras es facilitar una explotación eficiente de IG a todos los agentes involucrados en el mercado, ya sea nivel local, nacional o global. Y para el desarrollo de estas infraestructuras, la gestión de metadatos describiendo recursos y servicios resulta esencial.

La figura 1 presenta la arquitectura de componentes de una IDE típica ajustada a las necesidades de una institución, por ejemplo, una institución encargada de la gestión de los recursos hidrológicos de una cuenca (Arqued-Esquía et al. 2001, Nogueras-Iso et al. 2004). En la figura se muestra también la vista operacional de un entorno internet/intranet, mostrando igualmente la relación con las IDE de otras instituciones (por ejemplo, otras autoridades a cargo de recursos hidrológicos o el ministerio de medio ambiente a nivel nacional). Como se puede observar, cada institución controla su IDE local y a su vez es un nodo que se integra dentro de una IDE de nivel superior. Aparte de las características

¹<http://www.w3.org/TR/owl-features/>

²<http://www.w3.org/Submission/OWL-S/>

³<http://www.wsmo.org/>

⁴<http://www.w3.org/Submission/WSDL-S/>

especiales de cada componente (descritas brevemente más adelante), la característica más importante de estas infraestructuras es que no son un sistema ad-hoc para la resolución de un problema concreto. Por el contrario, lo que se pretende es que los servicios de información geográfica sean compatibles e interoperables con los servicios disponibles en otros nodos IDE, creados por cualquier tipo de organización a nivel local, nacional o global. Respecto a este objetivo de interoperabilidad, se debería mencionar la existencia de dos principales iniciativas de estandarización que han surgido en los últimos años: el comité de ISO con responsabilidad en información geográfica y geomática (ISO/TC 211) y el Open Geospatial Consortium (OGC). Desde una perspectiva más abstracta, ISO/TC 211 publica estándares y especificaciones definiendo los aspectos relevantes de la descripción y manejo de información geográfica y sus servicios asociados. Y por otro lado, OGC desarrolla y produce especificaciones técnicas a nivel de implementación para interfaces de procesamiento geoespacial, algunas de ellas adoptadas más tarde como estándares de ISO/TC211. Es decir, ambas iniciativas colaboran en el proceso de definición de interfaces estándares para los servicios principales que se integran habitualmente dentro de una IDE, con especial énfasis en aquellos servicios accesibles vía Web.

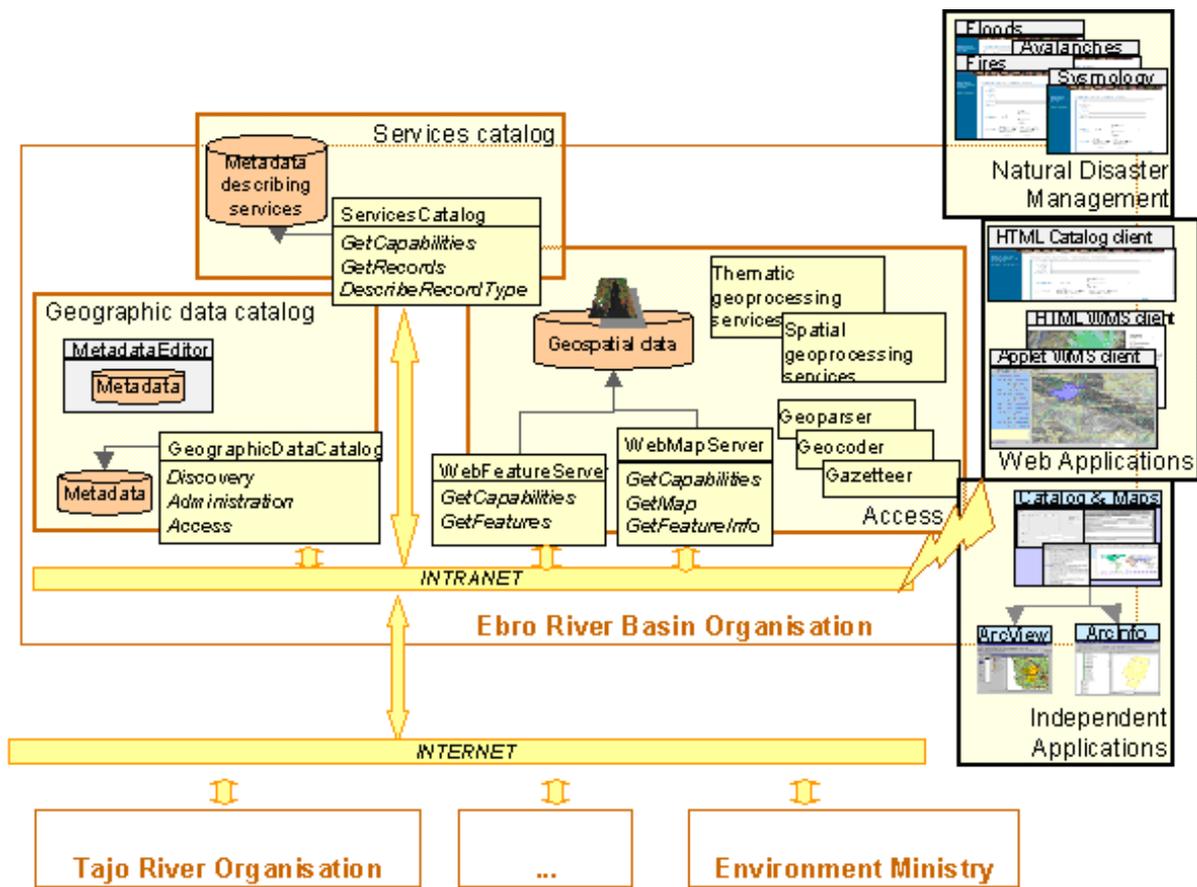


Figure 1. Arquitectura típica de una IDE.

Una de las partes más importantes de una IDE es el área correspondiente a los componentes de catálogo de datos geográficos. Los catálogos de datos geográficos son la solución para publicar descripciones (metadatos) de recursos de información geográfica de una forma estandarizada, facilitando que sus usuarios puedan localizar los datos de interés (Kottman Eds). Tal como se ha mencionado anteriormente, el uso de metadatos proporciona un vocabulario seleccionado contra el cual se pueden realizar consultas geoespaciales, facilitando así el entendimiento entre usuarios de la misma comunidad o de distintos dominios de aplicación. Esta área agrupa todos aquellos componentes necesarios para que los agentes involucrados creen y publiquen metadatos.

Una de las especificaciones líderes para la interoperabilidad entre catálogos es la especificación de Servicios de Catálogo de OGC (Nebert ed., Nogueras-Iso et al. 2005), que describe el conjunto de interfaces que soportan la Gestión (Management), el Descubrimiento (Discovery), y el Acceso (Access) a la información geográfica. En primer lugar los servicios de Descubrimiento permiten que los usuarios busquen dentro del catálogo usando un lenguaje de consulta con una sintaxis reconocida. Para la especificación de esas consultas se propone un lenguaje denominado OGC Common Query Language, muy similar al utilizado en las cláusulas WHERE de SQL. Además, se propone la codificación de este lenguaje sobre XML utilizando la especificación Filter Encoding Specification (Vretanos Eds). El uso de XML facilita la creación de interfaces de usuario orientadas a la creación de restricciones. En segundo lugar, los servicios de Gestión proporcionan la funcionalidad de gestión y organización de las entradas del catálogo en el dispositivo de almacenamiento local (ej., sistemas de ficheros o bases de datos relacionales). Y en tercer lugar, el rol de los servicios de Acceso es facilitar el acceso a los elementos que se habían previamente localizado con los servicios de Descubrimiento. Finalmente, otro aspecto importante en esta especificación es que se proporcionan diferentes perfiles de implementación de las interfaces de acuerdo a la plataforma y protocolo de distribución que se va a utilizar. En particular, esta especificación de OGC define perfiles de implementación para: CORBA (Orfali et al. 1999); el protocolo Z39.50 (protocolo muy utilizado en el ámbito de las bibliotecas digitales) (American National Standards Institute (ANSI) 1995) en su versión clásica con codificación BER (International Organization for Standardization (ISO) 1990) sobre TCP/IP, y en su reciente versión Web denominada Search/Retrieve Web Service (SRW/SRU); y un protocolo definido por OGC HTTP, también conocido como Catalog Services for the Web (CSW). En la implementación de estos servicios de catálogo es vital el uso de algoritmos y estrategias que permitan analizar los consultas y valorar los resultados de las mismas (ej., en base a frecuencia de aparición de términos, importancia de los elementos donde aparecen esos términos, importancia intrínseca del metadatos en base a estadísticas de acceso, ...) con el fin de ofrecer al cliente en primer lugar los resultados más relevantes de todos los encontrados.

Otra área importante en el desarrollo de las IDE es la correspondiente a los componentes de acceso a datos geoespaciales. Una vez que se han localizado los datos de interés, es necesario visualizar y evaluar los datos. Entonces, si estos datos son los deseados, los usuarios avanzados requerirán el acceso a los datos en su formato original. El componente Web Map Server (WMS) es un servidor de mapas que ofrece vistas gráficas (o mapas) de los datos de información geográfica a través de interfaces de servicio accesible en línea (Beaujardière Eds, Fernández et al. 2000). Con este tipo de componentes es posible evaluar y satisfacer gran parte de las necesidades de los usuarios sin requerir la descarga completa de los datos. Pero si este acceso final fuese necesario, se podría hacer uso de los componentes denominados Web Feature Server (WFS) (Vretanos Eds) y Web Coverage Server (WCS) (Evans 2003). El WFS es un servidor que permite acceso a fenómenos discretos (features) en formato GML (Geography Markup Language) (OpenGIS Consortium (OGC) Inc 2001). Por el contrario, el WCS es un servidor que permite el acceso a fenómenos continuos o coberturas en formatos raster. En la implementación de estos componentes es vital utilizar tecnología capaz de manejar grandes volúmenes de datos (superiores a 10 GB). Uno de los problemas con los que se topa al operar con datos geográficos es el gran tamaño de los mismos (hay que tener en cuenta que, por ejemplo, la fotografía aérea de todo Aragón es un único dato geográfico cuyo volumen puede exceder de los 1.000 CDs). Es necesario desarrollar tecnología que sea capaz de manejar estos volúmenes de información continua de manera eficiente y eficaz. Entran aquí en juego aspectos relativos al particionado transparente de la información, optimización de algoritmos de compresión, y recuperación y concatenación de los datos con el fin de poderlos visualizar desde cualquier aplicación tanto local, como vía Web.

