

Departamento de Ingeniería Topográfica y Cartografía  
Escuela Técnica Superior de Ingenieros en Topografía, Geodesia y Cartografía

# **El uso de los metadatos para el desarrollo de un modelo de interoperabilidad para las Infraestructuras de Datos Espaciales**

Autor: Miguel Ángel Manso Callejo (Ingeniero de Telecomunicación)

Directores:

Dra. Mónica Wachowicz

Dr. Miguel Ángel Bernabé Poveda

Año 2009



## AGRADECIMIENTOS

Este trabajo no hubiera sido posible sin la ayuda de mucha gente.

Deseo expresar mi agradecimiento a todos los compañeros del Departamento de Ingeniería Topográfica y Cartografía de la Universidad Politécnica de Madrid, a todos los compañeros del grupo de trabajo MERCATOR – Tecnologías de la GeoInformación y del laboratorio de Tecnologías de la Información Geográfica (LatinGEO) por los apoyos y estímulos recibidos.

Además, en particular, a mis directores: Dra. Mónica Wachowicz por las pautas, las discusiones y por los alientos y al Dr. Miguel Ángel Bernabé Poveda por la paciencia, comprensión y las sugerencias.

Deseo manifestar mi gratitud a la Universidad Politécnica de Madrid por el programa de contratación de profesores visitantes que ha hecho posible que la Dra. Wachowicz sea directora de esta tesis, y al Instituto Geográfico Nacional por las distintas colaboraciones.

Finalmente, deseo agradecer a mis padres por la educación y el continuo aliento de progreso, a Fabiola por el inestimable amor, apoyo y ayuda con la gramática y la alegría y sonrisas de mi hijo que me han ayudado en los momentos críticos.



## RESUMEN

Se propone un modelo de interoperabilidad basado en los metadatos para las Infraestructuras de Datos Espaciales (IDE), un método para crear automáticamente metadatos y una metodología que permite analizar la interoperabilidad aportada por los mismos. Los metadatos son una pieza fundamental para las IDE; catalogan la información geográfica (IG), describen sus características, la calidad, las condiciones, etc. y permiten las funciones: localizar, evaluar, acceder y explotar la IG. La interoperabilidad es un objetivo esencial para que se comparta, coopere, comunique e intercambie IG en las IDE. La formulación de modelos de interoperabilidad permite analizar el comportamiento de los sistemas desde distintos enfoques o niveles. La carencia de modelos de interoperabilidad aplicados a las IDE, de estudios que analicen la interoperabilidad proporcionada por los metadatos y de métodos de creación automática de metadatos, constituyen los objetivos de investigación de esta tesis.

El modelo de interoperabilidad propuesto para las IDE estudia los niveles definidos en los modelos aplicados a los sistemas de sistemas, por considerarlas un caso particular de ellos, y un nivel adicional para tratar los aspectos legales y de las organizaciones. En este contexto de los modelos de interoperabilidad basados en los metadatos, parece necesario disponer de un método original que facilite la creación automática de metadatos y de una metodología que permita analizar la interoperabilidad proporcionada por los mismos. El método propuesto para crear automáticamente metadatos estructura el proceso de compilación y tratamiento de la información, compone y almacena el metadato estandarizadamente y puede integrarse en los flujos de trabajo de la IG. El análisis de la interoperabilidad que proporcionan los metadatos ISO-19115 ha permitido interpretarlos desde un punto de vista alternativo al tradicional basado en funciones. La validación del modelo, mediante una encuesta a expertos, ha disipado la incertidumbre en torno a la subjetividad de la identificación de la interoperabilidad proporcionada por los metadatos. El análisis de la interoperabilidad potencial de un metadato ha permitido identificar los niveles favorecidos (semántico, dinámico y organizacional) y las carencias. Finalmente, el estudio de los metadatos que se crean automáticamente con el método propuesto, ha permitido conocer el potencial de interoperabilidad de los mismos y analizar si la creación automática satisface los requisitos de las organizaciones.

## **ABSTRACT**

An interoperability model is proposed based on the Spatial Data Infrastructure (SDI) metadata, as well as a method for automatic metadata creation and a methodology allowing analysis of the interoperability provided by them. Metadata constitute an essential piece for SDI's; they catalogue geographic information (GI), describe its characteristics, quality, conditions, etc, and their roles are: discovery, evaluation, access and exploitation of GI. Interoperability is an essential aim for GI to be shared, cooperated, communicated and exchanged in SDI's. The formulation of interoperability models allows analyzing system behaviour from different approaches or levels. The lack of interoperability models applied to SDI's, and the lack of studies analyzing the interoperability provided by metadata and of methods of automatic metadata creation constitute the research aims of this thesis.

The proposed interoperability model for SDI's considers the levels defined in the models applied to the systems of systems since they are considered a specific case of those, and an additional level to deal with legal and organizational aspects. In the context of the metadata-based interoperability models, it seems necessary to possess an original method advancing the automatic metadata creation as well as a methodology allowing analysis of the interoperability provided by them. The proposed method to automatically create metadata organize the process of information compilation and handling, it composes and stores metadata in a standard fashion and may be integrated into GI workflows. The analysis of the interoperability provided by the ISO-19115 metadata has allowed their interpretation from an alternative viewpoint different than the traditional function-based approach. The validation of the model with the help of an expert survey has dispelled the uncertainty around the subjectivity of the interoperability identification provided by the metadata. The analysis of the potential interoperability of metadata has resulted in the identification of favoured levels (semantic, dynamic and organizational) as well as the deficiencies. Finally, the study of the automatically created metadata with the proposed method has enabled knowledge of their interoperability potential and clearing up whether automatic creation fulfils the requirements of institutions and organizations.

# INDICE

INDICE .....	v
LISTA DE FIGURAS .....	ix
LISTA DE TABLAS .....	x
1 INTRODUCCIÓN.....	1
1.1 Los metadatos en la interoperabilidad de las Infraestructuras de Datos Espaciales .....	2
1.2 Descripción del proyecto de investigación.....	4
1.3 Cuestiones que se pretende responder por la investigación .....	6
1.4 Estructura de la tesis.....	7
2 METADATOS .....	9
2.1 Introducción.....	9
2.2 Definiciones de metadatos.....	9
2.3 Funciones que desempeñan los metadatos .....	12
2.4 Taxonomía de los metadatos .....	18
2.5 Procedimientos de creación de metadatos .....	21
2.6 Sumario .....	29
3 INTEROPERABILIDAD.....	31
3.1 Introducción.....	31
3.2 Importancia de la interoperabilidad para las IDE.....	31
3.3 Definiciones de interoperabilidad .....	33
3.4 Modelos de interoperabilidad .....	37
3.5 Niveles de interoperabilidad.....	42
3.5.1 Interoperabilidad Semántica .....	44
3.5.2 Interoperabilidad Sintáctica.....	49
3.5.3 Interoperabilidad Técnica .....	51
3.5.4 Interoperabilidad Pragmática .....	53
3.5.5 Interoperabilidad Organizacional .....	54
3.5.6 Interoperabilidad Esquemática o Estructural .....	56
3.5.7 Interoperabilidad Dinámica .....	57
3.5.8 Interoperabilidad Legal .....	57
3.5.9 Interoperabilidad Conceptual .....	58
3.5.10 Interoperabilidad Social .....	59

3.5.11 Interoperabilidad intracomunitarios .....	59
3.5.12 Interoperabilidad Política/Humana.....	59
3.5.13 Interoperabilidad internacional (lingüística) .....	60
3.5.14 Interoperabilidad Empírica.....	60
3.5.15 Interoperabilidad Física .....	60
3.6 Medidas de la interoperabilidad .....	61
3.7 Sumario .....	65
4 METODOLOGÍA .....	69
4.1 Introducción.....	69
4.2 Metodología de creación automática de metadatos.....	70
4.3 Modelo de interoperabilidad aplicado en las IDE.....	77
4.4 Clasificación de los ítems de metadatos según el modelo de interoperabilidad.....	83
4.4.1 Identificación de los niveles de interoperabilidad de los ítems de metadatos.....	84
4.4.2 Relaciones entre los niveles de interoperabilidad del modelo .....	88
4.5 Conclusiones .....	98
5 IMPLEMENTACIÓN Y RESULTADOS .....	101
5.1 Introducción.....	101
5.2 Formatos de almacenamiento usados en las IDE .....	102
5.3 Creación automática de metadatos .....	115
5.4 Funciones desempeñadas y niveles de interoperabilidad favorecidos por los ítems de metadatos creados automáticamente .....	127
5.5 Validación del modelo de interoperabilidad .....	133
5.5.1 Definición de las encuestas .....	134
5.5.2 Procesamiento de las encuestas .....	135
5.6 Conclusiones .....	148
6 CONCLUSIONES.....	153
6.1 Revisión de las cuestiones de investigación.....	154
6.1.1 Modelo de interoperabilidad para las IDE .....	155
6.1.2 Análisis de la interoperabilidad proporcionada por los metadatos.....	156
6.1.3 Metodología para la creación automática de metadatos.....	157
6.1.4 Validación del modelo de interoperabilidad .....	158

6.1.5 Interoperabilidad proporcionada por los metadatos creados automáticamente.....	159
6.2 Futuras líneas de investigación.....	160
7 BIBLIOGRAFÍA.....	162

## LISTA DE FIGURAS

Figura 2. 1: Posibles métodos de creación de metadatos (Adaptado proyecto ARIADNE).....	26
Figura 2. 2 Flujo de creación de metadatos (Fuente: proyecto NCGDAP).....	27
Figura 3. 1: Modelo de interoperabilidad Conceptual de Turnitsa y Tolk (2006)	39
Figura 3. 2: Histograma de los niveles de interoperabilidad	44
Figura 4. 1: Método de creación automática de metadatos: etapas y relaciones ...	71
Figura 4. 2: Propuesta sobre modelo integrado de interoperabilidad.....	78
Figura 4. 3: Esquema de datos del documento de análisis de interoperabilidad ...	85
Figura 4. 4: Histograma de los niveles de interoperabilidad en los ítems del núcleo ISO19115.....	86
Figura 4. 5: Histograma de los niveles de interoperabilidad en los ítems de la norma ISO19115 .....	87
Figura 4. 6: Distribución de las cantidades de ítems del núcleo de la norma de metadatos que sólo proporcionan un nº de niveles del modelo.....	90
Figura 4. 7: Distribución de las cantidades de ítems de la norma de metadatos que sólo proporcionan un nº de niveles del modelo.....	95
Figura 5.1: Porcentajes de ítems creados automáticamente por nivel de interoperabilidad.....	132
Figura 5.2: Apariencia y estructura de la hoja de cálculo de la encuesta.....	134
Figura 5.3: Histograma de las intensidades de interoperabilidad por encuestado .....	136

Figura 5.4: Tablas de las relaciones entre los niveles del modelo de interoperabilidad según los ítems del núcleo para cada encuestado.....	137
Figura 5.5: Diagramas radiales con las relaciones entre los niveles del modelo de interoperabilidad.....	138
Figura 5.6: Gráfico con las concordancias y discrepancias entre encuestados y modelo a validar .....	140
Figura 5.7: Grado de acuerdo en los niveles de interoperabilidad por ítems y grupos de encuestados .....	142
Figura 5.8: Tabla comparativa con los niveles de interoperabilidad de los ítems del núcleo de la norma de metadatos antes y después de la validación .....	145
Figura 5.9: Histograma de los niveles de interoperabilidad en los ítems del núcleo de la norma de metadatos una vez validado .....	146
Figura 5.10: Intensidades de las relaciones entre los niveles del modelo una vez validado .....	146

## LISTA DE TABLAS

Tabla 3. 1: Esquema de interoperabilidad de 8 niveles (Goodchild et al. 1997)...	41
Tabla 3. 2: Intermodel5, Modelo de Interoperabilidad de 5 niveles (Shanzhe, 1999).....	41
Tabla 3. 3: Tabla resumen de los niveles de Interoperabilidad identificados en la literatura.....	43
Tabla 4. 1 Frecuencias de los niveles de interoperabilidad en los ítems del núcleo fundamental .....	85
Tabla 4. 2: Frecuencias de los niveles de interoperabilidad en los ítems de la norma.....	86
Tabla 4. 3: Relaciones entre los niveles del modelo de interoperabilidad según los ítems del núcleo.....	88
Tabla 4. 4: Número de ítems del núcleo de la norma de metadatos que sólo proporcionan uno, dos, etc., niveles en el modelo .....	89
Tabla 4. 5: Niveles de interoperabilidad proporcionados simultáneamente .....	90
Tabla 4. 6: Representación gráfica del número de ítems que proporcionan dos niveles del modelo.....	92

Tabla 4. 7: Relaciones entre los niveles del modelo de según los ítems de la norma .....	94
Tabla 4. 8: Número de ítems de la norma de metadatos que sólo proporcionan uno, dos, etc., niveles en el modelo.....	95
Tabla 5.1: Formatos ráster genéricos .....	105
Tabla 5.2: Formatos ráster para grandes volúmenes de datos .....	106
Tabla 5.3: Formatos ráster utilizados en teledetección .....	107
Tabla 5.4: Formatos ráster utilizados en modelos digitales del terreno .....	108
Tabla 5.5: Formatos vectoriales tipo CAD.....	109
Tabla 5.6: Formatos vectoriales tipo SIG.....	110
Tabla 5.7: Bases de datos con extensiones espaciales.....	111
Tabla 5.8: Ejemplos de descripción numérica estandarizada: GeoTIFF, MrSID	113
Tabla 5.9: Ejemplos de descripción numérica no estandarizada.....	113
Tabla 5.10: Ejemplos de descripción nemotécnica .....	113
Tabla 5.11: Ejemplos de descripción mediante texto cuasi-estructurados.....	113
Tabla 5.12: Ejemplos de descripción mediante texto estructurado (WKT) .....	113
Tabla 5.13: Ejemplos de descripciones numéricas en bases de datos espaciales	114
Tabla 5. 14: Tipos y cantidades de identificadores diferentes en los formatos..	115
Tabla 5. 15: Clasificaciones de ítems de metadatos creados automáticamente ..	130
Tabla 5.16: Número de ítems metadatos que la favorecen los niveles de interoperabilidad.....	131
Tabla 5.17: Intensidades de los niveles de interoperabilidad por encuestado al analizar los ítems del núcleo fundamental de la norma ISO19115 de metadatos	135
Tabla 5. 18: Relaciones entre los niveles del modelo de interoperabilidad una vez validado .....	146



# 1 INTRODUCCIÓN

El objetivo global de este trabajo de investigación es aumentar el conocimiento sobre los modelos y niveles de interoperabilidad en las Infraestructuras de Datos Espaciales (IDE), y evidenciar cómo los metadatos, creados manual o automáticamente, que describen la Información Geográfica (IG), pueden ayudar a mejorar dicha interoperabilidad. Más concretamente, los objetivos son: (a) definir un modelo de interoperabilidad para las IDE basado en niveles de interoperabilidad proporcionados por los metadatos y evaluar la aportación de cada elemento de un registro de metadatos a cada nivel del modelo, (b) definir un método que permita automatizar la creación de metadatos sobre la IG, y (c) definir una metodología que permita evaluar la interoperabilidad que proporcionan los metadatos. Con el uso de la metodología y el modelo de interoperabilidad se va a poder adquirir información sobre la intensidad y variabilidad de niveles de interoperabilidad proporcionados por los metadatos. Independientemente del modo en el que hayan sido creados, se podrán analizar distintos perfiles de metadatos para evaluar las bondades y carencias a nivel de interoperabilidad, de tal modo que podrá utilizarse en las revisiones de las normas, para incluir nuevos elementos de un registro de metadatos (en adelante ítems) que maximicen la interoperabilidad de las IDE.

A lo largo de esta investigación, se ha definido un modelo original de interoperabilidad fundamentado en los modelos generales de sistemas y adaptado a las características particulares de las IDE; se ha desarrollado una metodología original para evaluar la interoperabilidad, basada en el contenido de los ítems de los metadatos para el contexto de las IDE; se ha desarrollado una nueva metodología que permite automatizar la creación de metadatos; se ha validado el modelo de interoperabilidad y, finalmente, se ha analizado la interoperabilidad que pueden proporcionar los metadatos creados con la nueva metodología, de acuerdo con el nuevo modelo.

El objetivo de éste capítulo de introducción es proporcionar una visión general sobre el uso de los metadatos para el desarrollo de un modelo de interoperabilidad para las Infraestructuras de Datos Espaciales. El primer apartado explica los

motivos del proyecto de investigación y el contexto. Después, se profundiza más en la tesis con la descripción del proyecto de investigación, seguido del tercer apartado, en el que se describen los principales objetivos a los que se apunta en esta tesis. En el cuarto apartado se presenta la estructura de la misma.

## **1.1 Los metadatos en la interoperabilidad de las Infraestructuras de Datos Espaciales**

La localización y acceso mediante el uso de buscadores especializados a datos que se encuentran distribuidos por la Web, implica que éstos se encuentran catalogados mediante descriptores homogéneos, esto es, mediante sus metadatos. Esta es la razón por la que una pieza fundamental de las IDE sean los metadatos, que se consideran como una herramienta y un formalismo necesarios para poder acceder a los datos espaciales tanto de forma manual como de una forma automatizada (Najar y Giger 2006), en este último caso, de manera que los sistemas que manejan información espacial puedan interoperar. Los metadatos son “datos acerca de datos” (ANZLIC 1996; Kildow 1996; ANZLIC 1997), que se aplican profusamente en todo tipo de recursos electrónicos de información (Milstead y Feldman 1999) y, en el caso especial de la información geográfica (IG), se utilizan para describir el contenido, la calidad, las condiciones y otras características de los datos, de forma que los usuarios de IG puedan acceder a ellos y explotarlos.

Tradicionalmente, los metadatos han ocupado un papel secundario y generalmente, se crean con posterioridad a la producción o adquisición de los datos, ya que las organizaciones consideran que la creación de los mismos supone, en muchos casos, un sobre coste (Najar 2006). Esta afirmación es rebatida con el siguiente argumento: *“si usted piensa que el coste de creación de los metadatos es demasiado alto, es que no ha calculado los costes derivados de su inexistencia: pérdida de información por cambio de personal, duplicidad de datos, conflictos de datos, responsabilidad, mal uso y toma de decisiones argumentadas pobremente”* (CGIAR-CSI 2004).

Una consecuencia de que los metadatos se creen manualmente y transcurrido un tiempo de la creación de los datos espaciales que describen, es la propensión a los

errores y a la carencia de información, conduciendo a que en algunas ocasiones, esta tarea sea difícil o imposible de realizar (Beard 1996). Se debe poner énfasis en la búsqueda de herramientas de creación automática de metadatos que, además de reducir el tiempo y los esfuerzos de catalogación de la IG, eviten los errores mecanográficos de la transcripción y los errores de interpretación y tratamiento de propiedades de los datos geográficos, tales como el sistema de referencia espacial o las coordenadas.

La interoperabilidad, entendida como la capacidad de intercambiar información entre dos sistemas, puede ser analizada desde distintos puntos de vista: de los datos, de los servicios, de las aplicaciones, de las organizaciones, otros, .. (William 2002, Gordon 2003). También puede ser analizada a distintos niveles: tecnológico, sintáctico, semántico (Tolk 2003; ISO19101; Turnitsa 2006). En el contexto de los Sistemas de Información Geográficos (SIG/GIS), la interoperabilidad se ha convertido en un objetivo de investigación desde hace una década, con la finalidad de implementar aplicaciones informáticas orientadas al intercambio de datos. Se identificaron ocho niveles de interoperabilidad: 1) usuarios e instituciones, 2) empresa o compañía, 3) aplicación, 4) herramientas y utilidades, 5) intermediarios, 6) almacenes de datos, 7) computación distribuida y 8) red (Goodchild *et al.* 1997).

La formulación de modelos de interoperabilidad (arquetipos en los que estructurar los objetivos o clases de interoperabilidad) que permitan realizar medidas de la conectividad de diferentes niveles de interoperabilidad de una forma crítica, es una tarea esencial para el desarrollo de abstracciones. En 2004, NATO definió su primer modelo de interoperabilidad para definir infraestructuras para arquitecturas de sistemas (NC3SAF) basado en la directiva para el desarrollo de las mismas. Posteriormente, el Instituto de ingeniería informática (SEI: Carnegie Mellon University) ha propuesto varios modelos de interoperabilidad tales como: niveles de interoperabilidad para sistemas de información (LISI), niveles de interoperabilidad conceptual (LCIM), interoperabilidad para sistemas de sistemas (SOSI) y niveles para la interoperabilidad conceptual (LCI) (Tolk 2003). Aunque existe abundante literatura sobre modelos de interoperabilidad y sobre los respectivos niveles, no se ha realizado una investigación en el desarrollo de

modelos de interoperabilidad para la implementación de las Infraestructuras de Datos Espaciales (Groot y McLaughlin 2000; Bernard *et al.* 2005), y tampoco se han encontrado referencias sobre el uso de los ítems almacenados en los metadatos de la información geográfica, en relación con los modelos de interoperabilidad. No se conocen modelos de interoperabilidad basados en metadatos (MIBM) con aplicación a las IDE.

## **1.2 Descripción del proyecto de investigación**

El principal objetivo de esta investigación es explorar el uso de un MIBM entre sistemas, para aplicarlo en la creación automática de metadatos para las IDE.

Actualmente, en la literatura se pueden distinguir dos conceptos y formas de presentación de metadatos de la IG:

- Como conjunto de reglas incluidas en las aplicaciones informáticas que manejan información geográfica para describir la estructura interna y esquemas de los datos (Codd 1990; Korth y Silberschatz 1991; Wilson 1998) y,
- Como productos autónomos asociados a los conjuntos de datos geográficos que proporcionan información a los catálogos de datos, de servicios y las IDE, de modo que se posibilite el uso compartido de los datos, además de organizar y mantener un inventario de los mismos o proporcionar información que posibilite su transferencia y su uso (Phillips 1998; FGDC 2000; Najar 2006).

La hipótesis de esta tesis es que los metadatos son el soporte que permite definir con eficiencia, completitud y precisión los diferentes niveles de interoperabilidad para las IDE.

La innovación de la investigación reside en: la definición de un modelo de interoperabilidad para las IDE; en el estudio de la interoperabilidad facilitada por la norma internacional ISO19115 de metadatos; y la definición de una nueva metodología de creación automática de metadatos de la IG.

Las contribuciones de la investigación realizada son:

- El diseño de un nuevo modelo de interoperabilidad entre sistemas aplicable a las IDE. En este diseño se han considerado, además de los aspectos comunes de los sistemas, los aspectos relacionados con las organizaciones.
- El diseño de una nueva metodología para la creación automática de metadatos relativos a la IG, susceptible de ser integrada en los flujos de trabajo de creación de metadatos.
- El estudio de los metadatos de la IG desde el punto de vista de la interoperabilidad, aportándose una metodología de análisis de la interoperabilidad favorecida por los ítems de la norma internacional de metadatos ISO19115, aplicable a los distintos perfiles que de ella se realicen, y extensible a nuevas normas
- La aplicación y validación del análisis de la interoperabilidad, proporcionada por los ítems de la norma de metadatos
- La aplicación de la metodología de creación automática de metadatos para distintos tipos de IG (ráster, vector, bases de datos) y distintos formatos de almacenamiento (se describe objetivamente el alcance de la metodología a nivel de ítems de metadatos, que pueden generarse desde un punto de vista teórico y agregado para las distintas tipologías de IG)
- La evaluación de la metodología para la creación automática de metadatos, desde el punto de vista de la interoperabilidad (se analizan los ítems creados atendiendo a los niveles de interoperabilidad del modelo, y se comparan los resultados con los máximos posibles a nivel de norma y de los ítems del núcleo fundamental de la misma)

### 1.3 Cuestiones que se pretende responder por la investigación

Esta tesis responde y amplía la motivación original de la investigación, siendo el objetivo principal definir un modelo de interoperabilidad adecuado para las IDE, sobre el que analizar la capacidad de los metadatos para facilitar los distintos niveles del modelo y estudiar cómo los metadatos generados automáticamente facilitan la interoperabilidad, además de evitar la monotonía de su creación y los errores consecuentes.

La aplicabilidad de los modelos generales de interoperabilidad definidos para los sistemas en el contexto de las IDE, la utilidad de los metadatos en el contexto de la interoperabilidad, los prejuicios existentes contra los metadatos de la IG y la dureza de su creación manual, han motivado que en esta investigación se hayan planteado diversas cuestiones de investigación.

Como resultado, las cinco cuestiones de investigación en torno a la que gira este trabajo son:

- ¿Se puede formalizar un modelo de interoperabilidad de sistemas para las infraestructuras de datos espaciales?
- ¿Cuál es la aportación, en términos de interoperabilidad, de la información contenida en los metadatos?
- ¿Se pueden crear metadatos útiles de la información geográfica de un modo automático y eficaz?
- ¿Qué propuesta de validación es más adecuada para validar un modelo de interoperabilidad de sistemas en el contexto de las infraestructuras de datos espaciales?
- ¿Cuáles son las fortalezas y las debilidades de los metadatos generados manual y automáticamente, desde el punto de vista de la interoperabilidad de los sistemas que los explotarán (IDE)?

Se evidencia, a partir de esta lista de cuestiones de investigación, que la tesis se centra en cuestiones metodológicas y es una investigación aplicada. En resumen, el objetivo principal de esta tesis es el diseño, desarrollo y evaluación de un modelo de interoperabilidad basado en metadatos creados automáticamente en el contexto de las IDE. Este objetivo implica: (a) Crear una metodología de

evaluación de la interoperabilidad proporcionada por los metadatos; (b) Crear una metodología para automatizar la creación de metadatos de la IG.

#### **1.4 Estructura de la tesis**

Para finalizar esta introducción se describe la estructura de esta tesis. Tras de esta introducción (Capítulo 1), se trata sobre el estado del arte y conocimiento en torno a los metadatos (Capítulo 2) y la interoperabilidad (Capítulo 3). A continuación, dando respuesta a las tres primeras cuestiones de investigación, se desarrolla la metodología de creación automática de metadatos; se define el modelo de interoperabilidad para las IDE y se realiza la clasificación de los ítems de la norma internacional de metadatos (ISO19115), según el modelo de interoperabilidad (Capítulo 4). La última parte de la tesis, que responde a las cuestiones cuarta y quinta de investigación, presenta la implementación con los resultados de aplicar la metodología de la creación automática de metadatos, el análisis de la interoperabilidad facilitada por los mismos y la validación del modelo (Capítulo 5). La tesis concluye con las respuestas a las cuestiones de investigación y la propuesta de trabajos futuros (Capítulo 6).



## 2 METADATOS

### 2.1 Introducción

Los metadatos son una pieza fundamental para las IDE, describen los recursos (datos, servicios y otros objetos) y facilitan que los usuarios y las aplicaciones los consulten en catálogos. Los metadatos se pueden considerar un formalismo y una herramienta necesaria que habilita el acceso a los datos y los servicios de un modo automatizado (Najar y Giger, C. 2006), facilitando de esta manera la interoperabilidad de los sistemas que utilizan la información espacial. Los metadatos además de definirse como “datos acerca de datos” (ANZLIC 1996; Kildow 1996; ANZLIC 1997), y utilizarse extensamente en todo tipo de recursos electrónicos de información (Milstead y Feldman 1999), se utilizan en el dominio de la IG para describir el contenido, la calidad, las condiciones y otras características de los datos. En este capítulo se hace una revisión bibliográfica en torno a las definiciones del término metadatos, las funciones que desempeñan y las distintas formas de almacenarse, percibirse y tratarse. También se han revisado los procedimientos de creación de metadatos, poniendo especial énfasis en los automáticos o los que lo automatizan.

### 2.2 Definiciones de metadatos

A continuación se presentan diferentes definiciones del término “metadato”. Estas definiciones exponen los distintos puntos de vista del término, incluyendo aquellas que buscan su significado en sus raíces, las definiciones acuñadas en el contexto de la biblioteconomía, las actuales definiciones relacionadas con Internet y la Web semántica y, finalmente, aquellas directamente relacionadas con la IG.

Los metadatos (*metadata*, del griego μετά, «después de» (RAE) y de “data” plural del latín datum-i, «lo que se da», «dato» (RAE)), que literalmente significa « más allá de los datos», describen otros datos, entendiéndose por lo general, que un grupo de metadatos describe a un grupo de datos o recursos.

Lack E. Myers (1969) acuñó el término “*metadata*” para describir conjuntos de datos, productos. La primera vez que apareció impreso el término en un folleto fue en 1973. Desde entonces, tanto *metadata* como *meta-data* ha sido adoptado por los dominios del conocimiento: informática, estadística, bases de datos y la biblioteconomía, con el significado “informaciones que describen a los datos” (también se ha popularizado la expresión “datos acerca de los datos”). Desde este enfoque, el término *metadata* engloba todos los atributos de los datos que lo describen, proporcionan contexto, indican la calidad o simplemente documentan las características de un objeto o un dato.

Desde un punto de vista informático los metadatos se consideran un conjunto de reglas incluidas en las aplicaciones de manejo de información geográfica que describen la estructura interna de los esquemas de datos (Codd 1990; Korth and Silberschatz 1991; Wilson E. 1998)

Algunos autores como Milstead y Feldman (1999) o Caplan (1995) diferencian las acciones de crear metadatos y catalogar recursos. La principal diferencia la identifican en los metadatos, indicando que sólo se refieren a información electrónica.

Para Caplan (1995), el término “*metadato*” se utiliza de un modo neutral para apartarse de los posibles prejuicios que puedan tener todas aquellas personas menos próximas al mundo bibliotecario, y coloca a todos los sectores profesionales implicados en su desarrollo en una posición de igualdad.

En el informe de Biblink (Heery, 1996), el *metadato* se define como información sobre una publicación que complementa su contenido. No sólo incluye descripción bibliográfica, sino que también contiene información relevante como materias, precio, condiciones de uso, etc.

Ercegovac (1999) afirma que un *metadato* describe los atributos de un recurso, teniendo en cuenta que el mismo puede consistir en un objeto bibliográfico, registros e inventarios archivísticos, objetos geoespaciales, recursos visuales y de museos o implementaciones de software. Aunque puedan presentar diferentes

niveles de especificidad o estructura, el objetivo principal es el mismo: describir, identificar y definir un recurso para recuperar, filtrar, informar sobre condiciones de uso, autenticación y evaluación, preservación e interoperatividad.

Algunos autores como Sheldon (2001) proponen la definición “informaciones sobre datos”, Steinacker (2001) “datos sobre informaciones”, o Swick (2002) “informaciones sobre informaciones”.

Otros autores como Duval *et al.* (2002) y Woodley *et al.* (2003), que aportan definiciones del término *metadato* específicos para determinados tipos de recursos de información que se encuentran tanto en forma física como digital (archivos, museos, librerías), enfatizan en las funciones que desempeñan los metadatos y sugieren la definición “datos estructurados sobre datos” para ese término.

Desde el punto de vista de la IG el Recetario IDE (traducción del documento SDI CookBook v2), indica que el concepto “metadatos” también es familiar a la mayoría de aquéllos que manejan temas espaciales. Se indica, por ejemplo, que la leyenda de un mapa contiene, entre otras, información sobre el editor, fecha de publicación, tipo de mapa, descripción, referencias espaciales, escala y exactitud, etc. que son metadatos aplicados a la documentación de la información geoespacial. Se indica que en muchos ámbitos de la información se describe ésta de forma que se puede contestar a preguntas del tipo "qué", "quién", "dónde", "por qué", "cuándo" y "cómo", sobre los datos que describen y que la única diferencia importante respecto a ellos es el énfasis en el componente espacial -o el aspecto "dónde"-.

Finalmente para otro grupo de autores los metadatos pueden considerarse productos autónomos asociados a los conjuntos de datos geográficos que proporcionan información a los catálogos de datos, de servicios y las IDE, de modo que se posibilite el uso compartido de los datos, además de organizar y mantener un inventario de datos o proporcionar información que posibilite la transferencia y el uso de los datos (Phillips A. 1998; FGDC 2000; Najar C. 2006).

En esta tesis se ha adoptado la siguiente definición de metadato: “Conjunto estructurado de datos que describen a otros datos, a su estructura interna y a sus servicios, cuyo propósito es incrementar el conocimiento y contestar a preguntas del tipo ‘qué’, ‘quién’, ‘dónde’, ‘cuándo’, ‘cuánto’ y ‘cómo’. También pueden considerarse productos autónomos asociados a los datos que permiten mantener un inventario de los mismos, facilitar su publicación y consulta a través de catálogos en las IDE y facilitar la reutilización de los datos y la explotación de los servicios”.

### 2.3 Funciones que desempeñan los metadatos

Los metadatos de la IG pueden ser clasificados según el rol o la función que desempeñan. El resultado de la revisión, relacionada con las funciones propuestas por distintos autores, es coincidente en un alto porcentaje. En la tabla 2.1 se presentan por orden cronológico y por autores las funciones identificadas para los metadatos.