Adicionalmente, esta área también integra componentes que proporcionan servicios para el geoprocesamiento y que están actualmente en proceso de estandarización. Algunos ejemplos de estos servicios son los que se encargarían de: procesamiento espacial para transformación de coordenadas, conversión de formatos y combinación de distintitos recursos geoespaciales; procesamiento temático para cálculo de parámetros, clasificación temática o muestreo basado en valores temáticos; procesamiento temporal para muestreo basado en valores temporales; y servicios de asociación de referencias geográficas como Geolocalizadores (Gazetteers) (Atkinson and Fitzke 2002), Geocodificadores (Geocoders) (Margoules 2001) o

Geoanalizadores (Geoparsers).

Otro componente vital para enlazar las IDE es el Catálogo de Servicios (Nogueras-Iso et al. 2009). Para que los servicios ofrecidos por una IDE sean accesibles al público general o a otras IDE, es necesario mantener un registro de los servicios disponibles. El componente de Catálogo de Servicios ofrece este directorio donde se publican los distintos servicios (WMS, WFS, WCS, catálogos,...) accesibles a través de la red. Por ejemplo, este componente puede monitorizar los catálogos de datos conectados a la IDE y puede servir a un portal que redirige consultas a los diferentes nodos de la IDE. Respecto a la especificación a utilizar para este catálogo de servicios, OGC propone la misma que para los catálogos de datos geográficos. Tal como se menciona en (Nebert ed.), un catálogo de datos geográficos se puede extender para almacenar y mantener metadatos de servicios.

Una conclusión que se puede extraer de los servicios y componentes descritos anteriormente es de que todos ellos dependen en mayor o menor medida del uso de metadatos, los cuales o bien se usan para describir la funcionalidad del servicio o constituyen los propios datos manejados por el servicio. Mientras que un catálogo de datos geográficos es un ejemplo de servicio que maneja directamente como datos los propios metadatos, la funcionalidad y el contenido ofrecido por servicios WMS o WFS se describe a través de un fichero de *capabilities*, que se puede considerar como los metadatos del servicio. Es decir, los metadatos se usan para enlazar todos los elementos de una IDE.

Finalmente, otro aspecto importante en el desarrollo de una IDE es el concerniente a las aplicaciones clientes de la infraestructura. La parte derecha de la figura 1 muestra distintos tipos de clientes que pueden beneficiarse de los servicios ofrecidos por la infraestructura: clientes genéricos y conformes a las especificaciones estándar establecidas por OGC; portales que proporcionan una puerta de entrada genérica a los servicios ofrecidos por una IDE denominados habitualmente Geoportales; o aplicaciones clientes desarrolladas de forma específica para un contexto particular (Riesgos naturales, Aplicaciones forestales, Medio ambiente, Economía, Climatología / Meteorología, Redes de suministro, Logística, ...). La integración de sistemas de información tradicionales con servidores de mapas, features, coberturas en Web, geolocalización (gazetteers), geocodificación, geoanálisis, geoprocésamiento y el encadenamiento de dichos servicios abre un panorama totalmente nuevo de posibilidades (Friis-Christensen et al. 2006, Cantán et al. 2001b,a).

4.4 Recursos web

Como consecuencia inmediata de la puesta en marcha de INSPIRE lo que se va a conseguir es un notable incremento del número de informaciones y recursos geográficos que pasarán a estar disponibles en la Web. Para dar una idea de los volúmenes de información y servicios que se pueden estar planteando, cabe mencionar que según la Asociación Europea para la Información Geográfica (EUROGI, <http://www.eurogi.org>) el 80 por ciento de toda la información almacenada en soporte electrónico por las Administraciones Públicas está relacionada con alguna localización geográfica (información georreferenciada) o es susceptible de estarlo. Se trata de grandes volúmenes de datos que resulta inviable transferir completos a los ordenadores personales encargados de trabajar con ellos de manera local (una imagen aérea de Aragón en color a resolución de medio metro por píxel ocupa del orden de 700 GB). El volumen y heterogeneidad de contenidos de esta información crece día a día gracias a los nuevos avances tecnológicos que se van produciendo en las técnicas de captura de la misma mediante sensores remotos de alta resolución ubicados en satélites. Adicionalmente, la información contenida en estos datos mantiene su vigencia durante muchos años. Hace más de 35 años que el hombre llegó a la luna y todavía se sigue utilizando información capturada entonces. Es razonable pensar que la información que se está recopilando en estos momentos es susceptible de ser utilizada al menos durante los próximos 50 años. Esto crea la necesidad de dotarse de servicios de publicación y acceso a estas informaciones que sean robustos y que puedan evolucionar tecnológicamente de manera viable. Junto con todo esto, la tecnología existente en estos momentos posibilita el uso de la información que se está capturando en tiempo real, siempre que se sepa dónde está y cómo acceder a ella. Todos estos volúmenes de información estarán disponibles en la Web superficial y, principalmente, en la Web profunda (Deep Web) Bergman (2001), también conocida como Web oculta (Hidden Web) Raghavan and Garcia-Molina (2001).

El tamaño de la Web superficial (Surface Web) o Web trivialmente indexable ha crecido de 200 millones de páginas en 1997, a 11.000 millones de páginas en 2005 Gulli (2005). El crecimiento exponencial de la información continúa en nuestros días, y no se prevé todavía su estabilización. La Web superficial está indexada por los denominados "motores de búsqueda comerciales. Estos asumen que Internet es una Web de Documentos enlazados. Inicialmente estos motores de búsqueda solo buscaban documentos HTML. Actualmente buscan una gran variedad de tipos de documento Web (p.e. imágenes, documentos PDF). Todo lo que está fuera del alcance de los motores de búsqueda se denomina convencionalmente Web Profunda. El tamaño estimado de la Web profunda es 550 veces el de la Web superficial Bergman (2001). La información puede no ser accesible por un mal diseño del portal, por restricciones a los motores de búsqueda, o por generarse como resultado a consultas a bases de datos. Estas últimas reciben el nombre de bases de datos profundas (Deep Web Databases). Se está operando un cambio de paradigma desde la Web de Documentos hacia la Web de Documentos y Datos y finalmente la Web de Información. Este cambio de paradigma es el impulsor del creciente interés por parte de investigadores y compañías comerciales en la indexación de los contenidos de las bases de datos profundas para que puedan ser localizados mediante los motores de búsquedas de la Web. Si en la Web de Documentos la extracción de información se basa en la emulación de la navegación, en la Web de Documentos y Datos la extracción de información implica descubrir los interfaces de las bases de datos profundas, qué términos aplicar en las preguntas, y cómo maximizar la cobertura del contenido de las bases de datos profundas con el menor número de preguntas Ntoulas (2005), Castillo (2006), Madhavan (2008).

La Deep Web no es una amalgama uniforme de información. Hay temáticas sectoriales muy diferenciadas en su naturaleza y contenido (p.e. medicina, turismo). Dentro de los diferentes tipos de la Deep Web hay que destacar la denominada Deep Geo Web. Podemos definir la Deep Geo Web como la parte de la Deep Web que está formada por bases de datos que publican información estructurada con contenido georreferenciado utilizando interfaces especializadas en la temática geoespacial (p.e. los interfaces OGC). Esta se caracteriza por la ingente cantidad y variedad de los tipos de información contenidos, así como en el elevado grado de heterogeneidad estructural y semántica de los esquemas de datos. El interés en la Deep Geo Web es doble. Por un lado su tamaño. Posiblemente la mayor parte de la información que contiene la Deep Web es información de temática geoespacial. Se pueden realizar estimaciones que indican que se trata de un área de trabajo que genera elevadas expectativas. En 2006 el tamaño de las bases de datos geográficas de los portales Geo Web desarrollados por los dos motores de búsqueda comerciales más populares (Google Earth de Google, y TerraServer de Microsoft Live) era de 70.5 terabytes y 16.5 terabytes, respectivamente. Estas bases de datos contienen una pequeña parte de la información disponible en las bases de datos científicas con información geográfica y que son accesibles. Por ejemplo, el Sistema de Observación e Información de la Tierra de la NASA (Earth Observing and Information System, EOSDIS) almacena un volumen de 4.9 petabytes de datos y produce 1 petabyte de información georreferenciada al año, toda ella fuera del alcance de los motores de búsqueda citados. Finalmente, hay que tener en cuenta su papel transversal. La Deep Geo Web contiene la información requerida para poder situar cualquier tipo de recurso de cualquier temática en una posición en la Tierra en un momento temporal determinado. Actualmente se están desarrollando numerosas iniciativas que están explotando la componente espacial de la información en Internet y que descubren que no pueden encontrar mediante medios convencionales la información espacial que sus proyectos requieren.