	Año	Buscar / localizar / descubrir	Recuperar / extraer / acceder	Transferir	Evaluar / explorar	Administrar	Describir / interpretar	Preservar	Usar / utilizar / explotar	Técnicos	Gestionar	Identificar
Beard	1996											
Gilliland-Swetland	2000											
Danko	2002											
Senso y Rosa Piñero	2003											
Jones y Taylor	2003											
ECNBII	2003											
ISO 19115	2003											
Nebert	2004											
Oosteron	2004											
Johnston	2005											
Ostensen <i>et al.</i>	2005											
Moellering and Brodeur	2006											
Gayatri y Ramachandran	2007											
Díaz <i>et al.</i>	2008											
		11	9	1	10	2	3	2	11	2	3	2

Tabla 2.1: Funciones que desempeñan los metadatos según los distintos autores y fechas

Una vez identificadas las distintas funciones, se presentan las definiciones que aportan los citados autores para describir las funciones de los metadatos de un

modo más detallado. El objetivo que se persigue es identificar el máximo número de funciones comunes localizadas en la bibliografía.

Funciones identificadas por Kate Beard (1996):

- Búsqueda: los metadatos deben proporcionar suficiente información, bien para descubrir si existen datos de interés dentro de la colección de datos disponibles, o simplemente, para saber que existen.
- Recuperación: los metadatos deben proporcionar información a los usuarios para que puedan adquirir la que sea de su interés. La analogía con una biblioteca consistiría en el procedimiento a seguir para sacar un libro. El componente que recupera los datos desde el metadato puede ser tan simple como proporcionar un URL (*Universal Resource Locator*) que identifique la localización de un conjunto de datos digitales, y tan complejo como para cubrir cuestiones de seguridad o realizar una transacción financiera para poder acceder a la información. En este sentido, también se considera la “función recuperación” a aquella información que describe cómo localizar fuera de línea los datos, la persona de contacto, los formatos de distribución de los datos o cualquier restricción de acceso a los datos, así como la información sobre los costes.
- Transferencia: los metadatos deben facilitar la información necesaria para que los usuarios hagan uso de los archivos recuperados en sus máquinas. Este componente incluiría información sobre el tamaño del conjunto de datos (y sus metadatos), la estructura tanto lógica como física de los datos y metadatos.
- Evaluación: esta función de los metadatos es quizá la más compleja. Los metadatos de apoyo a la evaluación pueden consistir en cualquier información que asista a los usuarios a determinar si los datos van a ser útiles para una aplicación.

Para Gilliland-Swetland (2000) se pueden diferenciar cinco tipos de metadatos: administrativos, descriptivos, de preservación, de uso y técnicos. El uso comprende la propiedad intelectual; los tres últimos términos suelen englobarse en administrativo y se suele añadir el tipo estructural. Los metadatos descriptivos suelen identificarse como los que facilitan la localización (Caplan 2003). Éste afirma que los metadatos describen las características y las funcionalidades de los

objetos a lo largo del ciclo de vida, desde la creación, las distintas versiones, continuando con la organización (registro, catálogo e indexación), búsqueda y recuperación, uso (reproducción, modificación) y preservación (refresco, migración, chequeo de integridad), finalizando con su posible eliminación.

Senso y Rosa Piñero (2003), además de realizar una revisión bibliográfica relacionada con la definición del término y las funciones que desempeñan los metadatos, entre las que destacan la taxonomía de Iannella y Waugh (1997), proponen las siguientes funciones: de uso, técnica, de preservación, descriptiva y administrativa. Las definiciones propuestas para estas funciones son:

- Uso: Relativos al nivel y tipo de uso que se hace con los recursos informáticos.
- Técnica: Relativos a cómo funcionan los sistemas o el comportamiento de los metadatos.
- Preservación: Para salvaguardar los recursos de la información.
- Descriptiva: Utilizados para representar recursos de la información.
- Administrativa: Usados en la gestión y administración de recursos de la información.

Jones y Taylor (2003) proponen las siguientes definiciones para las funciones de los metadatos:

- Búsqueda en la Web (*discovery*): Los metadatos de búsqueda en la Web proporcionan suficiente información para permitir discernir el contenido, formato y alcance de un conjunto de datos. La información responde en términos generales a las cuestiones “qué”, “quién”, “dónde”, “cómo” y “cuándo”, permitiendo decidir si el conjunto de datos es potencialmente útil. Para poder explotar esta funcionalidad de los metadatos, se necesita un sitio en el que realizar las búsquedas y localizar los datos existentes.
- Exploración: Una vez localizado un conjunto de datos, se tiene que evaluar su idoneidad para satisfacer los requisitos. Esta *adecuación al objetivo* es una función de la calidad de los datos. La Asociación de Información Geográfica (AGI) ha elaborado un conjunto de directrices para describir el contenido y la calidad de la información geográfica (Parker *et al.* 1996), e identifica cinco aspectos para la misma: completitud, precisión temática, precisión temporal,

precisión posicional y consistencia lógica. Los metadatos de exploración deben contener descripciones suficientemente detalladas de los cinco aspectos de la calidad de los datos como para permitir la evaluación de la idoneidad de un conjunto de datos.

- Explotación: El tercer nivel de los metadatos se relaciona con el proceso de obtención y uso de un conjunto de datos. Estos pueden contener información relacionada con la fuente de los datos y las restricciones de uso (Parker *et al.* 1996). Detalles técnicos, como el formato de los datos, orientarán al usuario en la selección de los conjuntos de datos que sean compatibles con su sistema.

ECNBII (Infraestructura de información biológica nacional de medioambiente Canadá, 2003) propone las funciones de:

- Búsqueda en la Web: La mayoría de las partidas de información medioambiental se localizarán en este nivel que es el más elemental, ya se trate de una colección de objetos simples o una base de datos.
- Acceso: Utilizando el perfil geoespacial completo y/o el perfil biológico, este nivel asegurará la descripción detallada y la localización de los datos.
- Uso: Este nivel permitirá el uso de los metadatos biológicos o geoespaciales a nivel de puesto de trabajo, para su visualización y extracción por medio de servicios Web.

Nebert (2004) plantea las funciones como las respuestas a cuestiones para las que el usuario necesita respuestas:

- Metadatos de localización - ¿Qué conjuntos de datos contienen el tipo de datos en el que estoy interesado? Esto permite a las organizaciones conocer y hacer pública la cartera de datos disponibles.
- Metadatos de exploración - ¿Contienen los datos identificados suficiente información como para permitir hacer un análisis razonable según los objetivos? Ésta es la documentación que ha de ser facilitada con los datos para garantizar que otros utilicen los datos correcta y juiciosamente.
- Metadatos de explotación - ¿Cuál es el proceso de obtención y utilización de los datos que se requieren? Esto ayuda a los usuarios finales y a las organizaciones proveedoras a almacenar eficientemente, reutilizar, mantener y archivar sus colecciones de datos.

Danko (2004), Oosteron (2004), Ostensen y Danko (2005) abordan las funciones de los metadatos desde el punto de vista de los elementos añadidos a los metadatos para permitir desarrollar las funciones:

- Localizar: Son ítems de metadatos que permiten a los usuarios localizar la IG que buscan y a los productores publicitar sus datos. Permiten que las organizaciones localicen datos externos y buscar socios con quienes compartir costes de captura y mantenimiento de la información. Además estos ítems facilitan la gestión, almacenamiento, recuperación y reutilización de los datos.
- Evaluar: Una vez localizados los datos, se necesitan otros ítems de metadatos que permitan determinar si los datos se adecuan al uso que pretende dar. Entre los ítems que permiten evaluar los datos están: la calidad y exactitud, el esquema espacial y temporal, el contenido y la definición de las entidades geográficas y los sistemas de referencia espacial utilizados.
- Extraer: En muchas ocasiones los usuarios necesitan acceder a los datos una vez localizados y evaluada su adecuación. Los ítems de metadatos que permiten extraer los datos permiten conocer dónde está localizado un conjunto de datos, su tamaño, formato, medios, precio y restricciones de uso.
- Emplear: Una vez descargados los datos los usuarios necesitan conocer como manejar los datos. Por esta razón se incluyen algunos ítems de metadatos adicionales que describen, entre otros, cómo fusionar y combinar los datos con los suyos propios, aplicarlos correctamente y comprender plenamente sus propiedades y limitaciones.

Johnston (2005) propone los siguientes fines para las funciones:

- Localizar: encontrar los recursos de interés, utilizando las descripciones de los recursos que han sido publicadas por sus creadores, distribuidores o terceras partes.
- Evaluar: deben permitir al usuario evaluar la pertinencia del recurso.
- Acceder: deben proporcionar información sobre los mecanismos necesarios para acceder a los recursos una vez que han sido evaluados.
- Interpretar: pueden ayudar a un usuario a interpretar los datos.

- Gestionar: pueden ayudar en otras tareas de gestión a los responsables de los datos, permitiéndoles preservar toda la información relacionada con la propiedad intelectual y los derechos.

Moellering y Brodeur (2006):

- Localizar: encontrar la ubicación de un conjunto de datos geográficos en referencia a un conjunto específico de características, por ejemplo topografía de un área; en muchos casos, este proceso de localización tiene lugar en la Web, en un entorno de red IDE.
- Evaluar: asegurar que los datos geográficos de la base de datos espacial tienen las características deseadas por el usuario: precisión, vigencia, etc.
- Extraer: transferir la base de datos espacial desde su ubicación, habitualmente a través de la Web, a una localización adecuada para el usuario de los datos.
- Emplear (aplicar): utilizar los metadatos para procesar con éxito la base de datos geográfica, analizar, y quizá resolver, el problema que se tiene entre manos.

Gayatri y Ramachandran (2007) proponen:

- Encontrar / localizar y acceder a recursos: Los metadatos ayudan a localizar o descubrir información relevante de acuerdo con los criterios dados. También pueden ayudar a filtrar tipos similares de recursos y separar recursos diferentes; ayudan en las búsquedas específicas. Asimismo pueden descubrir información sobre la localización de los recursos. En algunas ocasiones, facilitan una vista previa de datos como muestra.
- Identificación digital: Los identificadores digitales como el nombre del archivo URL y DOI (Identificador Digital de Objeto), que forman parte de los elementos de los metadatos, ayudan en la identificación de los recursos.
- Gestión y organización de los recursos: Los metadatos ayudan en la organización de varios vínculos asociados con los recursos, basándose en la petición del cliente. Estos recursos se crean dinámicamente a partir de las bases de datos de los metadatos. Esto ayuda a navegar fácilmente a través de la información obtenida.
- Interoperabilidad: los metadatos facilitan la interoperabilidad, puesto que se han definido estándares de metadatos y existen protocolos compartidos, el

descubrimiento de recursos de información ha sido integrado perfectamente, sin fisuras. Protocolos como el Z39.50 han ayudado en las búsquedas simultáneamente ante sistemas distribuidos. También el protocolo desarrollado para la recolección de metadatos de la Iniciativa de Archivos Abiertos (OAI-PMH) ayuda en la tarea de recolección.

- Archivo y conservación: Los objetos digitales se corrompen y pueden alterarse fácilmente, de ahí la necesidad de su conservación. Los metadatos son una pieza clave para garantizar que los recursos sobrevivan y continúen siendo accesibles en el futuro (NISO, 2004).

Para finalizar esta revisión, Díaz, *et al.* (2008) identifican 3 roles básicos para los metadatos: localización, evaluación y facilitación del uso de los recursos.

En esta tesis se han adoptado como funciones que desempeñan los metadatos: la *localización* de recursos, la *evaluación* de su idoneidad para los propósitos específicos, el *acceso* u obtención de los datos y, finalmente, el *uso* de los mismos. Se han seleccionado éstas por responder a las necesidades y objetivos, tanto de usuarios como de aplicaciones finales que desean usar los datos y los servicios además de ser las más importantes según la revisión literatura realizada.

## **2.4 Taxonomía de los metadatos**

Distintos autores clasifican los metadatos siguiendo diferentes criterios. A continuación, se enumeran las distintas taxonomías de metadatos encontradas en la literatura, para posteriormente sintetizar y estructurar dichas clasificaciones.

Jokela (2001) entiende por metadatos *implícitos* aquellos que están fuertemente ligados a los datos. Pueden ser: esenciales, aquéllos que son necesarios para poder usar los datos, (i.e. n° de filas, columnas y bandas de una imagen, tipo de compresión de los datos), o no esenciales para el uso de los datos. Por el contrario, en el contexto de las bases de datos, Morgensten (1998) entiende por metadatos implícitos aquellos que no se declaran y que dependen de la interpretación que se le dé al contexto.

Balfanz (2002) sugiere otra interpretación de los conceptos implícito y explícito. Propone como ejemplo de *explícito*, el nombre del formato de almacenamiento de los datos, y como implícitos, aquellos que son parte del geodato (por ejemplo: tipos de geometrías almacenadas, recuento de filas y columnas) o pueden calcularse, como por ejemplo, la extensión geográfica.

Díaz *et al.* (2008a) interpretan como metadatos implícitos aquellos que pueden inferirse (obtener metadatos a partir de otros metadatos o de los propios datos), relacionando los métodos propuestos por Beard y Goodchild. Para Beard (1996), los metadatos inferidos son aquellos que se obtienen definiendo reglas lógicas que permiten deducir valores a partir de datos conocidos, como por ejemplo, si se está en una determinada región con un tipo de clima dado y las temperaturas son inferiores a los valores típicos se puede inferir que la fecha es invierno. Para Goodchild (2007), los metadatos inferidos pueden ser aquellos que se obtienen por técnicas de minería de datos (*data mining*) o de recuperación automática de los mismos.

Para Wilson (1998), los metadatos implícitos también pueden ser el conjunto de reglas subyacentes que le indican a las aplicaciones cómo manejar los datos. Éste puede ser el caso de la información y el conjunto de reglas implementadas en los SIG para gestionar los sistemas de referencia espacial por coordenadas.

Codd (1990), Korth y Silberschatz (1991) indican que los sistemas de gestión de las bases de datos (SGDB) disponen, desde hace mucho tiempo, de metadatos para describir la estructura interna de los esquemas de datos, pudiéndose considerar estos metadatos también implícitos.

Atendiendo al ciclo de vida de los metadatos, Jokela (2001) propone clasificar los metadatos como *estáticos* y *dinámicos*. Los estáticos son aquellos metadatos cuyo contenido no varía a lo largo de su vida y dinámicos, aquellos que cambian con el tiempo y que deben de ser refrescados o recalculados. Este autor también cita los metadatos *temporales*, como aquellos que se crean con algún fin en un determinado momento y, que trascurrida una ventana temporal, se destruyen.

Como ejemplo de metadatos *temporales* se citan el estado y la programación del flujo de trabajo de la información.

Desde el punto de vista de los roles que desempeñan los metadatos, según la clasificación de Boll *et al.* (1998) citada en Jokela (2001), se proponen tres categorías: *estructurales*, *de control* y *descriptivos*. Para NISO (2004) las anteriores categorías se denominan: *estructurales*, *descriptivos* y *administrativos*. Los metadatos *estructurales* están fuertemente relacionados con los esenciales o necesarios para poder utilizar la información. Los metadatos de *control* se utilizan para gestionar el flujo de los contenidos; son datos que describen si el contenido está preparado para la siguiente fase y, en algunos casos, pueden considerarse metadatos temporales (i.e. control de máquina, calidad del servicio y gestión de errores). Según Jokela (2001), los metadatos *descriptivos* se clasifican a su vez en: *contextuales* y *semánticos*, basados en el contexto. Estos metadatos contemplan los aspectos relacionados con la propiedad intelectual y los derechos y privilegios de acceso. Los metadatos *contextuales* se refieren a la información temporal y al sistema usado para gestionar información, mientras que los metadatos *semánticos* describen las cualidades semánticas de los contenidos, permitiendo responder a preguntas relativas al significado de las cosas (i.e. asunto, localización, nombres, palabras clave). Para NISO (2004), los *administrativos* se subdividen en metadatos para la gestión de la propiedad intelectual y para la preservación de la información.

Para Durval *et al.* (2002) los metadatos, desde el punto de vista de su almacenamiento o de la forma de acceso, pueden ser:

- Almacenados en el *propio recurso*: utilizando marcas se crean simultáneamente a los datos y añaden valor y visibilidad;
- Almacenados en *archivos acoplados* a los recursos: aportan las ventajas en la creación independiente de los datos y las desventajas en la gestión simultánea de los datos, y los archivos que almacenan los metadatos;
- Almacenados en un *repositorio independiente*: generalmente en una base de datos. Esto dificulta su consulta directa. También puede dificultar el mantenimiento si lo realiza una organización que no es la que tiene el control de los datos.

Durval *et al.* (2002) también distinguen entre metadatos *objetivos* y *subjetivos*. Los *objetivos* son aquellos que están relacionados con el autor, la fecha, etc., toda aquella información que almacenan algunas aplicaciones, como los procesadores de texto, en forma de propiedades asociadas a los archivos. Los *subjetivos* son aquellos que pueden ser interpretados desde distintos puntos de vista (palabras clave, resúmenes), y se hacen más subjetivos cuando éstos dependen del dominio, del contexto o de la cultura.

Desde el punto de vista de la creación de metadatos, los datos que una vez estructurados forman un metadato, pueden obtenerse de distintas formas que definen una clasificación o taxonomía de los mismos. En esta tesis se han considerado los siguientes tipos de metadatos: *implícitos*, como aquellos que están fuertemente ligados a los datos y a su uso; *explícitos*, como aquellos ligados al tipo de datos, al almacenamiento y al contexto; *calculados*, como aquellos que pueden obtenerse mediante algún tipo de cálculo o tratamiento; e *inferidos*, como aquellos que pueden obtenerse mediante reglas lógicas que permiten deducir valores a partir de otros valores; *contextuales*, como aquellos que se pueden obtener o imponer por el contexto en el que se crean: fecha, aplicación, máquina, etc. No se han considerado el resto de tipos por no estar relacionados con la creación automática de metadatos.

## **2.5 Procedimientos de creación de metadatos**

Dado que:

- La creación manual de metadatos es un proceso pesado y propenso a errores (Batcheller 2008, Wyoming, West y Hess 2002, Guptill 1999);
- Los procedimientos automáticos de creación de metadatos no pueden proporcionar la información que, tanto los productores de datos como los compiladores de metadatos, pueden aportar (Campbell 2008, Guy *et al.* 2004, JORUM 2004);
- La creación manual de metadatos, por parte del autor de los datos o la creación automática de los mismos no pueden aportar la experiencia en catalogación de los expertos en la gestión de la información (Currier *et al.* 2004, Guy *et al.* 2004, JORUM 2004);

es por lo tanto oportuno realizar una revisión de los procedimientos y metodologías de creación de metadatos (manuales, automáticas y mixtas) existentes, para, posteriormente, proponer una metodología que intente automatizar, en la medida de lo posible, la creación de metadatos en el dominio de la IG (geodatos).

Desde el punto de vista de la catalogación, Coleman (2002) distingue dos tipos de catalogación: descriptiva y temática (*subject*). Varios son los procedimientos y las metodologías de creación de metadatos propuestos en la literatura, implementados en los flujos de trabajo y de gestión de los metadatos y, por tanto, también los usados en las distintas disciplinas que precisan de los mismos.

Beard (1996), en un taller sobre metadatos ambientales, propuso cinco métodos para crear o compilar un metadato:

- a) Manualmente, por teclado (método tradicional),
- b) Ampliando la información almacenada con valores que se obtienen mediante una búsqueda en una tabla de referencia (p.e. conocida la extensión geográfica, buscar un identificador geográfico usando un nomenclátor),
- c) Mediante medidas y observaciones automatizadas,
- d) Extrayendo y calculando e
- e) Inferir nuevos metadatos a partir de otros elementos.

Los 5 métodos propuestos pueden ser combinados de la mejor forma posible para dar respuesta a las necesidades. Por ejemplo, si se trata de metadatos relativos a objetos en movimiento que recolectan información ambiental, además de la posición, pueden combinarse los métodos b) y c) para obtener cierta información estática y complementarla con la dinámica derivada del movimiento. Por el contrario, si se desea crear un metadato sobre un conjunto de datos geográficos existentes, la combinación de los métodos d) y e) puede ser el mejor modo para crear metadatos, como indica Beard (1996).

Bailer y Schallauer (1998) examinan las funciones de los metadatos en el proceso de producción de medios audiovisuales: preproducción (concepción del producto), producción (creación de contenidos), almacenamiento de datos y metadatos,

postproducción (edición modificación), explotación (entregar los recursos) y visualización.

Cuando los autores analizan los metadatos en el proceso de creación, identifican como posibles fuentes de información: a) la captura (mediante dispositivos que tomen datos del contexto: cámaras, GPS, etc.), b) la información legada de los materiales antiguos, c) las anotaciones manuales, d) la información que se extrae del contexto, y e) el análisis semántico de textos.

A continuación se describen las fuentes de información:

- En el proceso de captura, son importantes los metadatos que describen la configuración del dispositivo usado para captar la fotografía o escena, típicamente en Teledetección y fotogrametría. Otros metadatos que pueden capturarse son la fecha, hora y lugar (con coordenadas globales o relativas). Cuando lo que se está digitalizando es material analógico (por ejemplo un mapa), la información relativa a la calidad y los defectos del material pueden capturarse.
- La información legada entre los materiales antiguos, tales como resúmenes, comentarios, artículos de prensa, libros o documentales, pueden ser importantes de preservar o mantener en los metadatos.
- Las anotaciones manuales también son muy importantes, ya que proporcionan abstracción semántica del significado; sin embargo, es un proceso demasiado costoso. En algunas ocasiones se hace una descripción global del contenido.
- El análisis del contenido permite crear grandes cantidades de metadatos a bajo coste; sin embargo, los resultados logrados de extraer dicha información desde los niveles básicos y medios, no han sido suficientemente buenos. Este hecho se conoce como “brecha semántica”. Los resultados pueden ser mejores si se dispone de información del contexto.
- El análisis semántico del texto es la técnica comúnmente utilizada para extraer semántica. En este tipo se incluyen las referencias a entidades con nombre (personas, organizaciones, lugares), el enlace con otras entidades pertenecientes a ontologías, la detección de los temas junto a la clasificación de sus contenidos en segmentos y el enlace de los elementos con información heredada o relacionada.

Balfanz (2002) apuesta por la creación automática y la visualización de metadatos en el contexto geográfico. Para este autor, la creación de metadatos debe obedecer a unos objetivos y debe de medirse el grado de éxito del mismo. El tipo de metadatos que se pueden extraer y calcular automáticamente es el implícito, y los medios para obtenerlos pueden ser varios, cada uno de los cuales debe estar caracterizado por una estimación del tiempo requerido y un indicador de la calidad de los resultados. De este modo, el operador del sistema de creación y visualización de los metadatos puede seleccionar el método que mejor se acomoda en cada caso y para cada tipo de necesidad.

Jane Greenberg (2004) identifica dos métodos de creación automática de metadatos: la extracción y la recolección (*harvesting*). La extracción usa técnicas de minería de datos e indexación para recuperar ítems o etiquetar contenidos. En el *harvesting* se utilizan técnicas de recolección de contenidos etiquetados ya existentes. Posteriormente, Greenberg *et al.* (2006) hacen, aunque con tintes bibliotecarios, una revisión de las herramientas y las aplicaciones desarrolladas para la creación automática de metadatos sobre recursos electrónicos, y afirman que el uso de estos métodos automáticos permite orientar el esfuerzo de los recursos humanos a aspectos más intelectuales. Dependiendo de lo automático que sea el proceso de creación de metadatos o de los requisitos humanos del mismo, se distinguen entre generadores (*generators*) y editores de metadatos, en los que se integran los procesos automáticos y los humanos.

El proyecto *AMeGA* hizo una revisión de las limitaciones de las aplicaciones que generan metadatos, para proponer las funcionalidades deseables para las mismas. Este proyecto realizó una encuesta para conocer las opiniones del personal involucrado en los metadatos. Ésta sondeaba sobre la creación automática y las funcionalidades deseables de las aplicaciones que posibilitan su creación en el contexto del *Dublín Core Metadata Initiative* (DCMI). Las cuestiones planteadas en la encuesta estaban relacionadas con la opción más apropiada para cumplimentar los ítems de un metadato (DCMI: 15 ítems): automática, semi-automática y manual. Las conclusiones de este estudio apuestan por la investigación de métodos que automaticen la creación de metadatos, aunque también la mayoría de los encuestados concluyen que las herramientas deben

ayudar al operador humano y no sustituirlo. Las funcionalidades más demandadas por los encuestados fueron:

- Posibilidad de ejecutar los algoritmos automáticos, de modo que una persona pueda evaluar los resultados, y
- La integración de estándares de contenidos como son los tesauros.

En relación con los algoritmos de creación automática de metadatos, se citan como ejemplos las investigaciones en indexación automática, basadas en el uso de tesauros en dominios específicos como los conceptos médicos (e.g., Nadkarni *et al.* 2001), en la investigación para creación automática de resúmenes de textos (e.g., Johnson 1995) y en la investigación sobre clasificación automática (e.g., Losee 2003).

Los equipos de EDINA y MIMAS, que han participado en el proyecto *JORUM* (2004) encaminado a la elaboración de metadatos sobre recursos para los entornos educativos electrónicos (e-learning), identifican varios modelos para la creación de metadatos:

- Que el autor cree el metadato.
- Que cree el metadato un especialista en información.
- Que se use una herramienta colaborativa, de modo que el autor incluya cierta información y el especialista en información otra.
- Que cree el metadato otra persona y que, antes de ser publicado, alguien perteneciente al proyecto lo valide.

Cada uno de los modelos tiene sus ventajas e inconvenientes razón por la que dichos grupos proponen el uso de modelos híbridos.

Guy *et al.* (2004), en el marco del proyecto *ARIADNE*, proponen un flujo de trabajo para la producción de metadatos, en el que se prevé la creación totalmente automática, semi-automática, manual, por el autor de los datos y, finalmente, por un especialista en gestión de información, como puede observarse en la figura 2.1. Dicho flujo de trabajo prevé todas las posibilidades: a) creación automatizada, b) creación automatizada mejorada por el autor de los datos, c) creación automatizada mejorada por un autor y por un especialista en información, d)

creación manual por el autor y mejorada por un especialista en información o e) creación por un especialista en gestión de información.

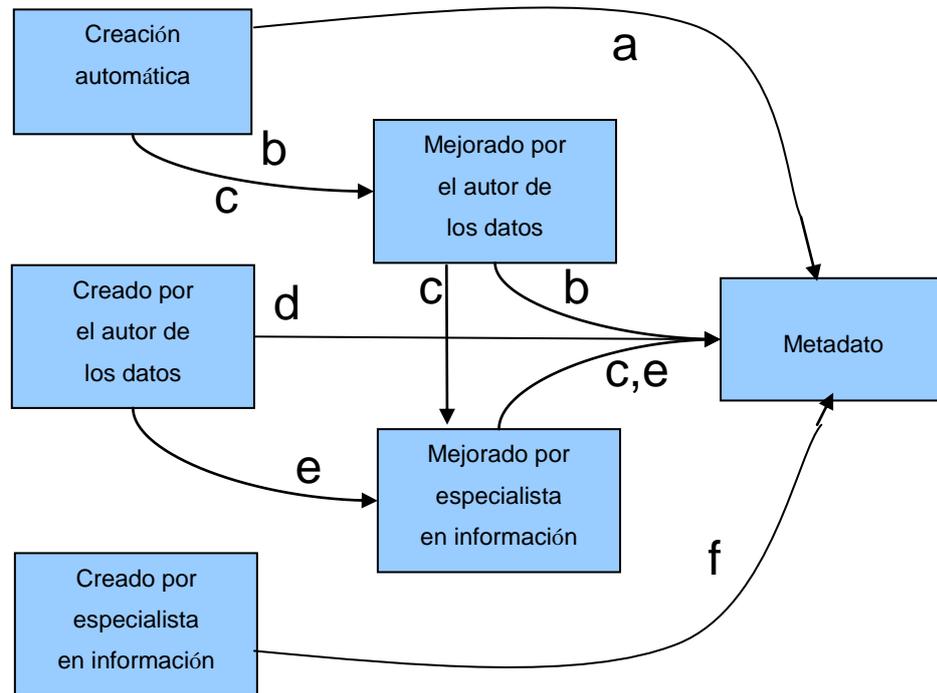


Figura 2. 1: Posibles métodos de creación de metadatos (Adaptado proyecto ARIADNE)

Baird (2006), en su revisión de los métodos que permiten automatizar el flujo de trabajo de los metadatos existentes o potenciales, identifica tres estados en el flujo: contribución, catálogo y revisión. El proyecto *Jorum*, en la revisión del flujo de trabajo de agosto de 2006, complementa dichos estados con un cuarto: publicación. Se trata inicialmente de un método de creación de metadatos manual, en el que la mayoría de los metadatos son proporcionados por el creador de los datos y son puestos en manos del flujo que los gestiona. Los metadatos pueden ser publicados, aunque con ciertas reservas, a la espera de que el catalogador los revise y los valide. Terminada esta tarea, los metadatos son válidos hasta que se revisen o hasta que cambien los datos que describen. Una situación similar proponen Currier *et al.* (2004) como proceso de recolección de metadatos en el contexto de los contenidos electrónicos educativos: el que crea los contenidos crea los metadatos y el especialista en gestión de la información los revisa, los complementa y los cataloga. Baird (2006) propone que ciertas operaciones de los catalogadores, e incluso de la creación de los contenidos, sean automatizadas. Se puede decir que la propuesta de flujo de trabajo con los metadatos es un proceso colaborativo.

Morris *et al.* (2007), en el proyecto de almacenamiento de datos geospaciales de Carolina del norte (NCGDAP), han propuesto un marco para la creación de metadatos (figura 2.2), en el que se parte con la definición de una plantilla o perfil (a) para la organización, para después, b) personalizarla para una colección de datos geográficos concreta. Definida la plantilla para un conjunto de datos, si existe el metadato, c) se procesa para adaptarlo a la plantilla y si no, d) se crea. A continuación e) se añade toda la información de linaje y f) finalmente, se aplica un proceso de sincronización con una herramienta de extracción de metadatos comercial. Otros autores como Hedorfer y Bianchin (1999) también sugieren el uso de plantillas como mecanismo que facilitan la creación de metadatos.

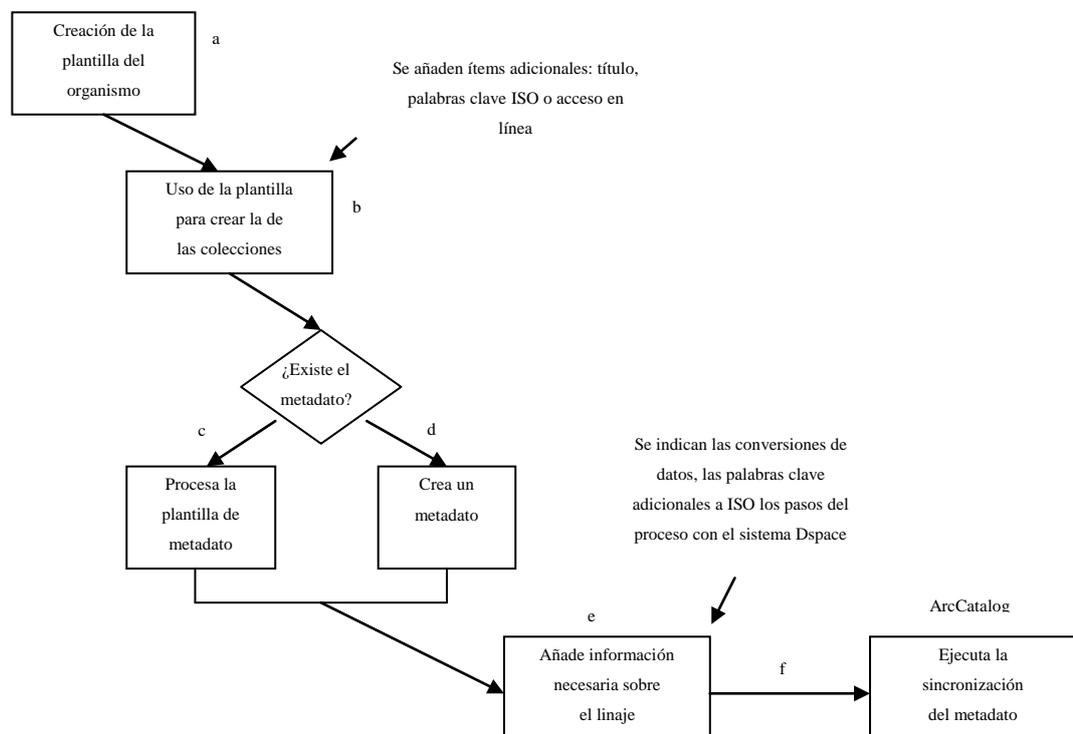


Figura 2. 2 Flujo de creación de metadatos (Fuente: proyecto NCGDAP)

Mustacoglu (2007), en una propuesta de trabajo de investigación, hace una breve revisión de los entornos colaborativos en los que se precisa tener descritos los recursos mediante metadatos. Se mencionan entornos tales como *YouTube*, *LibraryThing*, *Flickr* y *43things*, que utilizan profusamente herramientas de marcado mediante anotaciones (*tagging*). También menciona otros entornos como *CiteULike*, *Connotea*, *Bibsonomy* o los sistemas *Delicious* (*del.icio.us*), tratándolos como herramientas que facilitan el uso compartido de anotaciones sobre los recursos y sus enlaces (*URL*). Atendiendo a la creación de metadatos, el autor identifica dos mecanismos: metadatos profesionales (especializados), que

trabajan con esquemas de catalogación complejos y los metadatos que crean los propios autores de los contenidos, aunque sean noveles. En el primer caso, se cita como problema de falta de escalabilidad de los sistemas de catalogación cuando se tienen que aplicar a las grandes cantidades de datos en la Web. Los metadatos creados por los autores son vulnerables, ya que las descripciones propuestas no suelen ser adecuadas. Estas propuestas de colaboración se las denomina *folksonomy* y crean metadatos mediante marcas (*tagging*), depositando toda la responsabilidad de los metadatos en los propios usuarios.