De modo complementario a lo anterior, la base sobre la que se están desarrollando la mayor parte de esfuerzos esta fundamentada en el trabajo sobre la semántica de los recursos en la Web. La Web semántica es una extensión de la actual Web en la que a la información disponible se le otorga una semántica bien definida Hendler and Miller (2002). Está basada en la idea de proporcionar en la Web datos definidos y enlazados, permitiendo que aplicaciones heterogéneas descubran, integren, razonen, y utilicen la información presente en la Web. Una ramificación de la Web semántica la constituye la Web GeoSemántica (también conocida como GeoSematic Web o Semantic Geospatial Web Egenhofer (2002)). En este caso se trata de poner el empeño fundamental de la caracterización semántica de los recursos en los aspectos geográficos. No solamente se trata de poder vincular recursos a unas coordenadas geográficas, sino de hacer uso de sistemas con carga semántica más avanzados como los esquemas de referencia espacial de diferente naturaleza, por ejemplo topónimos, callejeros, etc.

4.5 *Sistemas de Referencia Espaciales y Bases de Conocimientos Geoespaciales*

En general, la información puede contener indicación relativa a su localización espacial en forma de dato o de texto. Las referencias espaciales se pueden clasificar de dos formas:

- Las referencias expresadas mediante coordenadas. Estas coordenadas están referenciadas respecto a una referencia sobre el planeta Tierra. Así tenemos coordenadas Geográficas, Lambert, etc.
- Las referencias expresadas mediante identificadores geográficos. Estas no se basan en coordenadas explícitas; se basan en una relación entre el identificador geográfico y una entidad en el mundo real. La relación entre la referencia y la posición que se quiere expresar puede ser, por ejemplo, de contención, relativa o aproximada.

Un Sistema de Referencia Espacial (Spatial Reference System o SRS) es una conceptualización compartida que especifica cómo se define una referencia espacial expresada mediante coordenadas o mediante identificadores geográficos, su traslación a una referencia sobre el planeta Tierra, cómo extenderlo para aplicaciones concretas, cómo realizar transformaciones entre sistemas de referencias, su almacenamiento y su interpretación. Por ejemplo, el estándar ISO 19111:2007 define sistemas de referencia basados en coordenadas. También define cómo realizar conversiones entre estos sistemas de referencia. Por otra parte, el estándar ISO 19112:2003 define sistemas de referencia espaciales usando identificadores geográficos y su almacenamiento en nomencladores. El problema de la conversión entre sistemas de referencia utilizando identificadores geográficos no está todavía resuelto aunque ya existen propuestas para su estandarización Plews (2008).

El término Base de Conocimientos Geoespacial (Geospatial Knowledge Base o GKB) se refiere a:

- un conjunto de sentencias en un lenguaje formal Franconi (2002) aplicables a cualquier cosa situada en el espacio Pittman (2006), y
- la porción del conocimiento de una organización sobre información geoespacial almacenada en dispositivos de almacenamiento persistentes y en la memoria no persistente de los computadores de dicha organización.

En el primer sentido, el contenido de una GKB está limitado por los objetivos de ésta. Por ejemplo, Martins (2005b) define una GKB especializada en representar conceptualizaciones básicas sobre entidades geográficas (nombre, tipo, posición, relaciones) para el soporte de aplicaciones de minería de textos y procesamiento de lenguaje natural. Sin embargo, una base de conocimientos madura de amplio espectro de aplicación como Cyc Lenat (1989) incluye una potente parte espacial Pittman (2006) especializada en representar conceptos complejos como:

- Mereotopología, que incluye conceptos como parte-de, solape y conexión en sentido espacial.
- Topografía, que provee conceptos como métrica y geometría aplicados a la descripción de la superficie terrestre.
- Cartografía, que permite definir los límites naturales o sociales de regiones terrestres (p.e. ríos, cordilleras, ciudades, estados) y caracterizar estas (p.e. climatología, población, acidez del suelo).

Ideas similares se pueden encontrar en Bishr (2008). En ambos casos serían conocimientos geoespaciales de carácter general. Para su aplicación en un dominio concreto se necesita la formalización de la descripción de datasets espaciales, de la minería de datos espaciales y de los servicios geoespaciales Zhao (2006). En el segundo sentido, una GKB incluye componentes que ayudan al modelado, búsqueda, trazabilidad de la información, procesos de Extracción, Transformación y Carga, y al establecimiento de políticas de acceso.

Los procesos de Extracción, Transformación y Carga (Extract, Transform, and Load; o ETL) en los contextos de gestión de bases de datos, de data warehouses y de integración con sistemas legados, implican:

- La extracción de datos de fuentes externas. Una parte necesaria de este proceso es el análisis de los datos extraídos comprobando si estos cumplen un patrón o estructura esperado.
- La transformación para que se ajusten a los requisitos operacionales. Esta etapa aplica una serie de reglas y funciones a los datos extraídos para derivar los datos cargados en el destino.
- La carga en el sistema destino. La carga puede ser destructiva, de actualización o acumulativa. Durante

esta fase se interactúa con el sistema destino aplicando estas restricciones que tenga definidas.

Los ETL son procesos costosos en recursos y tiempo consumidos. Vassiliadis (2002) cita fuentes profesionales que estiman que los costes económicos entre el 33 y el 55 por ciento mientras que el tiempo consumido se sitúa en el 80 por ciento. En la revisión del estado de la investigación en el campo de los datawarehouses Rizzi (2006) señalan diferentes líneas de investigación abiertas en el campo de las ETL: su integración en el modelado conceptual, el uso de metamodelos como CWM para su descripción, el diseño de metodologías para la propagación de cambios en la estructura de las fuentes, la integración con los sistemas de seguridad y autenticación de las fuentes y destinos y la evolución de sistemas batch hacia sistemas ETL en tiempo real. Se ha analizado los ETL desde los puntos de vista funcional Vassiliadis (2002), dinámico Bouzeghoub (1999) y semántico Calvanese (1998), pero Rizzi (2006) afirma que todavía estas investigaciones no han alcanzado la madurez de otras áreas relacionadas con las bases de datos.

4.6 Georreferenciación semántica

La georreferenciación, entendida como la transformación de una descripción de localización (p.e. "Zaragoza") comprensible por el usuario en una descripción comprensible por una máquina, la cual se resuelve en una posición sobre el mundo físico (p.e. "41.656,-0.877", "población", "capital de provincia", "50287"), hace la información más clara y posibilita un uso compartido en la Web GeoSemántica. Por ejemplo, la descripción de la localización de un elemento de patrimonio cultural (por ejemplo un molino) se puede apoyar en diferentes nombres (p.e. su nombre oficial) y por diferentes identificadores (p.e. el código catastral de la parcela donde se sitúa, el código INE del municipio), así como en elementos subjetivos que incluyen la percepción humana (p.e. la descripción de su localización por un viajero), las convenciones sociales (p.e. las diferentes denominaciones del lugar a lo largo del tiempo), y combinaciones de medidas objetivas (p.e. a 5 km al norte) con juicios subjetivos (p.e. a 5 min de un camino abandonado de Zaragoza). Cada uno de estos nombres e identificadores son la puerta de acceso a diferentes sistemas de información permitiendo publicar, compartir y acceder a información de comunidades de usuarios heterogéneas (p.e. una búsqueda distribuida a través de una red formada por servicios de distintas administraciones basada en las diferentes formas de describir la localización). La georreferenciación permite no solo compartir información hacia al exterior, sino, lo que es más relevante, integrar y complementar la información existente con información georreferenciada procedente de otros servicios (p.e. el uso actual del entorno del molino, el propietario actual, el nombre actualizado del municipio, el responsable de patrimonio de la comarca).

Las propuestas de georreferenciación existentes se basan en la aplicación de técnicas estadísticas y el uso de grandes colecciones de vocabulario geográfico complementadas con geo-ontologías generales Martins (2005b). Algunos de las aproximaciones más relevantes son:

- Asignación del ámbito geográfico basado en ranking de grafos. El ámbito geográfico de un documento puede ser establecido mediante la aplicación de algoritmos de ordenación de grafos desarrollados. Martins (2005a) propone el uso de los algoritmos PageRank Brin (1998) y HITS Buckland (2004), Kleinberg (1999) adaptados para grafos ponderados para esta tarea. Propone para cada documento un grafo ponderado formado por las referencias geográficas, una ontología geográfica y las asociaciones de las referencias geográficas y los conceptos en dicha ontología. relaciones entre ambas. Aplicado un algoritmo de ordenación de grafos las referencias geográficas con mayor peso serían el ámbito del documento.
- Similitud semántica. La similitud entre dos conceptos geográficos puede calcularse, por ejemplo, mediante la distancia topológica entre conceptos, la distancia espacial y la población que comparten Martins (2008). Una variante del análisis de similitud entre conceptos se basa en la comparación con el contenido informativo del antecesor común. Cuando un mismo concepto puede pertenecer a diferentes clasificaciones se ha planteado proporcionan interpretaciones distintas. Un ejemplo de estas medidas es GraSM Couto (2005) que se aplica a las ontologías biomédicas. Esta misma idea podría aplicarse a ontologías geográficas.
- Clasificación de documentos. La clasificación de documentos puede hacerse mediante la fórmula Okapi BM25 Robertson (1994) (o variaciones de esta aplicadas a documentos estructurados Robertson (2004)) que sigue representando el estado del arte en este campo. En el contexto de la información geográfica

los resultados pueden ser organizados en un cluster de dos niveles, representando primero la agrupación temática (basada en el texto contenido) y después la agrupación espacial (basada en el ámbito espacial del contenido) Martins (2008). Esta clusterización puede aprovecharse de algoritmos como el propuesto en Sanderson (2000) que se basan en la idea de inclusión estadística para simplificar el número de clusters resultantes.