Algunas de las ventajas de este tipo de técnicas son: a) facilidad de acceso y uso de las herramientas; b) facilidad de descubrimiento de nuevos contenidos; y c) soporte a la creación de nuevos nichos de comunidades. Como inconvenientes se citan: a) la ausencia de estándares para usar palabras clave; b) la dificultad para encontrar erratas, detectar sinónimos y acrónimos; y c) la dificultad para usar relaciones jerárquicas de inferencia entre las etiquetas, al no existir taxonomías.

Batcheller (2008) propone la personalización de herramientas comerciales para generar metadatos de un modo más eficiente, que desempeñen las siguientes funciones: a) recolectar los ítems de metadatos existentes; b) extraer los ítems de metadatos implícitos y almacenados junto a los datos; c) recolectar las plantillas de metadatos preparadas por el responsable; d) integrar los anteriores datos posibilitando su edición y visualización; y e) facilitar las herramientas de importación y exportación en formatos estandarizados.

Campbell (2008) propone una metodología de creación de metadatos basada en una herramienta comercial que contempla tres fases: a) extracción, b) investigación y c) compilación y exportación:

- La extracción requiere que los metadatos ya existan antes de que sean recuperados y puestos en archivos con formatos apropiados, por ejemplo en documentos Word o en hojas de cálculo.
- La investigación requiere obtener la información que no existe y, para ello, se deben de realizar entrevistas personales con quienes crearon los datos o con personas que estaban en la organización en el momento de crear los metadatos.

- La compilación y exportación han de tomar la información recuperada por las fases anteriores y plasmarla en el documento de la forma apropiada; además, se ha de comprobar la consistencia y completitud de los metadatos para todos los conjuntos de datos.

Aunque no lo menciona explícitamente en las fases de creación de los metadatos, la extracción automática de ciertos ítems se realiza con la herramienta comercial ArcCatalog.

## **2.6 Sumario**

Como se ha descrito anteriormente, los metadatos se utilizan para describir el contexto, la condición y las características de los datos y para que los usuarios puedan localizarlos y entenderlos. Ya en la definición, se evidenciaba que los metadatos desempeñan distintas funciones. Una vez identificadas las principales funciones que para los distintos autores desempeñan los metadatos, se procede a hacer una síntesis de las mismas:

- Descubrir/localizar: se trata de un aspecto primordial compartido por los autores, aunque con distinta denominación. El objetivo es encontrar aquellos datos que sean de interés. Se materializa en los ítems de metadatos que permiten a los usuarios de la información geográfica localizarla y a los productores publicitar sus datos.
- Evaluar: los metadatos además de ayudar a realizar las búsquedas, deben ayudar a los usuarios a determinar si los datos descritos por los mismos son aptos cubrir sus expectativas con su uso.
- Acceder a los datos: tras encontrar y evaluar la aplicabilidad de los datos el tercer paso suele consistir en transferir los mismos para explotarlos conjuntamente con los propios o con otros obtenidos por este procedimiento, o directamente en acceder o establecer la conexión para explotarlos en línea. Los usuarios precisan de ítems de información en los metadatos que describan la forma de acceder, o transferir, los datos para su utilización.
- Utilizar: el fin último del proceso de búsqueda es el uso de los datos. Los metadatos deben contener ítems de información que permitan identificar el uso más adecuado de los datos, fusionándolos y combinándolos

adecuadamente, para lo cual se requiere un pleno conocimiento de sus propiedades, limitaciones y restricciones de uso.

- Preservar: un menor número de autores identifican las funciones de gestión, de almacenamiento y de preservación. El objetivo de los ítems de metadatos que satisfacen estas funciones permiten organizar y mantener un inventario de los datos, de modo que se preserve el conocimiento y proporciona soporte a las organizaciones tanto para dirimir conflictos derivados de un uso inadecuado de los datos, como para salvaguardar los recursos de información.

También se ha evidenciado, tanto en las definiciones del término metadato como en las funciones que éstos desempeñan, que parte de esta información que describen los datos puede estar almacenada o puede obtenerse de diferentes formas. Como síntesis del análisis realizado sobre las taxonomías de los metadatos se puede concluir que, los metadatos pueden ser clasificados siguiendo los siguientes criterios:

- Desde el punto de vista de su existencia, como *implícitos* y *explícitos*;
- Teniendo en cuenta su ciclo de vida, como *estáticos*, *dinámicos* y *temporales*;
- Dependiendo del rol que cumplen, como *estructurales*, de *control* y *descriptivos*;
- De acuerdo a su interpretación, como *objetivos* y *subjetivos*;
- En función de su almacenamiento (o de su acceso), como *embebidos*, *archivos asociados* y *en repositorios externos*.

## **3 INTEROPERABILIDAD**

### **3.1 Introducción**

Para poder disponer de una percepción conjunta y general del significado del término, a continuación se sintetizan las distintas definiciones de interoperabilidad propuestas por un conjunto de autores y organizaciones. Dado que la interoperabilidad en muchos casos se trata de un modo estructurado, varias organizaciones e iniciativas han definido diferentes modelos de la misma. Estos modelos de interoperabilidad, así como las definiciones, modelan la interoperabilidad en niveles. Finalmente se revisan y analizan los niveles de interoperabilidad de forma individualizada. El objetivo de la definición de los distintos niveles es modular el problema para poder solucionarlo.

Para analizar el grado de interoperabilidad alcanzado se pueden definir medidas de interoperabilidad. Estas medidas permiten detectar las debilidades y las fortalezas, si bien se han de definir métricas que lo permitan medir.

Los objetivos que se pretenden alcanzar con este conjunto de revisiones se concretan en:

- Definir el término interoperabilidad
- Analizar los modelos de interoperabilidad existentes
- Analizar los niveles de interoperabilidad identificados
- Realizar una revisión concisa de las medidas de interoperabilidad

Este estado del arte servirá para proponer un modelo de interoperabilidad válido y aplicable en el contexto de las Infraestructuras de Datos Espaciales que responda a la primera cuestión de investigación planteada: ¿Se puede formalizar un modelo de interoperabilidad de sistemas para las infraestructuras de datos espaciales?

### **3.2 Importancia de la interoperabilidad para las IDE**

Las IDE son iniciativas institucionales que facilitan el acceso a la IG a través de la Internet, y comprenden aspectos de organización, culturales, políticos y tecnológicos. Esta infraestructura se puede definir como el conjunto de las reglas,

las normas, los procedimientos, las directrices, las instrucciones, las políticas y la tecnología que permiten crear, recoger, procesar, almacenar, mantener, intercambiar, acceder y utilizar datos espaciales (Crompvoets y Bregt, 2003 ). Algunos autores y usuarios de la IG, consideran a las IDE como el avance tecnológico introducido en los Sistemas de Información Geográficos (SIG) tradicionales, que les permite acceder, a través de aplicaciones Web distribuidas, a la IG (Najar, 2006). Actualmente, las IDE pueden considerarse un proceso que comprende los aspectos antes mencionados, que pueden llevarse a cabo de distintas formas, a distintas velocidades, con distintos costes y con distintos impactos (Longhorn, 2008).

Los SIG han evolucionado tecnológicamente (Dangermond 1991, Longley et al. 2000), y cada vez se encuentran más empresas que desarrollan o construyen aplicaciones informáticas en estos entornos a través de Internet. Una consecuencia de este crecimiento es que aumentó el número de sistemas y formatos de datos en uso, aunque los usuarios pueden no ser conscientes de esa multiplicidad y heterogeneidad. Los organismos de normalización y estandarización implicados en las IDE (*ISO, OGC*), tienen como principal objetivo la interoperabilidad en sus diferentes connotaciones.

El término interoperabilidad tiene muchas connotaciones, incluyendo los objetivos de la comunicación, el intercambio de información, la cooperación y el uso compartido de recursos entre distintos tipos de sistemas. De hecho, la esencia de la interoperabilidad es garantizar las relaciones entre los sistemas, donde cada relación es una forma de compartir, de comunicar, de intercambiar y de cooperar (Carney *et al.* 2005). Nuestra atención se centra en las IDE, en las que las tecnologías, sistemas, redes, estándares y normas, los datos, las personas, las políticas, los acuerdos, los aspectos organizacionales y finalmente, los mecanismos de entrega de datos a los usuarios finales (GSDI 2004 p. 8, Georgiadou *et al.* 2005, Williamson 2004), deben de facilitar la localización, la evaluación, el acceso y el uso de la IG de un modo transparente para los usuarios, ya sean agentes humanos o aplicaciones informáticas, siendo este último caso en el que la interoperabilidad es un requisito más restrictivo.

En la literatura se pueden encontrar diferentes definiciones del término interoperabilidad que, por lo general, difieren en la descripción de las relaciones y de los componentes del sistema. Por otro lado, Georgiadou *et al.* (2005), de acuerdo con Bernard *et al.* (2005) afirman que las IDE son un caso especial de las Infraestructuras de Información (II) específicamente orientado a la IG. Béjar *et al.* (2008), de acuerdo con Maier (1996), proponen otros marcos de apoyo a la investigación para las IDE considerándolas componentes de los Sistemas de Sistemas (*SoS*). La principal conclusión de la última propuesta es que los conceptos analizados son similares, aunque analizados desde distintas perspectivas. Si bien se postula que el marco de referencia de *SoS* es un concepto más amplio que el de las II. En este marco de referencia es en el que se va a proceder a analizar la Interoperabilidad para las IDE. Finalmente, William (2002) y Gordon (2003) proponen que la interoperabilidad puede analizarse desde distintos puntos de vista: el de los datos, los servicios, las aplicaciones y las organizaciones. Para otras organizaciones como ISO (ISO19191:2002) o autores como Tolk (2003) y Turnitsa (2006), la interoperabilidad puede ser analizada a distintos niveles: tecnológico, sintáctico y semántico.

### **3.3 Definiciones de interoperabilidad**

El Instituto de ingeniería eléctrica y electrónica (IEEE 1990) define interoperabilidad como la capacidad de dos o más sistemas, o componentes, para intercambiar información y usar la información intercambiada.

Las agencias y organismos de seguridad *DoD*, *NATO*, *ADF C&C Information Systems Plan* (1995/6) describen interoperabilidad como “*la capacidad de los sistemas, unidades o fuerzas para proporcionar servicios y aceptar servicios de otros sistemas, unidades o fuerzas, intercambiados de modo que puedan operar conjuntamente de un modo eficiente, sin alterar o degradar la información*” (*glossary of the C2S Study Phase 1 report*) (C2SS WG 1996). El mismo estudio aporta otras dos definiciones más simples: la capacidad de una entidad para servir a otra y la necesidad de un grupo de interactuar de algún modo con otro.

El *Open Geospatial Consortium* (OGC 1998), en el documento “*OpenGIS guide*” define interoperabilidad en el contexto de las especificaciones *OGC*, como componentes de aplicaciones informáticas trabajando recíprocamente unos con otros, para evitar pesadas y sistemáticas tareas de conversión, evitar los obstáculos de las importaciones y exportaciones de datos y las barreras de acceso a los recursos distribuidos, impuestas por los entornos de procesamiento y por la heterogeneidad de los datos.

Miller (2000) entiende la interoperabilidad como un proceso en el que se encuentran embarcadas las organizaciones, para asegurar que los sistemas, los procesos y la propia cultura de la organización sean gestionados óptimamente, para que se maximicen las oportunidades de intercambiar y reutilizar la información, tanto interna como externamente.

Flater (2002), en el documento *el impacto de los estándares orientados por modelos*, propone una definición de interoperabilidad basada en dos conceptos: “compatibilidad” e “adaptabilidad” (*integrability*). Interoperabilidad es el proceso que persigue hacer integrables los sistemas para que éstos sean compatibles. Se entiende por compatibilidad a la cualidad de los sistemas para cooperar o coexistir sin interferir unos con otros. Se entiende por “adaptabilidad” la capacidad de adaptar posibles sistemas incompatibles, o los datos que intercambian, para que cooperen.

El organismo internacional de normalización (*ISO*), a través de su comité técnico 211 (TC211) define, en la norma ISO 19101 (2002), el modelo de referencia para toda la familia de normas ISO19100, y aporta la siguiente definición de interoperabilidad: interoperabilidad es la capacidad de los sistemas o componentes de intercambiar información y de poder garantizar el procesamiento cooperativo entre aplicaciones. La interoperabilidad hace referencia a las siguientes habilidades:

- Capacidad de encontrar información y herramientas de procesamiento cuando sea necesario, e independientemente de su localización física;

- Entender y usar la información y las herramientas descubiertas sin limitaciones, debidas al tipo de plataformas a utilizar, ya sean en contextos locales o remotos;
- Poder desarrollar entornos de procesamiento para uso comercial sin imponer limitaciones en el mercado por monopolio;
- Poder basarse en la información y el procesamiento ofrecido por las infraestructuras de terceras partes, permitiendo las necesidades de los distintos nichos de mercado, sin temor al fracaso cuando la infraestructura de apoyo madure y evolucione.
- Participar en un mercado libre y transparente, en el que los bienes y servicios respondan a las necesidades de los consumidores y que los canales se abran a medida que el mercado crezca lo suficiente para apoyarlos.

La misma norma ISO19101 describe los aspectos de la interoperabilidad entre sistemas, y enumera los siguientes niveles de interoperabilidad: protocolos de comunicación en redes, sistema de ficheros, llamadas a procedimientos remotos (RPC) y búsqueda y acceso a las bases de datos. Finalmente, la norma menciona dos clases de interoperabilidad: sintáctica y semántica.

Para Gordon (2003), la interoperabilidad es la medida de cómo las aplicaciones, los datos y las soluciones que residen en lugares diferentes unos de otros, son capaces de compartir información y funcionalidades de un modo correcto y proporcionando más valor que un único producto, en el que todos ellos se encontraran integrados.

Rawat (2003) propone el concepto de interoperabilidad en el dominio de los SIG, como la capacidad para intercambiar IG y datos procedentes de distintas organizaciones, para que la sociedad pueda aplicarlos, a través de las redes, a cualquier tipo de aplicación.

La asociación de bibliotecas Americanas (ALCTS 2004), define interoperabilidad como la capacidad de dos o más sistemas o componentes para intercambiar y utilizar la información intercambiada, sin que ello suponga un esfuerzo especial para ninguno de los dos.

Para *NISO* (2004) (organismo nacional de normalización de la información norteamericano), la interoperabilidad es la capacidad de intercambiar y compartir datos entre distintos sistemas que utilizan diferentes arquitecturas de equipos y programas informáticos, distintas interfaces y estructuras de datos.

Taylor (2004) afirma que dos sistemas son compatibles cuando ambos pueden intercambiar información y usarla sin necesidad de un tratamiento especial, acuñando el concepto “compatibilidad”.

La agencia para la organización de las tecnologías de la información “*Police Information Technology Organisation*” (PITO) define interoperabilidad como la capacidad de dos sistemas, o componentes, para intercambiar y usar la información, y la capacidad de los mismos para proporcionar o recibir servicio de otros sistemas, de modo que utilicen los servicios intercambiados, para operar conjuntamente de un modo efectivo (ALCTS 2004).

El glosario de *Dublin Core Metadata Initiative* (DCMI) (Woodley *et al.* 2005) define interoperabilidad como la capacidad de los distintos tipos de ordenadores, redes, sistemas operativos y aplicaciones informáticas, para trabajar conjuntamente de una forma efectiva, sin requerir comunicaciones previas, de modo que se intercambie información de un modo útil y válido. Finaliza identificando tres aspectos de la interoperabilidad: sintáctica, estructural y semántica.

*ISO*, en su norma ISO19119 (2005), propone una definición de interoperabilidad que se aplica a todo tipo de información relacionada con el espacio y los datos geográficos. “*Interoperabilidad geográfica es la capacidad de los sistemas de información para intercambiar libremente todo tipo de información espacial relacionada con la tierra, los objetos y los fenómenos que existen encima, debajo y sobre la superficie terrestre, y ejecutar programas capaces de manejar dicha información de forma cooperativa, sobre redes de comunicaciones*”.