Dentro de los resultados destacados de otros grupos de investigación en las líneas que se abren en estos trabajos, cabe destacar el trabajo del grupo XLDB (liderado por Mário J. Silva) en LaSIGE (Departamento de Informática de la Facultad de Ciencias de la Universidad de Lisboa) en los proyectos GREASE y GREASE-II. Otros trabajos destacados corresponden al grupo de investigación en geoinformática y gestión de datos espaciales de la Facultad de Ciencias de la Computación de la Universidad de Cardiff, liderados por Christopher Jones y Alia Abdelmoty, cuyos resultados se asocian en los proyectos SPIRIT y TRIPOD. También son una referencia en este campo el grupo MUSIL (liderado por Werner Kuhn) en el IFGI (Universidad de Münster) con su participación en proyectos europeos como SWING, y el grupo OntoGeo (liderado por Marinos Kavouras) de la Universidad de Atenas que con proyectos como GeoNLP esta integrando análisis de lenguaje natural y geo-ontologías.

4.7 *Geoparsing, Geocoding y Reverse Geocoding*

Geoparsing se entiende como el proceso de identificación en los documentos de descripciones de localización conformes a un SRS basado en identificadores geográficos, u otro tipo de descripciones de lugares (p.e. direcciones). El Geoparsing de los contenidos, como por ejemplo páginas Web Zong (2005), permite añadir la semántica por anotación de documentos con identificadores geográficos procedentes de un SRS. La tarea de Geoparsing se ha automatizado recientemente obteniendo resultados comparables con un procesamiento manual Zheng (2002). Los avances más relevantes en el campo del Geoparsing provienen de las investigaciones en el campo del Procesamiento del Lenguaje Natural. Por ejemplo, existen workshops especializados en el análisis de las referencias geográficas en las conferencias de este campo donde se presentan avances en el Geoparsing Densham (2003), Southall (2003), Bilhaut (2003), Uryupina (2003), Waldinger (2003).

Geocoding es el nombre que recibe el proceso de asignación de una geometría espacial (por ejemplo, latitud y longitud) a una dirección o cualquier otro tipo de descripción de lugar Shekhar (2004). Este proceso puede utilizar como base de conocimiento uno o varios SRS basados en identificadores. Este acto de convertir una descripción en una referencia geográfica se ha convertido en una pieza crítica en cualquier proceso que implique la explotación de información textual Goldberg (2007). Un sistema de Geocoding robusto y global es un reto al no existir una forma única u homogénea de describir lugares Goldberg (2008). En particular, los servicios de Geocoding en línea deben tratar con preguntas incompletas, con errores ortográficos, con inversión de términos y variaciones similares aplicando técnicas de Geoparsing Sengar (2007), Joshi (2008).

Reverse Geocoding es la operación opuesta a Geocoding: dada una localización espacial asigna un identificador geográfico, una dirección o una descripción de lugar Shekhar (2004). De forma similar al Geocoding, este proceso puede utilizar como base de conocimiento uno o varios SRS basados en identificadores. Los servicios de Reverse Geocoding en línea deben tratar con problemas derivados de la precisión de la localización (p.e. derivados de la precisión del GPS) para asignar la descripción más adecuada. Para esta tarea se utilizan técnicas geométricas Bernstein (1996), probabilísticas Pyo (2001), basadas en lógica difusa Syed (2004) o, incluso, semánticas Liu (2006).

4.8 *Crawlers de Repositorios de Información (Datos/Servicios) Publicados*

Un crawler es un programa, comúnmente usado por motores de búsqueda, cuyo resultado de ejecución es una serie de URLs generalmente vinculadas a páginas Web Pinkerton (1994). Básicamente, un crawler recibe como entrada una URL inicial de la cual extrae cualquier otra referencia externa para añadirla a su lista de URLs a visitar indefinidamente hasta que se programa su parada.

4.8.1 *Crawlers superficiales (Surface Web Crawlers).* Tal y como se puede intuir, no todas las URLs devueltas tendrán el mismo interés, y por otro lado, el tiempo que cuesta realizar esta búsqueda deberá ser el menor posible. Para conseguir maximizar el número de URLs de máximo interés recibido y minimizar el tiempo de búsqueda se han definido a lo largo de la historia muchas técnicas, como por ejemplo la ordenación de las mismas Cho (1998). Debido a los grandes avances en la tecnología informática, y siempre buscando esa minimización de tiempo de búsqueda, es interesante conocer y tener en cuenta para este proyecto el concepto de crawlers paralelos Cho (2002). La arquitectura de un crawler paralelo se basa en la idea de múltiples procesos explorando la red. Cada proceso comienza la exploración con un conjunto de URLs iniciales propio. Muy posiblemente existirán sitios visitados y por lo tanto descargados por varios procesos, para evitar esto, todos los procesos deben mantener comunicación. Como es obvio, la información en Internet no es estática, es decir, se actualiza con asiduidad dependiendo del sitio Web. Por lo tanto, es necesario realizar visitas continuas para actualizar la información que se tiene de cada uno de los sitios Web. Esta problemática ha sido objetivo de estudio en numerosos trabajos como Ntoulas (2004), Cho (2003), o Tan (2007). La viabilidad de funcionamiento de un crawler reside en su diseño, un software de este tipo, con una faceta iterativa tan marcada puede convertirse en un programa capaz de derruir incluso las arquitecturas más robustas. Castillo (2006) presenta una estrategia que reduce al 50 por ciento el tamaño máximo de la cola de descarga de páginas Web mientras que se mantiene la calidad de los resultados obtenidos.

4.8.2 *Crawlers profundos (Deep Web Crawlers).* Extraer información de la Deep Web implica afrontar varios retos en el diseño del Deep Web Crawler (DWC) Ntoulas (2005), Madhavan (2008). Requiere una aproximación completamente automática, escalable y eficiente en un entorno donde los recursos están limitados. Debe también ser capaz de seleccionar los términos que van a ser utilizados para instanciar las preguntas considerando aspectos como el coste de la recuperación. También ha de construir las preguntas, tanto si hay un único campo en el formulario o varios. El algoritmo genérico de un DWC una vez encontrada la entrada a una base de datos profunda consiste primero en decidir por qué termino debe preguntar, después debe realizar la pregunta, recuperar la página con los resultados y seguir los enlaces a cada uno de los documentos para recuperarlos. Ntoulas (2005) describe una serie de factores críticos a tener en cuenta en el diseño: qué preguntar primero (solo hay metainformación disponible), qué preguntar a continuación (hay metainformación y los resultados de las preguntas anteriores), qué impacto tiene la limitación en el número de resultados devueltos y que consumo de recursos hace el DWC al examinar la base de datos profunda. El problema de la limitación de recursos por parte del DWC se trata en Ntoulas (2005), Castillo (2006) o Madhavan (2008). Los DWC pueden estar enfocados en tareas concretas. Cai (2008) describe un DWC especializado en recuperar información de foros. Este tipo de base de datos profunda está caracterizada por poseer un gran número de enlaces rotos, regiones enteras de texto duplicadas en diferentes páginas, y problemas de acceso. Duda (2008) describe un crawler capaz de recuperar información de aplicaciones AJAX/RIA. Yang (2008) propone por su parte crawlers que eligen su navegación utilizando el conocimiento establecido en una ontología. Finalmente, Al-Masri (2008) ha desarrollado un DWC denominado Web Service Crawler Engine capaz de obtener metainformación sobre Servicios Web buscando en registros UDDI, en portales Web y en otros buscadores. Convencionalmente se considera que los crawlers tienen como objetivo obtener una replicación al 100 por 100. Este no es un requisito de los DWC. Por ejemplo Madhavan (2008) plantea replicar solo parte de la base de datos oculta con mecanismos DWC. El objetivo es que los documentos recuperados sean el origen de tráfico hacia la base de datos profunda. Es decir, que sean puntos de partida para que un crawler normal explore el resto del contenido de la base de datos. Otro de los objetivos de los DWC es la recopilación de información para la construcción de metabuscadores. Un metabuscador identifica, en función de la pregunta del usuario, las bases de datos profundas más relevantes a las que dirigir la pregunta Liu (2004). En este caso, el DWC tiene como objetivo recopilar información útil para que el metabuscador pueda realizar esta tarea. Otra idea convencional es que los DWC no han de tener ayuda por parte de la base de datos oculta. Para hacer visible contenidos de la Deep Web, además de facilitar el acceso a los crawlers tradicionales y disminuir el ancho de banda que estos consumen, autores como Brandman (2000) han planteado la posibilidad de ofrecer meta-información en los servidores sobre su contenido. En particular, la identificación de los recur-

sos que han cambiado y la enumeración de los recursos disponibles en un único documento. Un ejemplo de aplicación se encuentra en Nelson (2006). Este trabajo plantea utilizar el estándar OAI-PMH para intercambio de información entre bibliotecas digitales para publicar la meta-información de un portal Web. Otra aplicación similar es el protocolo Google Sitemaps utilizado por Google para que los administradores de portales Web y bases de datos profundas notifiquen que documentos son relevantes a los crawlers de Google (2009).

4.9 Integración semántica de servicios

En general, la integración de servicios requiere que estén adecuadamente descritos. El estándar de W3C Web Services Description Language 1.1 (WSDL) permite describir la sintaxis de servicios Web genéricos. En el campo de los servicios geográficos la especificación OGC Web Services Common Specification 1.1.0:2007 indica cómo hacerlo para servicios OGC, aunque OGC está convergiendo hacia el uso de WSDL en sus especificaciones. Un campo de investigación actualmente en auge es el orientado a la descripción de servicios Web semánticamente, de manera que se pueda facilitar el procesado automático de las operaciones que ofrecen, y así ayudar a construir de manera más o menos automática encadenamientos o coreografías de servicios para resolver tareas complejas. Web Ontology Language (OWL) es una de las opciones para formalizar estas descripciones semánticas, y más concretamente OWL-S, que es una ontología de nivel superior basada en OWL que modela características de servicios Web, y que se utiliza para anotar las descripciones sintácticas de los servicios con información semántica. Otra opción común es WSDL-S. Lemmens (2006) presenta un ejemplo de aplicación de estas técnicas a una aplicación de gestión de riesgos basada en algunos servicios integrados en una IDE. Klien (2006) aplica ontologías al descubrimiento de servicios geográficos. Otra de las opciones existentes para la descripción de servicios es la WSMO (Web Services Modeling Ontology), un modelo conceptual para describir semánticamente todos los aspectos relevantes de los servicios Web con el objetivo de facilitar la automatización de su descubrimiento, invocación y composición. Para definir este modelo conceptual se utiliza como lenguaje el WSML (Web Services Modeling Language). Dentro del contexto de la información geográfica, esta descripción de servicios se ha utilizado en proyectos como SWING (Semantic Web Services Interoperability for Geospatial decision making).

Un tipo de servicios especialmente relevante en el ámbito de la información geoespacial lo constituyen los Servicios de Geoprocesamiento. Éstos podrían definirse como el motor que mueve todas las funcionalidades de las aplicaciones de usuario. Mientras que por aplicación de usuario se entiende el software que normalmente percibirá el usuario, es decir, las interfaces para las búsquedas y consultas, las herramientas para administrar contenidos de los almacenes de información y algunas aplicaciones de análisis. Estos servicios deben tener la capacidad de proporcionar la información requerida por sus clientes de manera adecuada. Para apoyar el desarrollo de este tipo de servicios, son necesarios también servicios genéricos de procesamiento de datos geográficos (Geoprocesamiento). La aproximación a este tipo de servicios desde el OpenGeospatial Consortium es la especificación del Servicios de Procesado en Web (Web Processing Service Consortium (2007)). La especificación de WPS proporciona un marco para la construcción de servicios de procesamiento accesibles por Internet. Su objetivo es facilitar la publicación y descubrimiento de servicios, aumentando su interoperabilidad tanto para otros servicios como para los clientes finales. Un WPS puede ser configurado para ofrecer un tipo de funcionalidad SIG a través de la red proporcionando acceso a cálculos preprogramados o modelos de computación que operan con datos referenciados geoespacialmente. Los datos requeridos por el WPS pueden ser obtenidos a través de la red o estar disponibles en el mismo servidor. El cálculo puede ser tan simple como la resta de valores entre variables concretas de los juegos de datos (p.e. la determinación de la diferencia en casos de gripe entre dos estaciones de diferentes), o tan complicado como un modelo de cambio de clima global Consortium (2007).

Los servicios OGC son independientes de los estándares propuestos por W3C por lo que existen incompatibilidades entre estos Weiser (2007). Ultimamente OGC ha ido acercándose a los estándares propuestos por W3C debido a que estos están siendo ampliamente adoptados. En la versión 1.0 de WPS ha incluido la recomendación y algunas pautas para ofrecer el servicio utilizando la descripción WSDL y el protocolo SOAP. El trabajo en la adaptación SOAP/WSDL para servicios OGC no es nuevo, aunque la mayor parte

de lo que se ha llevado a cabo presenta métodos de adaptación automáticos que deben ser comprobados con posterioridad. Moses Gone (2007) presenta un análisis del empleo de BPEL en comparación con Web Service Modeling Ontology (WSMO) y una revisión del estado del arte en este tipo de aproximaciones. Sancho-Jiménez et al. (2008) Ofrece un método que permite añadir cualquier interface opcional automáticamente a un WPS sin necesidad de acceder al código fuente ni a los binarios del servicio, tan solo requiere tener acceso a su interface obligatorio. Este método permite a los WPS que cumplan con los requisitos mínimos cumplir las especificaciones de W3C sobre SOAP/WSDL proporcionando una serie de características: encadenamiento de servicios, interoperabilidad, flexibilidad y escalabilidad Leite (2006)

5. Conclusiones

En este trabajo se ha hecho una revisión de diversos aspectos que nos parecen de interés para la investigación en temas relacionados con las infraestructuras de datos espaciales. Estos temas y otros muchos que están siendo abordados por otros colegas, ponen de manifiesto el punto álgido en el que se encuentra el progreso científico y tecnológico relacionado con las infraestructuras de datos espaciales. Todos estos aspectos son de interés para proyectos de investigación pero hay que resaltar también que es un área de trabajo en la que está resultando de vital importancia la transferencia de resultados de investigación y en la que se está produciendo una interesante sinergia entre los distintos sectores involucrados que ha conducido a un destacado posicionamiento nacional en el ámbito europeo de las infraestructuras de datos espaciales, lo que es más meritorio teniendo en cuenta la retrasada posición de partida.

6. Agradecimientos

Los trabajos referidos están parcialmente soportados por la Comisión Europea (proyectos 2008/S 139-185315, EuroGEOSS, EuroGeoSource y EurAdin), el Gobierno de España (proyectos TIN2007-65341, TIN2009-10971, PET2008-0026, TSI-020301-2009-13, TSI-020301-2009-14, “España Virtualref. CENIT 2008-1030 y programa Torres Quevedo ref. PTQ06-2-0790 y PTQ06-2-0798), el Gobierno de Aragón (proyecto PI075/08), el Instituto Geográfico Nacional (IGN) y GeoSpatiumLab S.L.

Referencias

- Al-Masri, E.; Mahmoud, Q. H. (2008). Investigating web services on the world wide web. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 795–804.
- American National Standards Institute (ANSI) (1995). Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection. ANSI/NISO Z39.50-1995 (ISO 23950). <http://lcweb.loc.gov/z3950/agency/document.html>.
- American National Standards Institute (ANSI) (2001). The Dublin Core Metadata Element Set. ANSI/NISO Z39.85-2001.
- Annoni, A.; Craglia, M. S. P. (2002). Comparative analysis of nsdi. In *8th EC-GI&GIS Workshop*.
- Arqued-Esquía, V. M., Zarazaga-Soria, F. J., and Losada-García, J. A. (2001). El sistema de información gis-ebro. metadatos y catálogo de datos geográficos. *BOLE.TIC*, 19:83–97.
- Atkinson, R. and Fitzke, J., editors (2002). *Gazetteer Service Profile of the Web Feature Service Implementation Specification, v0.9*. Open Geospatial Consortium Inc, OGC 02-076r3.
- Beaujardière(Eds), J. (2002). Web Map Server Implementation Specification. Version 1.1.1. OpenGIS project document 01-068r3, Open GIS Consortium Inc.
- Bergman, M. K. (2001). The deep web: Surfacing hidden value. *Journal of Electronic Publishing*.
- Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web. *Scientific American*.
- Bernstein, D.; Kornhauser, A. (1996). An introduction to map matching for personal navigation assistants. Technical report, New Jersey TIDE Center.

- Bilhaut, F.; Charnois, T. E. P. M. Y. (2003). Geographic reference analysis for geographic document querying. In *Proceedings of the HLT-NAACL, Workshop on Analysis of Geographic References*.
- Bishr, Y. A. (2008). *Encyclopedia of GIS*, chapter Semantic Web: Applications, pages 391–398. Springer.
- Béjar, R., Latre, M., Nogueras-Iso, J., Muro-Medrano, P., and Zarazaga-Soria, J. (2009a). An architectural style for Spatial Data Infrastructures. *International Journal of Geographical Information Science (IJGIS)*, 23(3):271–294.
- Béjar, R., Latre, M., Nogueras-Iso, J., Muro-Medrano, P., and Zarazaga-Soria, J. (2009b). Systems of Systems as a Conceptual Framework for Spatial Data Infrastructures. *International Journal of SDI Research (IJS DIR)*, 4:201—217.
- Béjar, R., Nogueras-Iso, J., Latre, M. ., Muro-Medrano, P., and Zarazaga-Soria, F. J. (2009c). *Handbook of Research on Digital Libraries: Design, Development, and Impact*, chapter Digital Libraries as a Foundation of Spatial Data Infrastructures. Chapter XXXIX, pages 382–389. IGI Global, Singapore. ISBN 978-1-59904-879-6.
- Boardman, J.; Sauser, B. (2006). System of systems - the meaning of of. In *2006 IEEE/SMC International Conference on System of Systems Engineering*.
- Booth, D., Haas, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C., and Orchard(eds), D. (2004). Web Services Architecture. W3C Working Group Note 11 February 2004, W3C. <http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/>.
- Borgman, C. (2003). The invisible library: Paradox of the global information infrastructure. *Library trends*, 51(4):652–674.
- Bouzeghoub, M.; Fabret, F. M.-B. M. (1999). Modeling the data warehouse refreshment process as a workflow application. In Stella Gatzui, Manfred A. Jeusfeld, M. S. and Vassiliou, Y., editors, *Proceedings of the Intl. Workshop on Design and Management of Data Warehouses, DMDW'99*, volume 19 of *CEUR Workshop Proceedings*.
- Brandman, O.; Cho, J. G.-M. H. S. N. (2000). Crawler-friendly web servers. *ACM SIGMETRICS Performance Evaluation Review*, 28(2):9–14.
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., and Maler, E. (2000). Extensible Markup Language (XML) 1.0 (Second Edition). W3C Recommendation 6 October 2000, W3C. <http://www.w3.org/TR/2000/REC-xml-20001006>.
- Brickley, D. and Guha, R. V. (2004). RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation 10 February 2004, W3C. <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>.
- Brin, S.; Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems (Proceedings of the Seventh International World Wide Web Conference)*, 30:107–117.
- Buckland, M.; Lancaster, L. (2004). Combining place, time and topic: The electronic cultural atlas initiative. *D-Lib Magazine*, 10(5).
- Buehler, K. and McKee(eds.), L. (1996). The Opengis Guide. Introduction to Interoperable Geoprocessing. Part I of the Open Geodata Interoperability Specification (OGIS). OGIS TC Document 96-001, OGIS Project 6 Technical Committee of the OpenGIS Consortium Inc.
- Cai, R.; Yang, J. L.-W. W. Y. Z. L. (2008). irobot: an intelligent crawler for web forums. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 447–456. ACM.
- Calvanese, D.; De Giacomo, G. L. M. N. D. R. R. (1998). Information integration: Conceptual modeling and reasoning support. In *COOPIS '98: Proceedings of the 3rd IFCIS International Conference on Cooperative Information Systems*, page 280–291. IEEE Computer Society.
- Cantán, O., Bañares, J. A., Gutierrez, J., Nogueras, J., and Zarazaga, F. J. (2001a). Merging catalog services and gis applications by component interoperability mechanisms. In *Proc. of the 7th European Commission GI&GIS Workshop, Managing the Mosaic*, Potsdam, Germany.
- Cantán, O., Casanovas, M., Gutierrez, J., Nogueras, J., and Zarazaga, F. J. (2001b). Joining geographic catalog services and map servers with gis applications. In *Proc. of the 4th AGILE Conference on Geographic Information Science. GI in EUROPE: Integrative, Interoperable, Interactive*, pages 347–354, Brno, Czech Republic.
- Cantán, O., Nogueras-Iso, J., and Zarazaga-Soria, F. (2009). *Handbook of Research on Digital Libraries:*