Carney *et al.* (2005) amplían las definiciones anteriores centrándose en la finalidad (el objetivo para la interoperabilidad) y el contexto (el entorno en el que el sistema existe).

Para esta tesis se ha adoptado como definición del concepto interoperabilidad: la capacidad de una colección de componentes de un sistema, o un sistema en un sistema de sistemas, para intercambiar y compartir información específica y de tratar dicha información de acuerdo con una semántica compartida con el fin de lograr un objetivo concreto en un contexto dado. En el contexto de las IDE, la interoperabilidad debe posibilitar que usuarios y sistemas puedan compartir datos y servicios intercambiando información, de acuerdo con un conjunto de reglas semánticas compartidas según los distintos contextos y propósitos.

Esta definición motiva la concreción de modelos de interoperabilidad, con el objeto de garantizar que se produce la interoperabilidad entre los sistemas, de acuerdo a los distintos propósitos y contextos.

### **3.4 Modelos de interoperabilidad**

Los esfuerzos realizados en el pasado y en la actualidad para crear sistemas informáticos interoperables, han dado lugar a los modelos que ayudan a solucionar la heterogeneidad sintáctica, semántica y estructural de datos, interfaces de servicios y meta modelos que los describen (Lemmens 2006).

La formulación de modelos de interoperabilidad tiene sus orígenes en los sistemas de información y en la necesidad de integrar dichos sistemas. NATO (2004), en el documento *NATO C3 System Architecture Framework (NC3SAF)*, define un primer modelo de interoperabilidad en el que se proponen las directrices generales para el desarrollo de arquitecturas de sistemas. Posteriormente, el *SEI (Software Engineering Institute, Carnegie Mellon University)* ha definido varios modelos de interoperabilidad entre sistemas, entre los que destacan: *LISI (Levels of Information System Interoperability)*, *LCIM (Levels of Conceptual Interoperability)*, *SoSI (System of System Interoperability)* y *LCI (Levels of Coalition Interoperability)*, entre otros (Tolk 2003).

Desde el punto de vista de los modelos de interoperabilidad, la construcción de SIG tiene varios enfoques. Cada enfoque presenta un conjunto de ventajas y de inconvenientes, dependiendo del contexto en el que se desarrolle (Lewis *et al.*

2005). Las principales ventajas de los modelos de interoperabilidad son: la capacidad de definir un vocabulario común que permita el análisis y la discusión sobre el significado, proporcionar sugerencias relacionadas con la estructura de las soluciones y, por último, servir de soporte para evaluar nuevas ideas y distintas opciones (*SEI-CMU 2007*).

Cada modelo de interoperabilidad propuesto en la literatura trata propósitos de uso distintos. Sin embargo, la mayoría de estos modelos no aprovechan plenamente las posibles ventajas de algunos tipos de sistemas, tales como la independencia total de la plataforma o la interoperabilidad de las herramientas. Además, los modelos también poseen algunas limitaciones, especialmente cuando se trata de intercambiar los mismos o expresar más información sobre el contexto, como ocurre en las arquitecturas distribuidas.

En la actualidad, cada modelo de interoperabilidad define una taxonomía particular, con la finalidad de apoyar distintos objetivos de uso o alcanzar la interoperabilidad en diferentes contextos. Los conceptos o clasificaciones que habitualmente se usan para definir las taxonomías de los modelos son: capas, dimensiones, niveles y áreas. Dependiendo del tipo de contexto, se aplica uno u otro modelo de interoperabilidad para construir un sistema. A continuación, se presentan algunos ejemplos de modelos de interoperabilidad que han sido aplicados, con éxito, en contextos ajenos a los SIG o las IDE:

- La perspectiva más común y tradicional sobre modelos de interoperabilidad ha sido tratada y representada por el grupo de trabajo de arquitecturas C4ISR en el modelo LISI (1998) (Niveles de interoperabilidad para los sistemas de información). Los cinco niveles del modelo *LISI* son: aislado, conectado, distribuido/funcional, integrado/dominio y global.
- El modelo de madurez de la interoperabilidad empresarial (*EIMM*) (Athena 2005), ayuda a detectar procesos de colaboración dentro de la organización. Este modelo define seis áreas de interés en la evaluación: estrategia de negocios y procesos, organización y competencias, productos y servicios, sistemas y tecnología, entorno de seguridad jurídica y confianza, y por último, modelos empresariales, con una escala de madurez de cinco niveles: alcanzada, modelada, integrada, interoperable y mejorable.

- En el modelo de madurez de interoperabilidad organizacional (*OIMM*) de Clark y Jones (1999) se describen las etapas por las que los sistemas, los procesos o las organizaciones progresan o evolucionan, de acuerdo a cómo están definidas, construidas o mejoradas. El concepto de nivel es intrínseco a este modelo y se utiliza para caracterizar el estado de un sistema o de una organización. Los niveles del modelo son: unificado, combinado, en colaboración, a medida e independiente.
- El modelo de interoperabilidad de agilidad organizacional (*OIAM*), propuesto por Kignston *et al.* (2005), captura los aspectos dinámicos del trabajo en alianza. En este modelo se ha desarrollado una interoperabilidad de agilidad organizacional escalable, que discurre de menos a más ágil, lo que permite relacionar un conjunto de niveles con otro de atributos y factores. Los cinco niveles del modelo son los siguientes: estático, dócil, complaciente, abierto y dinámico.
- El modelo de interoperabilidad conceptual (*LCIM*), definido por Tolk (2003) y refinado por él y por Turnitsa (2006), define 7 niveles de interoperabilidad: no conectado, técnico, sintáctico, semántico, pragmático, dinámico y conceptual. Posteriormente los mismos autores (Turnitsa *et al.*, 2007) proponen el uso de los metadatos en el diseño de sistemas como mecanismo integrador o habilitador de la interoperabilidad.

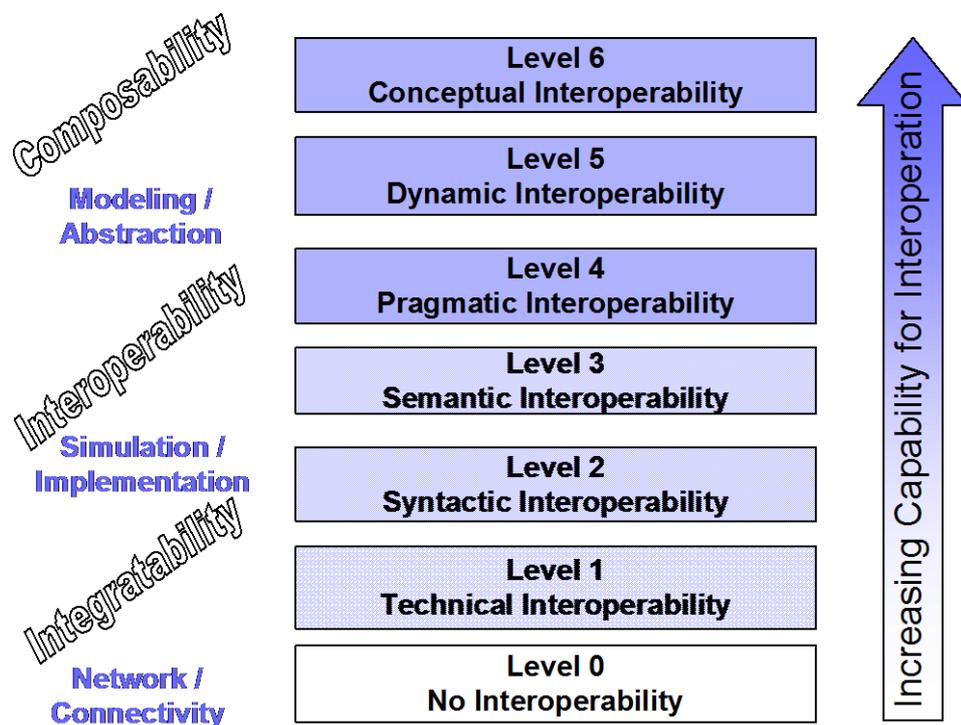


Figura 3. 1: Modelo de interoperabilidad Conceptual de Turnitsa y Tolk (2006)

En el contexto de los SIG, la interoperabilidad se comenzó a tratar hace una década en el panel “*On computation implementations of geographic concept*”s (Goodchild *et al.* 1997). En este panel se adopta un esquema de 8 niveles de interoperabilidad, formado por dominios e instituciones de información, empresas, aplicaciones, herramientas, intermediarios, almacén de datos, entorno de cálculo distribuido y red (“*Information community e institution, enterprise, application, tools, middleware, data store, distributed computing environment and network*”).

En el contexto de las IDE, los principales actores que están promoviendo la interoperabilidad de los datos y los sistemas son el *OGC* e *ISO*, esta última por medio de su comité técnico 211 (TC 211). Ambos, además de definir normas para los datos y los metadatos, están definiendo modelos de objetos y esquemas XML para el almacenamiento y la transferencia de información, así como las interfaces abiertas de servicios de procesamiento.

Goodchild *et al.* (1997) definieron un modelo integrado de Interoperabilidad para los SIG. En éste los autores identificaron que la voluntad para alcanzar la interoperabilidad depende de muchos factores: la voluntad o predisposición, los factores económicos, los problemas legales y organizacionales. De forma abstracta distinguieron cinco niveles en el modelo: ingeniería y redes, tecnología y plataformas, arquitectura computacional y aplicaciones informáticas, modelos conceptuales de datos y de información y finalmente de iniciativa o empresa. El mismo grupo de discusión define un esquema de interoperabilidad basado en 8 niveles, en el que se combinan distintas dimensiones de la organización, incluyendo la escala y la abstracción como se indicó anteriormente y puede verse en la tabla 3.1.

Para Yasher Bishr (1998), el modelo de interoperabilidad en el contexto de los SIG, que trata de evitar principalmente las barreras semánticas, está formado por 6 niveles: protocolos, hardware y sistemas operativos, formatos de datos espaciales, sistemas de gestión de bases de datos, modelos de datos y, finalmente, semántica de aplicación.

A	Intercambia con	B
Comunidades de usuarios, Instituciones	Políticas, valores, cultura	Comunidades de usuarios, Instituciones
Empresa	Acuerdos, consensos	Empresa
Aplicación	Cooperación y coordinación	Aplicación
Herramientas	servicios	Herramientas
Capa intermedia	Objetos distribuidos	Capa intermedia
Almacén de datos	datos	Almacén de datos
Entornos de proceso distribuidos		Entornos de proceso distribuidos
Redes		Redes

Tabla 3. 1: Esquema de interoperabilidad de 8 niveles (Goodchild et al. 1997)

El modelo *Intermodel5*, propuesto por Shanzhen *et al.* (1999), identifica los problemas de interoperabilidad en la conexión y el intercambio de datos entre bases de datos y sistemas. Para estos autores, la interoperabilidad se alcanza definiendo las interfaces mediante especificaciones y estándares. Shanzhen *et al.* (1999) identifican que la investigación en interoperabilidad se encuentra usualmente en torno a los procedimientos para la distribución y la integración de sistemas: igualdad, reciprocidad, intercambio, diversidad, independencia y pertenencia a un dominio. El modelo *Intermodel5* define los siguientes niveles de interoperabilidad: localización de recursos, transformación de los mismos, servicio de aplicación, semántico y finalmente institucional. Además, estos autores establecen relaciones entre el modelo y los datos o los *SII* (sistemas de información espaciales), en los que la arquitectura de los sistemas se definen cuatro niveles: datos, técnico, operacional e institucional.

Sistema A	Método de interoperabilidad	Sistema B
Institución	Políticas, Cultura, Valores	Institución
Semántico	Semántica traductores, metadatos Información Geográfica Formalización de sistemas	Semántico
Aplicación o servicio	Agentes de objetos distribuidos, CORBA, OpenGIS	Aplicación o servicio
Recursos de procesamiento	Bases de datos virtuales, MultiDatabase, OGC SDTS data Warehouse Framework	Recursos de procesamiento
Recursos de localización	Metadatos, librerías digitales, Catálogos, Clearinghouse	Recursos de localización

Tabla 3. 2: Intermodel5, Modelo de Interoperabilidad de 5 niveles (Shanzhe, 1999)

El proyecto *InterOP* (IST508011) para un modelo de Interoperabilidad en la sociedad de la información define un marco de interoperabilidad que distingue distintos puntos de vista, atendiendo los aspectos relacionados con: los datos, los servicios, los procesos y los negocios; considerando las posibles barreras se

pueden clasificar en: conceptuales, tecnológicas y organizacionales y, finalmente, atendiendo al tipo de enfoque del modelo de interoperabilidad (de acuerdo con ISO 14258) puede hablarse de: unificado, integrado y federado.

Una vez revisadas las definiciones del concepto interoperabilidad e identificados los modelos que tratan de analizar las relaciones y los aspectos de la misma, desde un punto de vista estructurado, se va a realizar una revisión en profundidad de los niveles de interoperabilidad que proponen o se citan en la bibliografía, de modo que se pueda disponer de una visión general del problema y de las soluciones propuestas hasta el momento.

Como se ha indicado en el apartado 3.2, las IDE pueden considerarse un caso particular de las II y éstas, a su vez, también un caso particular de los SoS. En este apartado relativo a los modelos de interoperabilidad se ha realizado una revisión de los mismos existentes: en el dominio geográfico (Goodchild *et al.*, 1997, Bishr e Intermodel5), en el dominio de los SoS (LISI y LCIM) y en el contexto organizacional (OIAM y OIMM). Cada modelo trata de responder a un conjunto de necesidades de interoperabilidad; sin embargo, sin hacer un análisis más profundo de los niveles de interoperabilidad propuestos en la literatura, seleccionar un modelo aplicable a las IDE es poco objetivo e injustificado. Una vez identificado el modelo de interoperabilidad a utilizar en el contexto de las IDE y justificados sus niveles, se pretende analizar cómo los metadatos pueden facilitar la interoperabilidad tal y como han propuesto Tolk, Diallo y Turnitsa (2007).

### **3.5 Niveles de interoperabilidad**

Los niveles de interoperabilidad de un modelo son los conceptos, las abstracciones o categorías que distinguen una taxonomía relacionada con la interoperabilidad. Estos niveles se utilizan para definir las capacidades que han de cumplir los sistemas para alcanzar un determinado grado de interoperabilidad.

La revisión bibliográfica propone bastantes niveles: semántico, técnico, legal, organizacional y de otros tipos, como puede apreciarse en la Tabla 3.3. A pesar

de que se distinguen muchos niveles de interoperabilidad, es difícil entenderlos de forma aislada, ya que las clasificaciones propuestas pueden ser similares o estar relacionadas y las definiciones de los distintos niveles pueden tener características comunes.

Autor	Año	Niveles de interoperabilidad														
		Técnico	Esquemática o estructural	Semántico	Organizacional	Físico	Empírico	Sintáctico	Pragmático	Social	Intracomunitario	Político / Humano	Legal	Internacional	Dinámico	Conceptual
ISO	1990	x		x				x								
Goh	1997		x													
Goodchild <i>et al</i>	1997	x	x		x			x	x				x		x	x
Bishr	1998		x	x				x								
Vckovski	1998			x				x								
Harvey, et al.	1999	x		x						x		x				
Shanzhen <i>et al</i>	1999			x	x			x	x						x	
Ouksel & Sheth	1999	x	x	x				x	x	x					x	
Miller	2000	x		x							x	x	x	x		
Nedovic-Budic & Pinto	2001				x											
Tolk	2003	x	x		x							x				x
Tolk & Muguirra	2003	x		x				x	x						x	x
Bermudez	2004			x												
Shekhar	2004		x	x				x								
Schekkerman	2004	x	x	x												
Stroetmann	2005			x				x								
Ding	2005			x				x								
Kuhn	2005	x		x	x			x		x						
Nowak <i>et al</i>	2005		x	x				x								
Mohammadi	2006	x			x					x		x	x			
Kalantari	2006	x		x							x		x			
Vas Assche	2006					x	x	x	x	x						
Turnitsa & Tolk	2006	x		x				x	x						x	x
Whitman <i>et al</i>	2006	x		x				x	x							
Dekkers	2007	x		x	x											
Chen, D.	2007	x		x	x			x								
Zeigler & Hammonds	2008			x				x	x							

Tabla 3. 3: Tabla resumen de los niveles de Interoperabilidad identificados en la literatura

Los 15 niveles de interoperabilidad identificados son: Semántico, Sintáctico, Técnico, Pragmático, Organizacional, Esquemático o Estructural, Dinámico, Legal, Conceptual, Social, Intracomunitario, Político/Humano, Internacional, Empírico y Físico.