- Design, Development, and Impact*, chapter DL and GIS: Path to a new collaboration paradigm. Chapter XL, pages 390–399. IGI Global, Singapore. ISBN 978-1-59904-879-6.
- Castillo, C.; Nelli, A. P. A. (2006). A memory-efficient strategy for exploring the web. In *WI '06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 680–686. IEEE Computer Society.
- Cho, J.; Garcia-Molina, H. (2002). Parallel crawlers. In *WWW '02: Proceedings of the 11th international conference on World Wide Web*, pages 124–135.
- Cho, J.; Garcia-Molina, H. (2003). Effective page refresh policies for web crawlers. *ACM Transactions on Database Systems*, 28:390–426.
- Cho, J.; Garcia-Molina, H. P. L. (1998). Efficient crawling through url ordering. *Computer Networks and ISDN Systems (Proceedings of the Seventh International World Wide Web Conference)*, 30:161–172.
- Christensen, E., Curbera, F., Meredith, G., and Weerawarana, S. (2001). Web Services Description Language (WSDL) 1.1. W3C Note 15 March 2001, W3C. <http://www.w3.org/TR/2001/NOTE-wsdl-20010315>.
- Coleman, D. J. and Nebert, D. D. (1998). Building a North American Spatial Data Infrastructure. *Cartography and Geographic Information Systems*, 25(3):151–160.
- Coleman, D. J.; McLaughlin, J. (1997). Defining global geospatial data infrastructure (ggdi): Components, stakeholders and interfaces. In *2nd Global Spatial Data Infrastructure Conference*.
- Comité Técnico de Normalización 148 de AENOR(AEN/CTN 148) (1998). UNE 148001 EXP: 1998, Mecanismo de Intercambio de Información Geográfica Relacional formado por Agregación (MIGRA), versión 1.
- Consortium, O. G. (2007). Opengis® web processing service. ogc 05-007r7 version: 1.0.0. 2007.
- Couto, F.M.; Silva, M. C. P. (2005). Semantic similarity over the gene ontology: family correlation and selecting disjunctive ancestors. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, page 343–344.
- Craglia, M., Annoni, A., and Masser, I., editors (1999). *Geographic Information Policies in Europe: National and Regional Perspectives. EUROGI-EC Data Policy Workshop, Amersfoort 15th November 1999*. European Commission - Space Applications Institute, European Communities. <http://www.ec-gis.org/reports/policies.pdf>.
- DeLaurentis, D.; Callaway, R. (2004). A systems-of-systems perspective for public policy decisions. *Review of Policy Research*, 21(6):829–837.
- Densham, I.; Reid, J. (2003). A geo-coding service encompassing a geo-parsing tool and integrated digital gazetteer service. In *Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References*.
- Dublin Core Metadata Initiative (DCMI) (2004). Homepage of the Dublin Core Metadata Initiative. <http://www.dublincore.org>.
- Duda, C.; Frey, G. K. D. Z. C. (2008). Ajaxsearch: crawling, indexing and searching web 2.0 applications. In *Proceedings of the VLDB Endowment archive*, pages 1440–1443.
- Egenhofer, M. J. (2002). Toward the semantic geospatial web. In *Proceedings of the 10th ACM international Symposium on Advances in Geographic Information Systems*, pages 1–4.
- European Commission (1998). Public Sector Information: A Key Resource for Europe. Green Paper On Public Sector Information In The Information Society. COM 98/0585.
- European Commission (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>.
- European Commission (2008). Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata (Text with EEA relevance). <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:326:0012:0030:EN:PDF>.
- European Committee for Standardization(CEN) - CEN/TC 287 (1998). Geographic Information European Prestandards, Euro-norme Voluntaire for Geographic Information - Data description - Metadata. ENV

12657.

- Evans, J., editor (2003). *Web Coverage Service (WCS), v1.0*. Open Geospatial Consortium Inc, OGC 03-065r6.
- Federal Geographic Data Committee (FGDC) (1998). Content Standard for Digital Geospatial Metadata, version 2.0. Document FGDC-STD-001-1998, Metadata Ad Hoc Working Group.
- Federal Geographic Data Committee (FGDC) (2000). *Content Standard for Digital Geospatial Metadata Workbook, version 2.0*. Metadata Ad Hoc Working Group.
- Fernández, P., Béjar, R., Latre, M. ., Valiño, J., Bañares, J. A., and Muro-Medrano, P. R. (2000). Web mapping interoperability in practice, a java approach guided by the opengis web map server interface specification. In *Proc. of the 6th European Commission GI&GIS Workshop, The Spatial Information Society - Shaping the Future*, Lyon, France.
- Fisher, D. (2006). An emergent perspective on interoperation in systems of systems. integration of software-intensive systems. technical report cmu/sei-2006-tr-003. Technical report, Carnegie Mellon University. Software Engineering Institute.
- Franconi, E. (2002). Description logics.
- Friis-Christensen, A., Bernard, L., Kanellopoulos, I., Noguera-Iso, J., Peedell, S., Schade, S., and Thorne, C. (2006). Building service oriented applications on top of a spatial data infrastructure - a forest fire assessment example. In *Proceedings of 9th AGILE International Conference on Geographic Information Science: Shaping the future of Geographic Information Science in Europe*, pages 119–127, Visegrád, Hungary.
- Georgiadou, Y. (2006). Sdi ontology and implications for research in the developing world. *International Journal of Spatial Data Infrastructures Research*, 1:51–64.
- Goldberg, D. W. (2008). *Geocoding in Cancer Surveillance: A Handbook of Best Practices*. The North American Association of Central Cancer Registries.
- Goldberg, D. W.; Wilson, J. P. K. C. A. (2007). From text to geographic coordinates: The current state of geocoding. *URISA Journal*, 19:33–46.
- Google (2009). Using the sitemap protocol.
- Groot, R. and McLaughlin, J. (2000). *Geospatial Data Infrastructure: concepts, cases and good practice*. Oxford University Press, New York, USA.
- Gulli, A. ; Signorini, A. (2005). The indexable web is more than 11.5 billion pages. In *WWW '05: Special interest tracks and posters of the 14th international conference on World Wide Web*, pages 902–903.
- Haas, H. and Brown, A. (2004). Web Services Glossary. W3C Working Group Note 11 February, W3C. <http://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/>.
- Hanseth, O.; Monteiro, E. (1998). *Understanding Information Infrastructure*.
- Hendler, J., B.-L. T. and Miller, E. (2002). Integrating applications on the semantic web. *Journal of the Institute of Electrical Engineers of Japan*, 122(10):676–680.
- International Organization for Standardization (ISO) (1990). Information technology - Open Systems Interconnection - Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN.1). ISO/IEC 8825:1990.
- International Organization for Standardization (ISO) (2003a). Geographic information - Metadata. ISO 19115:2003.
- International Organization for Standardization (ISO) (2003b). Information and documentation - The Dublin Core metadata element set. ISO 15836:2003.
- ISO (2003). *Geographic Information - Services*. International standard; ISO 19119. International Organization for Standardization, Genève, Switzerland.
- Joshi, T.; Joy, J. K. T. K. U. K. A. S. V. (2008). Crosslingual location search. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*.
- Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632.
- Klien, E.; Lutz, M. K. W. (2006). Ontology-based discovery of geographic information services—an application in disaster management. *Computers, Environment and Urban Systems*, 30(1):102–123.

- Kottman(Eds), C. (1999). The OpenGIS Abstract Specification. Topic13: Catalog Services (version 4). OpenGIS Project Document 99-113, OpenGIS Consortium Inc.
- Leite, F. L.; de Souza Baptista, C. d. A. S. P. R. d. S. E. (2006). *Advances in Geoinformatics. VIII Brazilian Symposium on GeoInformatics, GEOINFO 2006*, chapter WS-GIS: Towards a SOA-Based SDI Federation, pages 247–265. Springer Berlin Heidelberg.
- Lemmens, R. (2006). Integrating semantic and syntactic descriptions to chain geographic services. *IEEE Internet Computing*, 10(5):42–52.
- Lenat, D.B; Guha, R. V. (1989). *Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project*. Addison-Wesley Longman Publishing Co.
- Lieberman, J. (2003). OpenGIS Web Services Architecture. OpenGIS Discussion Paper OGC 03-025, Open GIS Consortium Inc.
- Liu, Z.; Luo, C. C. J. C. W. W. (2004). A probabilistic approach to metasearching with adaptive probing. In *ICDE '04: Proceedings of the 20th International Conference on Data Engineering*. IEEE Computer Society.
- Liu, J.; Wolfson, O. Y. H. (2006). Extracting semantic location from outdoor positioning systems. In *MDM '06: Proceedings of the 7th International Conference on Mobile Data Management*. IEEE Computer Society.
- Madhavan, J.; Ko, D. K. L. G. V. R. A. H. A. (2008). Google's deep web crawl. In *Proceedings of the VLDB Endowment*, pages 1241–1252.
- Maganto, A. S., Iso, J. N., and Ballari, D. (2008). Normas sobre metadatos (iso19115, iso19115-2, iso19139, iso 15836). *Mapping*, 123:48–57. ISSN 1-131-9-100.
- Maier, M. W. (2008). Architecting principles for systems-of-systems. In *6th Annual International Symposium of INCOSE*, page 567 – 574.
- Manola, F. and Miller, E. (2004). RDF Primer. W3C Recommendation 10 February 2004, W3C. <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>.
- Margoulies, S., editor (2001). *Geocoder Service Specification, v0.7.6*. Open Geospatial Consortium Inc, OGC 01-026r1.
- Martins, B. (2008). *Geographically Aware Web Text Mining*. PhD thesis, Faculty of Sciences, University of Lisbon.
- Martins, B.; Silva, M. (2005a). A graph-ranking algorithm for geo-referencing documents. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 741–744. IEEE Computer Society.
- Martins, B.; Silva, M. C. M. (2005b). Gkb - geographic knowledge base. Technical report, Department of Informatics, University of Lisbon.
- Moses Gone, S. (2007). Towards semantic composition of geospatial web services using wsmo in comparison to bpel. In *Proceedings of GI-Days: Young Researchers Forum*.
- Nebert, D., editor (2001). *Developing Spatial Data Infrastructures: The SDI Cookbook v.1.1*. Global Spatial Data Infrastructure. <http://www.gsdi.org>.
- Nebert(ed.), D. (2004). OpenGIS - Catalogue Services Specification (version: 2.0). OpenGIS Project Document 04-021, Open GIS Consortium Inc. In April 2004 it is only available for OGC members.
- Nedovic-Budic, Z.; Budhathoki, N. R. (2006). Technological and institutional interdependences and sdi - the bermuda square? *International Journal of Spatial Data Infrastructures Research*, 1:36–50.
- Nelson, M.L.; Smith, J. d. C. I. (2006). Efficient, automatic web resource harvesting. In *WIDM '06: Proceedings of the 8th annual ACM international workshop on Web information and data management*, pages 43–50.
- Nogueras-Iso, J., Barrera, J., Rodríguez, A. F., Recio, R., Laborda, C., and Zarazaga-Soria, F. J. (2009). Development and deployment of a services catalog in compliance with the inspire metadata implementing rules. In van Loenen, B., Zevenbergen, J., and Besemer, J., editors, *Spatial Data Infrastructure Convergence: Research, Emerging Trends, and Critical Assessment*, volume 48 of *Groene serie*, pages 21–34. The Netherlands Geodetic Commission (NGC), The Netherlands.
- Nogueras-Iso, J., Latre, M. A., Muro-Medrano, P. R., and Zarazaga-Soria, F. J. (2004). Building eGovernment services over Spatial Data Infrastructures. *3rd International Conference on Electronic Government - EGOV'04, Lecture Notes in Computer Science*, 3183:387–391.

- Nogueras-Iso, J., Zarazaga-Soria, F. J., Béjar, R., and Álvarez P. R. Muro-Medrano, P. J. (2005). Ogc catalog services: a key element for the development of spatial data infrastructures. *Computers and Geosciences*, 31(2):199–209.
- Ntoulas, A.; Cho, J. O. C. (2004). What’s new on the web?: the evolution of the web from a search engine perspective. In *WWW ’04: Proceedings of the 13th international conference on World Wide Web*, pages 1–12.
- Ntoulas, A.; Zerkos, P. C. J. (2005). Downloading textual hidden web content through keyword queries. In *JCDL ’05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, pages 100–109.
- OASIS (2004). Universal Discovery, Description and Integration of Web Services (UDDI) protocol. <http://www.uddi.org/>.
- Official Journal of the European Union (2003). Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. L 345 , 31/12/2003 pp. 0090-0096.
- OpenGIS Consortium (OGC) Inc (2001). Geography Markup Language (GML) 2.0, OpenGIS Implementation Specification. OpenGIS Project Document OGC 01-029.
- Orfali, R., Harkey, D., and Edwards, J. (1999). *Client/Server Survival Guide*. John Wiley & Sons, 3rd edition.
- Parliament, E. and Council, T. E. (2007). Directive of the european parliament and of the council establishing an infrastructure for spatial information in the european community (inspire). joint text approved by the conciliation committee provided for in article 251(4) of the ec treaty. 2004/0175(cod), pe-cons 3685/06.
- Pinkerton, B. (1994). Finding what people want: Experiences with the webcrawler. In *The Second International WWW Conference Chicago*, pages 17–20.
- Pittman, K.; Masters, C. (2006). Cyc’s comprehensive geospatial knowledge base: An overview. In *Networking Geospatial Information Technology for Interoperability and Spatial Ontology*.
- Plews, R.; Kawano, S. (2008). A conceptual framework for the description of place identifiers. In *Proceedings of the ISO Workshop on address standards: Considering the issues related to an international address standard*.
- Pyo, J.; Shin, D. S. T. (2001). Development of a map matching method using the multiple hypothesis technique. *Intelligent Transportation Systems*, pages 23–27.
- Raghavan and Garcia-Molina (2001). Crawling the hidden web. In *VLDB ’01: Proceedings of the 27th International Conference on Very Large Data Bases*, pages 129–138.
- Rizzi, S.; Abelló, A. L. J. T. J. (2006). Research in data warehouse modeling and design: dead or alive? In *DOLAP ’06: Proceedings of the 9th ACM international workshop on Data warehousing and OLAP*, pages 3–10. ACM.
- Robertson, S.; Zaragoza, H. T. M. (2004). Simple bm25 extension to multiple weighted fields. In *CIKM ’04: Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pages 42–49.
- Robertson, S. E.; Walker, S. (1994). Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In Croft, W. B. and van Rijsbergen, C. J., editors, *Proceedings of the 17th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Dublin, Ireland, July 03 - 06, 1994)*. Springer-Verlag New York.
- Sancho-Jiménez, G., Béjar, R., and Muro-Medrano, P. (2008). *Advances in conceptual modeling-challenges and opportunities: ER 2008 Workshops CMLSA, ECDM, FP-UML, M2AS, RIGiM, SeCoGIS, WISM, Barcelona, Spain, October 2008, Proceedings*, volume 5232 of *Lecture Notes in Computer Science*, chapter A method to derivate SOAP interfaces and WSDL metadata from the OGC Web Processing Service mandatory interfaces, pages 375–384. Springer-Verlag.
- Sanderson, M.; Lawrie, D. (2000). Building, testing, and applying concept hierarchies. *Advances in Information Retrieval. The Information Retrieval Series*, 7:235–266.
- Sengar, V.; Joshi, T. J. J. P. S. T. K. (2007). Robust location search from text queries. In *GIS ’07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*.

- Shekhar, S.; Vatsavai, R. R. M. X. Y. J. (2004). Navigation systems: A spatial database perspective. *Location-Based Services*, pages 41–82.
- Somers, R. (1997). *Framework Introduction and Guide*. Federal Geographic Data Committee (FGDC). <http://www.fgdc.gov/framework/frameworkintroguide>.
- Southall, H. (2003). Defining and identifying the roles of geographic references within text. In *Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References*.
- Star, S. L.; Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, 7(1):111–134.
- Syed, S.; Cannon, M. (2004). Fuzzy logic-based map matching algorithm for vehicle navigation system in urban canyons. In *Proceedings of the Institute of Navigation (ION) National Technical Meeting*, pages 982–993.
- Tan, Q.; Mitra, P. G. C. (2007). Designing clustering-based web crawling policies for search engine crawlers. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 535–544. ACM.
- Tolosana-Calasanz, R., Nogueras-Iso, J., Béjar, R., Muro-Medrano, P. R., and Zarazaga-Soria, F. J. (2006). Semantic interoperability based on dublin core hierarchical one-to-one mappings. *International Journal of Metadata, Semantics and Ontologies (IJMS&O)*, 1(3):183–188. Accepted for its publication.
- Uryupina, O. (2003). Semi-supervised learning of geographical gazetteers from the internet. In *Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References*.
- U.S. Federal Register (1994). Executive Order 12906. Coordinating Geographic Data Acquisition and Access: the National Spatial Data Infrastructure (U.S.). *The April 13, 1994, Edition of the Federal Register*, 59(71):17671–17674.
- Vassiliadis, P.; Simitis, A. S. S. (2002). Conceptual modeling for etl processes. In *DOLAP '02: Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP*, pages 14–21.
- Vretanos(Eds), P. (2001). Filter Encoding Implementation Specification, Version 0.0.7. OpenGIS project document OGC 01-067, OpenGIS Consortium Inc.
- Vretanos(Eds), P. (2002). Web Feature Server Implementation Specification. Version 1.0.0. OpenGIS project document OGC 02-058, Open GIS Consortium Inc.
- W3C (2004a). The Extensible Stylesheet Language Family (XSL). <http://www.w3.org/Style/XSL/>.
- W3C (2004b). The Semantic Web Activity. <http://www.w3.org/2001/sw/>.
- Waldinger, R.; Jarvis, P. D. J. (2003). Pointing to places in a deductive geospatial theory. In *Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references*.
- Weiser, A.; Zipf, A. (2007). Web service orchestration of ogc web services for disaster management. In *GI4DM 2007. 3rd International Symposium on Geoinformation for Disaster Management*. Delft.
- Yang, S. (2008). Developing of an ontological focused-crawler for ubiquitous services. In *AINAW '08: Proceedings of the 22nd International Conference on Advanced Information Networking and Applications - Workshops*, pages 1486–1491. IEEE Computer Society.
- Zarazaga, F. J., Bañares, J. A., Bernabé, M. A., Gould, M., and Muro-Medrano, P. R. (2000). La infraestructura nacional de información geográfica desde la perspectiva de bibliotecas digitales distribuidas. In *Actas de las I Jornadas de Bibliotecas Digitales*, pages 163–172, Valladolid, Spain.
- Zarazaga-Soria, F. J., Nogueras-Iso, J., Béjar, R., and Muro-Medrano, P. R. (2004). *Electronic Government, Lecture Notes in Computer Science (LNCS)*, volume 3183 of *Lecture Notes in Computer Science*, chapter Political aspects of Spatial Data Infrastructures, pages 392–395. Zaragoza, Spain.
- Zarazaga-Soria, F. J., Nogueras-Iso, J., and Ford, M. (2003). Dublin core spatial application profile. CWA 14858, CEN/ISSS Workshop - Metadata for Multimedia Information - Dublin Core.
- Zhao, P.; Di, L. (2006). Semantic web service based geospatial knowledge discovery. In *IEEE International Conference on Geoscience and Remote Sensing Symposium, IGARSS 2006*, pages 3490–3493.
- Zheng, G.; Su, J. (2002). Named entity tagging using an hmm-based chunk tagger. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 209–219.
- Zong, W.; Wu, D. S. A. L. E. . G. D. H. (2005). On assigning place names to geography related web pages. In *JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, pages 354–362. ACM.