En la figura 3.2 se muestra la frecuencia con la que cada nivel ha sido citado. Se puede observar que los niveles físico, empírico e internacional son citados por un único autor, mientras que el nivel semántico de interoperabilidad es el más

citado, seguido del nivel sintáctico, técnico, pragmático y organizacional respectivamente. Para el resto de niveles la frecuencia de las citas es baja.

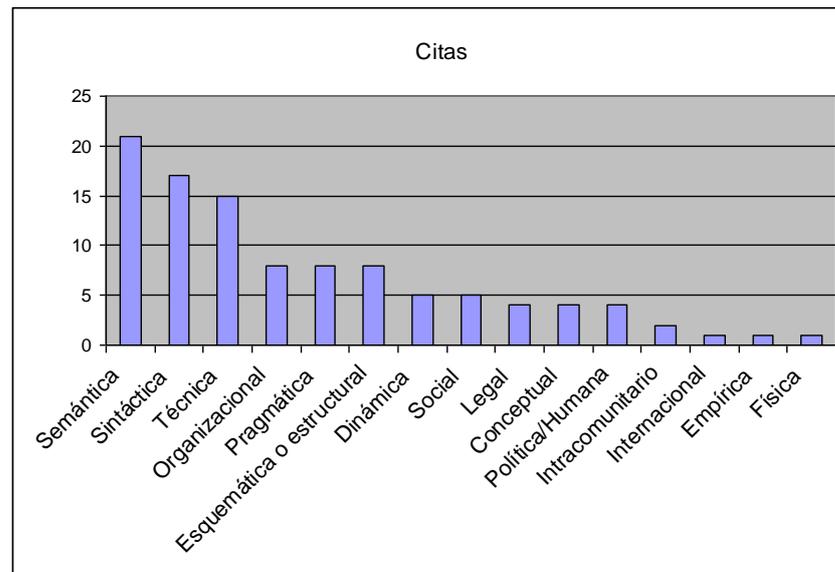


Figura 3. 2: Histograma de los niveles de interoperabilidad

Como se ha indicado anteriormente, para poder entender el problema en su conjunto, a continuación se presentan las definiciones de los niveles de interoperabilidad interpretadas para los autores revisados. Esta revisión se ha extendido a todos los niveles identificados, pertenezcan o no a un modelo de interoperabilidad.

### 3.5.1 Interoperabilidad Semántica

*SEI* en el modelo *LCIM*, identifica la semántica como la que permite transferir los datos y el contexto, de modo que se comparta el significado de los datos sin ambigüedad.

*ISO* en la familia de normas incluidas en la serie 19000, incluye en sus normas *ISO-19101 Modelo de referencia* y *ISO-19119 servicios*, los tipos de interoperabilidad y tratan la semántica.

Goodchild *et al.* (1997) indican que cualquier nivel de interoperabilidad semántica gestiona las dificultades que pueden surgir al compartir los significados.

La guía del OpenGIS (1998) identifica problemas semánticos en los datos geográficos o espaciales. Para disponer de integridad semántica, se propone el uso de lenguajes comunes, modelos conceptuales y se sugiere el uso de conceptos comunes.

Bishr (1998) identifica problemas semánticos en el modo en que las distintas disciplinas, o comunidades de usuarios, describen los objetos del mundo real en las bases de datos propiciando, en consecuencia, la heterogeneidad semántica. Por ejemplo, una red de carreteras, para un gestor de pavimento, posee diferentes descripciones semánticas que los datos de infraestructuras de transportes mantenidas en una base de datos SIG, diseñada para aplicaciones de cartografía topográfica de escalas pequeñas.

Vckovski (1998) define el concepto de diversidad semántica. Ésta suele producirse cuando cooperan las comunidades de información y la diversidad se produce por no haber definido semánticamente los conceptos, los objetos, etc.. Cada comunidad de información suele entender los datos y sus asunciones implícitas de un modo distinto al resto. Sin embargo, estas asunciones son muy peligrosas. Según el autor, la única forma de dar solución a este problema es el uso del razonamiento. Se sugiere usar conceptos en sistemas de apoyo a la toma de decisiones para resolver la heterogeneidad semántica. Otro aspecto que destaca el mismo autor es la falta de previsión sobre el futuro uso que se puede dar a los datos geográficos.

Harvey *et al.* (1999) indican que no se alcanzan los niveles de interoperabilidad semántica gracias a la estandarización, ya que el significado de los conceptos es diferente en distintos dominios. Como ejemplo, aparecen problemas de interoperabilidad al fusionar datos geográficos, ya sea por proceder de distintas fuentes o por existir una frontera cultural. Los autores identifican tres aspectos de este tipo de interoperabilidad, en los que se está investigando: cognitivo, computacional y lingüístico. Los mismos autores proponen el concepto de similitud semántica como una métrica que permite medir el grado de interoperabilidad alcanzado.

Ouksel (1999) identifica seis elementos en la infraestructura de sistemas abiertos para interacción social, relacionados con la interoperabilidad semántica: el significado, las proposiciones, la validez, la veracidad, la relevancia y las denotaciones.

Shanzhen (1999) identifica como objetivo de este nivel de interoperabilidad el intercambio semántico entre distintos dominios de conocimiento o comunidades de usuarios, usando estándares especiales de información y reglas de correspondencia entre comunidades. Para alcanzar este objetivo, es imprescindible disponer de un cuerpo de conocimiento común y fundamental en torno a la IG y su teoría.

Bermudez (2004) y Goh (1997), en sus respectivas tesis doctorales relacionadas con las ontologías en el contexto de los metadatos, han identificado problemas semánticos de interoperabilidad en los sinónimos y homónimos, en las escalas y en las unidades de medida, comprendiendo los aspectos relacionados con la interpretación errónea de un concepto.

El proyecto IDABC (2004) identifica como aspecto clave de la interoperabilidad semántica, la necesidad de interfaces multilingües en la oferta de servicios a los usuarios.

Schekkerman (2004) se centra en el significado de la información que se intercambia. En este caso, la interoperabilidad semántica está relacionada con el objetivo de asegurar que cualquier otra aplicación desarrollada para otros propósitos entienda el significado de los datos intercambiados de un modo preciso.

Shekhar (2004) define la interoperabilidad semántica como un requisito necesario para los sistemas de información dedicados a la minería de datos espaciales. Este autor destaca la necesidad de especificar, mediante un documento, el contenido de los datos espaciales, los modelos y las relaciones. Desde un punto de vista ideal, se propone compartir un vocabulario y una ontología de conceptos. En su defecto,

se propone usar un sistema de traducción bien definido, que permita relacionar los conceptos utilizados por los sistemas que envían y reciben los datos.

Hao Ding (2005), propone el almacenamiento de relaciones entre ontologías y el uso de relaciones cruzadas borrosas, como soluciones a los problemas de interoperabilidad semánticos.

Kuhn (2005) sugiere que la interoperabilidad semántica es el único tipo útil, aunque posteriormente cita otros aspectos de interoperabilidad como relevantes. El trabajo sugiere que la Web semántica no trata todos los aspectos semánticos, ya que no se trata el significado de las operaciones que ofrecen los servicios, dando más relevancia a la semántica de los parámetros y los resultados. El autor diferencia dos tipos de heterogeneidad semántica al realizar integraciones de datos (*matching*): sintácticos y estructurales. El mismo autor propuso con anterioridad el concepto de sistemas de referencia semánticos (Kuhn 2003) por semejanza con los sistemas de referencia espaciales, teniéndose que definir las reglas de transformación de un sistema a otro.

Nowak (2005) define como aspectos de interoperabilidad semántica los relacionados con la diferencia en el significado de una misma entidad global en distintas bases de datos (diferencia en el contexto de la información).

Pokraev (2005) afirma que los problemas de interoperabilidad semántica aparecen cuando el modelo del “dominio objeto” a nivel de usuario es distinto que a nivel de sistema. Se refiere como modelo de “dominio objeto” a la parte del mundo sobre el que trata el mensaje (Wieringa 2003).

Assche (2006) identifica como problemas de interoperabilidad semántica aquellos relacionados con el significado y la validez de lo que está expresado.

Antonovic y Novak (2006) proponen que los actores deben de ser capaces de entender la información que están intercambiando, es decir, compartir el significado de los elementos de información con el menor error y ambigüedad posible.

Kalantari *et al* (2006) inciden en la armonización de la terminología y la interpretación de los conceptos.

Probst (2006) identifica problemas de interoperabilidad semántica al analizar, desde un punto de vista ontológico, la especificación del *OGC* relativa a observaciones y medidas. Se propone un modelo conceptual y se indica que en la especificación hay aspectos que dependen de la interpretación humana al aportar las descripciones en forma de texto libre. Como consecuencia de esa circunstancia se identifican problemas en los procesos de localización y evaluación de las fuentes de objetos geográficos.

Turnitsa y Tolk (2006) enfatizan en la necesidad de un modelo de referencia común para el intercambio de información, y refuerzan la necesidad de compartir el significado de los datos.

Chen *et al.* (2007), en el proyecto *InterOp*, han identificado barreras de naturaleza conceptual, y consideran los problemas semánticos como una barrera importante para la interoperabilidad. Relacionan estos problemas con el hecho de que en los modelos, o en las aplicaciones informáticas, no se haya definido claramente la semántica a utilizar para representar la información y el conocimiento.

Dekkers (2007) define este tipo de interoperabilidad como la “*comprensión del significado de la información, teniendo presente datos de otras personas*”. Agrupa bajo este nivel las cuestiones relacionadas con el tipo de información en sí misma, las responsabilidades sobre los datos y su mantenimiento, además de los aspectos relacionados con las condiciones y las restricciones de uso; también relaciona todo lo anterior con la estructura (estándares de metadatos), con vocabularios, valores y clasificaciones.

Zeigler y Hammonds (2007), en su modelo de intercambio de información *Net-Centric*, indican que los lenguajes *XML*, *XML-Schema* y *UML* dotan de interoperabilidad sintáctica y semántica, pudiendo expresar la semántica desde un

punto de vista ontológico. Identifican la carencia de ítems de metadatos que permitan mejorar la semántica, relacionada con la dinámica de modo práctico.

### **3.5.2 Interoperabilidad Sintáctica**

Bishr (1998) entiende por problemas de interoperabilidad sintáctica los relacionados con la implementación de las bases de datos en distintos paradigmas (relacional, objetos), o la representación geométrica de los objetos (raster, vector).

Vckovski (1998) hace referencia a la diversidad sintáctica indicando que es la fuente de problemas para el intercambio de geodatos, y es debido a la diversidad de formatos de almacenamiento y transferencia, a pesar de los esfuerzos de estandarización. El autor destaca que estos problemas son especialmente importantes en el contexto de la gestión de la calidad de los datos. Muchos de los problemas de interoperabilidad sintáctica suponen la conversión manual de formatos, afectando en muchos casos a la calidad de los mismos; como ejemplo de dichos problemas se cita la capacidad de expresividad de los formatos. En algunos casos dichos formatos contienen información que las herramientas no saben manejar o que desconocen. Esto implica la necesidad de personal especialmente entrenado para realizar dichas conversiones. Otro aspecto que evidencia también el autor es que en algunos casos se pueden interpretar problemas semánticos como si fueran sintácticos, y sugiere como ejemplo los homónimos. También puede ocurrir que no se detecte la heterogeneidad sintáctica (por ejemplo: codificaciones del texto, orden de los bytes). El autor concluye diciendo que la fuente de estos problemas suele ser la falta de especificaciones explícitas de formatos. En muchos casos se hacen asunciones implícitas de un modo no conciso.

Ouksel (1999), en la infraestructura de sistemas abiertos para interacción social, identifica algunos aspectos sintácticos para la formalización de estructuras, lenguajes, lógica, datos, registros, deducciones, programas y ficheros.

También las normas ISO19101 (2002) e ISO19119 (2005) identifican el nivel de interoperabilidad sintáctico como uno de los objetivos a alcanzar.

Shekhar (2004) define la interoperabilidad sintáctica como la especificación de formatos comunes para los mensajes (por ejemplo usar marcas y etiquetas), para intercambiar datos espaciales, modelos y relaciones.

Nowak *et al.* (2005) identifican como aspectos sintácticos que motivan la falta de interoperabilidad, a la heterogeneidad de las plataformas hardware y software, a la representación geométrica y temática y a las relaciones de los objetos espaciales (sistemas de coordenadas, resoluciones geométricas, representación geométrica, calidad).

Turnitsa y Tolk (2006) mencionan que el nivel de interoperabilidad sintáctica debe apoyarse en una estructura común para el intercambio de información, en la que se requiere, por ejemplo, formatos comunes de datos.

Assche (2006) atribuye a este nivel de interoperabilidad aspectos relacionados con el lenguaje, la estructura y la lógica usada.

*SEI* ha introducido este nivel en el modelo *LCIM*, indicando que es el que posibilita el intercambio de datos utilizando formatos estandarizados.

Chen *et al.* (2007) identifican barreras de naturaleza conceptual en el proyecto *InterOp*, pudiéndose encontrar diferencias sintácticas cuando distintas personas o sistemas utilizan diferentes estructuras para representar la información y el conocimiento.

Zeigler y Hammonds (2007) hacen una breve retrospectiva e identifican carencias en aspectos sintácticos en los modelos de transmisión antiguos, en los que la información se ordena en serie y se espera una estructura concreta de información. Actualmente, estos aspectos han sido solventados gracias a las tecnologías *SOA* y la codificación de la información en formatos *XML*.

### 3.5.3 Interoperabilidad Técnica

Goodchild *et al.* (1997) consideran como aspectos técnicos de la interoperabilidad a la computación distribuida, a las redes de comunicaciones, a las tecnologías por sí mismas y a las plataformas de cálculo distribuido. Amplía la necesidad de mejorar la interoperabilidad técnica a los datos y a las capas intermedias de software (*middleware*).

Ouksel (1999) propone como solución a la heterogeneidad de sistemas el protocolo de Internet *Inter-ORB* (IIOP), para la interacción entre objetos y componentes distribuidos.

Miller (2000) identifica como aspectos técnicos de la interoperabilidad los estándares de comunicación, transporte, almacenamiento y representación; cita, como ejemplo, el protocolo de comunicación Z39.50 (ISO23950), destinado a facilitar la búsqueda y recuperación de información en distintos sistemas.

ISO en la norma ISO19101 (2002) *Información Geográfica - Modelo de referencia* se describe el nivel de interoperabilidad técnica, atribuyéndole las comunicaciones, transferencia de información y ejecución de programas en unidades funcionales.

El proyecto IDABC (*European Interoperability Framework for pan-European eGovernment services*) (2004) identifica como aspectos técnicos los relacionados con la conexión de servicios y sistemas informáticos a distintos niveles: interfaces abiertas, servicios de interconexión, integración de datos, intercambio y presentación de datos, accesibilidad, así como servicios de seguridad.

Schekkerman (2004) afirma que la interoperabilidad técnica está básicamente relacionada con la conexión de los sistemas informáticos y los servicios.

Turnitsa y Tolk (2006) también identifican los aspectos técnicos de la interoperabilidad con el intercambio de datos. Sugieren el uso de protocolos de comunicación comunes, con el objetivo de que los sistemas intercambien bits y bytes.

Kalantari *et al* (2006) entienden por interoperabilidad técnica el desarrollo de estándares de comunicación, intercambio, modelado y almacenamiento de los datos, así como los portales y los servicios Web que interoperen equipados con interfaces amigables.

Antonovic y Novak (2006) introducen aspectos claves para alcanzar la interoperabilidad técnica, como son: interfaces abiertas, interconexión de servicios, integración de datos, capas intermedias de software, presentación e intercambio de datos, accesibilidad y servicios de seguridad.

Mohammadi *et al.* (2006) identifican y catalogan los siguientes problemas como aspectos de interoperabilidad técnica: disparidad en los entornos de computación, falta de estándares, topología 3D (vertical), sistemas de referencia espaciales, escalas, calidad, modelo de datos, metadatos, formatos y semántica.

*Software Engineering Institute (SEI) - Carnegie Mellon University* - propone tener en cuenta los aspectos técnicos de la interoperabilidad al crear modelos, y coincide con Tolk y Turnitsa (2006) resaltando el intercambio de información y de bits, a nivel de protocolos de comunicación.

Chen *et al.* (2007) identifican problemas de interoperabilidad técnica en el proyecto *InterOp*, relacionándolos principalmente con las incompatibilidades de las tecnologías de la información (arquitecturas, plataformas, infraestructura, etc.).