MINISTERIO
DE CIENCIA
E INNOVACION



REUNIÓN NACIONAL DE LA INFRAESTRUCTURA DE DATOS ESPACIALES DE ESPAÑA (IDEE)

LAS INFRAESTRUCTURAS DE DATOS ESPACIALES (IDE) Y SUS
APLICACIONES EN PROYECTOS DE INVESTIGACIÓN CIENTÍFICO-
TECNOLÓGICOS

Lugar de celebración:

Centro de Ciencias Humanas y Sociales (CCHS) del Consejo Superior de Investigaciones Científicas (CSIC)

**24 – 25 febrero
Madrid, 2010**

Organizan



Consejo Superior Geográfico
Infraestructura de Datos Espaciales de España



Centro de Ciencias
Humanas y Sociales

24 DE FEBRERO/JORNADA 1

09:00-10:00 Recepción de participantes y entrega de documentación.

10:15-10:30 APERTURA DE LA JORNADA

Eduardo Manzano Moreno (*Director del Centro de Ciencias Humanas y Sociales, CCHS*). Consejo Superior de Investigaciones Científicas (CSIC).

Sebastián Mas Mayoral (*Presidente de la Comisión de Infraestructura de Datos Espaciales de España del Consejo Superior Geográfico - Instituto Geográfico Nacional, IGN*).

10:30-11:00 “EuroSDR , Red Pan Europea de Investigación de Datos Espaciales”.

Antonio Arozarena Villar. Presidente de EuroSDR.

European Spatial Data Research.

11:00-11:30 Pausa – café.

BLOQUE I IDE y sus aplicaciones en proyectos de investigación científico-tecnológicos

11:30-13:00 SESION 1: Patrimonio Histórico

(Moderador: Juan M. Vicent García)

“SILEX: IDE del proyecto arqueológico Casa Montero. Servicios Web de información temática basados en arquitectura REST”.

Alfonso Fraguas, Antonio Menchero, Antonio Uriarte, Juan Vicent, Susana Consuegra, Pedro Díaz-del-Río, Nuria Castañeda, Cristina Criado, Enrique Capdevila y Marta Capote.

Instituto de Historia. CCHS-CSIC.

“Adopción de las recomendaciones del GT-IDEE en la creación de IDE de Arqueología del paisaje: El ejemplo de la IDE de Zona Arqueológica de Las Médulas (IDEZAM)”.

Miguel Lage Reis-Correia ⁽¹⁾. Juan Luis Pecharroman Fuente ⁽²⁾, María Ruíz del Arbol Moro ⁽²⁾, Francisco Javier Sánchez-Palencia Ramos ⁽²⁾.

⁽¹⁾ **Fundación Las Médulas** ⁽²⁾ **Instituto de Historia. CCHS-CSIC.**

“El Archivo Epigráfico de Hispania: como acercar el patrimonio arqueológico a la investigación multidisciplinar a través de las IDE”.

Mercedes Farjas⁽¹⁾, Isabel Velázquez⁽²⁾, Alberto López⁽²⁾, Joaquín L. Gómez-Pantoja⁽³⁾ y Fernando Luis Álvarez ⁽³⁾.

⁽¹⁾**Univ. Politécnica de Madrid.** ⁽²⁾ **Univ. Complutense de Madrid.** ⁽³⁾ **Univ. de Alcalá de Henares.**

“La gestión espacial del Patrimonio y la problemática de la integración en IDE”.

José Julio Zancajo Jimeno y Teresa Mostaza Pérez.

Universidad de Salamanca.

“La Información Espacial del Patrimonio Cultural en el Instituto Andaluz del Patrimonio Histórico”.

Carmen Pizarro Moreno y David Villalón.

Instituto Andaluz de Patrimonio Histórico.

24 DE FEBRERO/JORNADA 1

13:00-14:30 **SESION 2: Otras iniciativas IDE en la investigación científica**
(Moderadora: Ana M. Crespo Solana)

"SerGEO: Gestor de visualización y descarga de datos geográficos en el CSIC".
Isabel del Bosque González ⁽¹⁾, Rocío Gutiérrez González ⁽¹⁾ y Lorenzo Mateos Corchero ⁽²⁾.

⁽¹⁾ **Centro de Ciencias Humanas y Sociales (CCHS)-CSIC.** ⁽²⁾ **ESRI España Geosystems.**

"Las IDE y su aplicación en proyectos de investigación de la Universidad de Zaragoza".
R. Béjar, Pedro Muro-Medrano, J. Zarazaga, J. Nogueras, M.A. Latre, F.J. López.
Universidad de Zaragoza (UNIZAR). Grupo de Sistemas de Información Avanzados.

"SIGMayores: Una herramienta para el Atlas de Recursos Sociales y Sanitarios de España".

Clara Bécares, Antonio Abellán, Joaquín Siabra, Equipo Portal Mayores y Unidad SIG.
Centro de Ciencias Humanas y Sociales (CCHS)-CSIC.

"Visualización y análisis de redes mercantiles en la Primera Edad Global (1400-1800): Una IDE histórica".

Esther Pérez Asensio, Roberto Maestre Martínez, Isabel del Bosque González, Ana Crespo Solana y Juan Manuel Sánchez-Crespo.

Centro de Ciencias Humanas y Sociales (CCHS)-CSIC.

"Resultados y situación actual de los proyectos EuroGEOSS y eSDINet+".

Carlos Granell, Laura Díaz, Joaquín Huerta.

Universidad Jaume I de Castellón.

14:30-15:30 **Pausa – comida.**

15:30-17:30 **SESION 3: IDE y Recursos Naturales**
(Moderador: F. Javier Sanz Cañada)

"La Infraestructura Mundial de Información en Biodiversidad (GBIF) en el contexto de las IDE".

Francisco Pando. GBIF-España.

Real Jardín Botánico (RJB) – CSIC.

"Servidor de cartografía digital de Doñana: pasado, presente y futuro".

Ricardo Díaz Delgado.

Estación Biológica de Doñana (EBD) – CSIC.

"Anthos.es. Una década de información sobre las plantas de España en internet".

Leopoldo Medina y Carlos Aedo.

Real Jardín Botánico (RJB) – CSIC.

"Infraestructura de Mapas Geológicos: Proyectos OneGeology y OneGeology-Europe".

Fernando Pérez Cerdán, Ángel Prieto Martín, María J. Mancebo Mancebo y Emilio González Clavijo.

Instituto Geológico y Minero de España (IGME).

24 DE FEBRERO/JORNADA 1

"Diseño de Servicios Web para generar mapas de estimación de daños causados por riesgos naturales".

Miguel A. Manso y Vladimir Gutiérrez.

Universidad Politécnica de Madrid (UPM).

"Hacia una IDE marina: Una IDE para buques oceanográficos y su aproximación desde los sistemas de información científicos oceanográficos".

J. Olivé, M. Farrán, J. Martínez, J. Sorribas, M. Martínez, F. Pérez, O. Chic, E. Arilla, S. Soto, D. Afonso, O. García, A. Hernández, J.L.Ruiz, X. Romero, A. Sandoval, J. A. Serrano, J. Guillén, J. Piera, E. García-Ladona.

Centro Mediterráneo de Investigaciones Marinas y Ambientales (CMIMA) – CSIC.

"IDEO: Desarrollo de una IDE científica en el Instituto Español de Oceanografía".

Nuria Hermida Jiménez, Elena Pastor García, Olvido Tello Antón.

Instituto Español de Oceanografía (IEO).

BLOQUE II Reuniones de los Subgrupos de Trabajo de la IDEE

17:30-19:30 **Reuniones paralelas de los siguientes subgrupos de trabajo:**

Foro de Direcciones: Sala "Herbert A. Simón" (3D18-3D6)

SGT-IDE Local: Sala "Gómez Moreno" (2C10-2C24)

SGT- Arquitectura y Normas: Sala "Juan Cabré" (2D21-2D7)

25 DE FEBRERO/JORNADA 2

Reunión ordinaria del Grupo de Trabajo para el establecimiento de la Infraestructura de Datos Espaciales de España (IDEE)

(Moderador: Sebastián Más Mayoral)

09:30-11:30 Reunión ordinaria del GT-IDEE.

11:30-12:00 **Pausa – café.**

12:00-14:00 Reunión ordinaria del GT-IDEE.

14:00 **Cóctel.**