Dekkers (2007), director de gestión de *DCMI*, define este nivel de interoperabilidad y lo relaciona con la interconexión, presentación, intercambio de datos, accesibilidad y seguridad. Agrupa bajo este nivel los aspectos relacionados con los protocolos de comunicaciones y las interfaces de los servicios, los formatos, las codificaciones, las medidas de accesibilidad y las soluciones de seguridad.

### 3.5.4 Interoperabilidad Pragmática

Ouksel (1999) identifica en el documento “*La infraestructura de sistemas abiertos para interacción social*” como aspectos de la interoperabilidad pragmática, las intenciones, comunicaciones, conversaciones y negociaciones.

Shanzhen (1999), en el modelo *Interomodel5* para computación distribuida, propone el nivel de servicio de aplicaciones relacionándolo con los modelos de computación distribuidos, que posibilitan el intercambio de servicios de geo-procesamiento y el análisis entre distintas comunidades y departamentos.

Para Pokraev *et al.* (2005) el problema de la interoperabilidad pragmática aparece cuando el efecto de un mensaje difiere del que se esperaba de él.

Assche (2006) atribuye a este nivel de interoperabilidad los objetivos, las responsabilidades y las consecuencias ocultas en la información o los mensajes. La interoperabilidad pragmática alcanza a todos los usuarios de servicios interoperables que tengan objetivos, responsabilidades y consecuencias compatibles, en relación con los propios servicios y la información intercambiada. En este contexto, la interoperabilidad pragmática está especialmente relacionada con el hecho de que todas las partes involucradas y usuarias de un servicio tengan definidas sus responsabilidades (roles).

Para Turnitsa y Tolk (2006), se alcanza este nivel de interoperabilidad cuando los sistemas conocen y son capaces de usar los métodos y procedimientos ofrecidos e implementados por los sistemas. En otras palabras, que el uso de los datos o del contexto de sus aplicaciones sea comprendido por los sistemas participantes.

*SEI*, en el modelo *LCI*, agrupa bajo un solo nivel las interoperabilidades pragmática y dinámica. Las identifica como aquéllas que permiten identificar el contexto de los datos y las posibles formas de aplicarlos, sin lugar a posibles ambigüedades.

Zeigler y Hammonds (2007) definen, en su modelo de intercambio de información, el nivel pragmático relacionándolo con aspectos de la transmisión, la

detección de errores, la corrección de los mismos o la negociación de la retransmisión.

### **3.5.5 Interoperabilidad Organizacional**

*SEI* propone considerar los aspectos organizacionales y culturales de la interoperabilidad para crear modelos que permitan evaluarla pero no como modelo de interoperabilidad entre sistemas.

Goodchild *et al.* (1997) identifican los aspectos relacionados con la institución, que pueden ser más problemáticos que el resto, ya que dependen de muchos factores, entre los que se pueden destacar: de comportamiento, económicos, problemas legales y organizacionales.

Ouksel (1999) identifica en el documento “*la infraestructura de sistemas abiertos para interacción social*” aspectos tales como: comportamientos, expectativas, contratos, leyes, cultura y compromisos.

Shanzhen (1999) describe que los problemas de interoperabilidad institucional aparecen cuando no existen, o no se conocen, las políticas de intercambio de datos entre los departamentos; estos intercambios pueden estar condicionados por las diferentes políticas, culturas, valores y aspectos ligados con la privacidad. También se afirma que es necesario coordinar los aspectos políticos, culturales y los valores en las relaciones entre distintas comunidades.

Nedovic-Budic y Pinto (2001) tratan monográficamente el aspecto de interoperabilidad organizacional en el contexto SIG. Identifican cuatro áreas de trabajo que tienen que ver con: interacciones, implementación, coordinación y resultados. En las interacciones se tratan los aspectos de las responsabilidades, las políticas de acceso a los datos y los acuerdos para la gestión de recursos. En la implementación se citan siete aspectos a tener en cuenta: los recursos, la experiencia, la estabilidad, la coordinación, el acceso a los datos, la capacidad para implementar, los aspectos del liderazgo y la gestión. En lo relativo a la coordinación se citan como objetivos los acuerdos, las reglas y los roles.

Finalmente, en la parte de los resultados, indican que el objetivo es mejorar las relaciones entre las organizaciones, además de alcanzar la satisfacción por los logros alcanzados.

Schekkerman (2004) identifica como aspectos de interoperabilidad organizacional los relacionados con los objetivos de negocio, los modelos de los procesos de negocio y el intercambio de información entre organizaciones. Se propone como posible solución a estos aspectos considerar los requisitos de la comunidad de usuarios, haciendo que los servicios estén disponibles, accesibles y fácilmente identificables, con una orientación al usuario.

Mohammadi *et al.* (2006) proponen, a nivel institucional, considerado equivalente al nivel organizacional, los aspectos relacionados con los modelos de colaboración y financiación y los vínculos entre las unidades que gestionan tanto los datos como el conocimiento de su existencia.

Para el proyecto SAGA (2006), el principal objetivo de la interoperabilidad organizacional es determinar cuándo y por qué se intercambian datos. Por tanto, el enfoque de este tipo de interoperabilidad es definir el marco legal de referencia para el intercambio de datos.

Dekkers (2007) incluye bajo este nivel la cooperación entre organizaciones, conocida como *BPI* (interoperabilidad de procesos entre empresas), los objetivos empresariales y el modelado de los procesos. Se incluye el intercambio de información entre socios (dominio científico, clientes y proveedores, agencias de la administración), así como toda la coordinación de las tareas empresariales, tales como las necesidades de comunicación de los usuarios o los resultados de los análisis de costes y beneficios.

El proyecto IDABC (2004) reitera los aspectos descritos por Dekkers (2007) teniendo en cuenta los requisitos de los usuarios relacionados con la identificación y el acceso a los mismos, así como su disponibilidad.

Chen *et al.* (2007), en el proyecto *InterOp*, identifican problemas de interoperabilidad organizacional, al identificar a la autoridad y sus responsabilidades dentro de la organización. Indican que los factores humanos y las tecnologías están relacionados con los comportamientos humano y organizacional, de tal forma que puede llegar a ser imposible la interoperabilidad (incompatibles).

### **3.5.6 Interoperabilidad Esquemática o Estructural**

Goh (1997) enumera los siguientes aspectos estructurales de la interoperabilidad: tipos de datos (distintas primitivas de datos en distintos sistemas), conflictos en las etiquetas (sinónimos y homónimos en distintos esquemas), discrepancias en las agregaciones (diferentes formas de diseño o asignación de atributos a las entidades) y conflictos de generalización (formas de relacionar unas entidades con otras).

Bishr (1998) identifica como heterogeneidad esquemática situaciones tales como que los objetos de una base de datos son considerados como propiedades en otra, o que las clases de objetos pueden tener distintos niveles jerárquicos de agregación y generalización, aunque describan los mismos objetos de la realidad.

Ouksel y Sheth (1999) proponen el uso del formato *RDF* (infraestructura para la descripción de recursos) desarrollado con propósitos generales de descripción de fuentes de información o de modelos de objetos, para el intercambio de información en la Web (Manola 1998). El formato MPEG-4 es otro ejemplo en el que se describe la estructura o nivel de objetos para vídeo, MHEF-5 para multimedia e hipermedia, *KIF* (formato de intercambio de conocimiento) para la representación del conocimiento o *OKBC* (librería abierta para el acceso a bases de datos de conocimiento) como base para la distribución del conocimiento.

Shanzhen (1999) afirma que los datos procedentes de distintas fuentes poseen diferentes estructuras y esquemas de datos y propone en "*Interomodel5*", el nivel de transformación de recursos como la solución para resolver este tipo de heterogeneidad de datos. En este sentido, el *FGDC* (Comité Cederal para datos geográficos norteamericano) ha especificado el estándar *SDTS* (para la

transferencia de datos espaciales). El OGC ha propuesto la especificación *SF* (fenómenos simples) contribuyendo positivamente a este nivel de interoperabilidad, para solucionar los problemas de heterogeneidad estructural.

Shekhar (2004) sugiere un nivel intermedio de interoperabilidad estructural, de tal modo que se proporcionen los medios que permitan especificar la semántica de los esquemas de datos (metadatos) para compartirlos.

Nowak *et al.* (2005) definen como aspectos relacionados con la interoperabilidad estructural, la heterogeneidad de los modelos de datos y los esquemas de las bases de datos (jerarquía de clases, estructura de atributos, etc.).

### **3.5.7 Interoperabilidad Dinámica**

Shanzhen (1999) propone, en el modelo *Interomodel5*, el nivel de localización de recursos, que se fundamenta en la existencia de estándares de metadatos y en la capacidad de localizar recursos para su explotación: primero buscamos objetos interoperables que resuelvan nuestras necesidades. En el nivel de localización de recursos, se definen los objetos y el método de búsqueda. El problema de este nivel está resuelto. Hay muchos métodos, tales como ha hecho el FGDC de Norteamérica, desarrollando estándares de metadatos para los almacenes de datos del proyecto *NSDI* (Infraestructura de datos espaciales nacional). Entre las descripciones contenidas en los metadatos está la información espacial y las potenciales aplicaciones de los datos.

Turnitsa y Tolk (2006) afirman que se alcanza este nivel de interoperabilidad cuando los sistemas son capaces de reaccionar ante los cambios de estado de los demás, o ante condiciones que afectan al intercambio de los datos y son capaces obtener ventajas de dicho cambio.

### **3.5.8 Interoperabilidad Legal**

Miller (2000) identifica como aspectos de interoperabilidad legal aquellos relacionados con los derechos de propiedad intelectual (*IPR*) y con las leyes o

normas que facilitan la disseminación de la información en el sector público, entre otros.

Kalantari *et al.* (2006) proponen la creación de directrices, reglas, parámetros e instrucciones para gestionar el flujo de trabajo, considerando la información e incorporando las comunicaciones en la gestión del territorio. Esta consideración puede ser extendida en el contexto de las IDE.

Mohammadi *et al.* (2006) identifican aspectos como los derechos de propiedad intelectual, las restricciones y responsabilidades, las licencias y las limitaciones de acceso o privacidad de los datos, como aspectos de interoperabilidad legal.

### **3.5.9 Interoperabilidad Conceptual**

*SEI*, en el modelo *LCIM*, propone que este nivel debe permitir establecer una vista común del mundo, basándose, por ejemplo, en la epistemología. Este nivel debe de contener, además del conocimiento implementado, las relaciones entre elementos.

Goodchild *et al.* (1997) afirman que la interoperabilidad empresarial se alcanzará cuando sea evidente la conceptualización a nivel individual.

Turnitsa y Tolk (2006) definen este tipo de interoperabilidad como aquél que puede alcanzarse cuando se documenta el modelo conceptual, mediante métodos usados en ingeniería, de modo que pueda ser interpretado y evaluado por una tercera parte. En este caso, si el modelo conceptual está alineado (por ejemplo las asunciones y las restricciones de la abstracción sobre una realidad), significa que se alcanza el mayor nivel de interoperabilidad. Adicionalmente, se debe cumplir que el método utilizado para documentar el modelo conceptual no influya en la implementación, ni dependa del propio modelo.

### **3.5.10 Interoperabilidad Social**

Ouksel (1999) identifica en “*la infraestructura de sistemas abiertos para interacción social*” algunos aspectos sociales como son: intereses, expectativas, contratos, leyes, cultura y compromisos.

Harvey *et al.* (1999) citan, simplemente, los aspectos sociales de interoperabilidad.

Assche (2006) atribuye a este nivel los aspectos relacionados con los intereses, las creencias y las conclusiones compartidas como resultados

Mohammadi *et al.* (2006) añaden a la propuesta anterior los aspectos culturales, la formación y la identificación o conocimiento de las responsabilidades.

### **3.5.11 Interoperabilidad intracomunitarios**

Miller (2000) define como aspectos de interoperabilidad intracomunitarios los relacionados con las soluciones comunes para distintos niveles de detalle en la descripción geográfica y en distintos dominios de conocimiento, ciencia, etc.

Kalantari *et al.* (2006) identifican este tipo de interoperabilidad con la coordinación y la alineación de los procesos empresariales y las arquitecturas de información, que comprende tanto a personas, como a los socios (nivel privado) y al sector público.

### **3.5.12 Interoperabilidad Política/Humana**

Harvey *et al.* (1999) citan, sin describir, los temas políticos como aspectos de interoperabilidad.

Miller (2000) identifica como aspectos políticos y humanos de la interoperabilidad aquellos relacionados con las políticas o directrices de las organizaciones, orientadas a la diseminación y el mantenimiento de información. Para garantizar la interoperabilidad, la información debe de mantenerse

actualizada y difundirse, o en su defecto, puede que desaparezca junto a las personas encargadas de su mantenimiento.

Mohammadi *et al.* (2006) identifican las carencias en materia de legislación, políticas de precios o en la identificación y en la asignación de prioridades.

Aunque se ha identificado otro documento en los que se referencia la interoperabilidad política, o humana, en el contexto de la salud, no se ha incluido la referencia por tratar los mismos aspectos ya descritos por los anteriores autores.

### **3.5.13 Interoperabilidad internacional (lingüística)**

Miller (2000) define los aspectos de interoperabilidad internacional como aquellos ligados al idioma en el que se proveen o describen los datos. El objetivo es evitar los problemas relacionados con la comprensión de determinados idiomas, por parte de los usuarios. También incluye en esta categoría los aspectos relacionados con las prácticas de uso, la cultura, las expectativas y las necesidades de los usuarios. No se han identificado otros autores que traten el nivel de interoperabilidad internacional.

### **3.5.14 Interoperabilidad Empírica**

Assche (2006) atribuye a este nivel de interoperabilidad los aspectos relacionados con la entropía, la confusión y la diversidad. No se han identificado otros autores que traten el nivel de interoperabilidad empírica.

### **3.5.15 Interoperabilidad Física**

Assche (2006) entiende por aspectos de interoperabilidad a nivel físico los relacionados con la apariencia física, el medio y la cantidad de contacto o interacción entre los sistemas. No se han identificado otros autores que traten el nivel de interoperabilidad física.

En la literatura revisada no se han identificado métodos que permitan validar los modelos de interoperabilidad. En algunos casos, como Whitman *et al.* (2006) proponen realizar la validación por medio de comparaciones de los resultados obtenidos con los esperados. Por este motivo, en esta tesis se propone revisar la literatura existente en el contexto de la medida de la interoperabilidad como primer paso para proponer un método que permita validar la interoperabilidad, ya sea sobre el modelo o sobre los distintos niveles.

### **3.6 Medidas de la interoperabilidad**

Medir el grado de la interoperabilidad existente entre dos sistemas permite conocer las fortalezas y debilidades de ambos cuando operan conjuntamente. Esta medida puede ayudar a mejorar la interoperabilidad y solucionar carencias. Ya se han realizado algunos avances en esta dirección (Pridmore y Rumens 1989, Hamilton *et al.* 2004, Kasunic y Anderson 2004, Janowicz *et al.* 2008), sin embargo los principales obstáculos surgen al definir las métricas. Métrica, en el contexto de la ingeniería del software, se define como cualquier medida o conjunto de medidas que permiten caracterizar un software o sistema de información. En este caso no se trata de caracterizar un sistema de información, sino la interoperabilidad entre sistemas de información. Daclin *et al.* (2006) destacan que lo arduo es identificar los parámetros que caracterizan la interoperabilidad, para poder aplicar una medida.

Pridmore y Rumens (1998), en el documento “*¿Cómo saber cuando se ha alcanzado la interoperabilidad?*”, indican que para saber si se ha conseguido ésta, primero se han de definir los requisitos de interoperabilidad y posteriormente se han de realizar medidas para comprobar si se ha logrado. Para medir la interoperabilidad hace falta una métrica. La métrica define un marco de referencia común sobre el que realizar las medidas. Además de definir la métrica se han de establecer las ponderaciones de los indicadores relativos de los logros de interoperabilidad dentro del sistema, en función de lo crítico o importante que sea la parte del sistema que se esta observando. El proceso de medida no finaliza ahí, sino que se han de establecer ponderaciones a cada una de las medidas para estimar la confianza que se tiene sobre los resultados.

Aplicar una métrica supone comparar el sistema real con un patrón o escala de medida (definido en base a los requisitos), y expresar los resultados mediante términos simples. Los pasos fundamentales para realizarlo son:

- Identificar cada medida con un requisito del sistema,
- Comparar el resultado de la medida con el objetivo,
- Normalizar el resultado y definir el grado alcanzado para el requisito,
- Estimar la confianza del resultado para cada comparación,
- Ponderar los resultados con valores relativos del sistema en su conjunto,
- Combinar los resultados para formar la medida de la interoperabilidad y su confianza.

Los autores concluyen su trabajo con una reflexión crítica: se podría llegar a interpretar que una métrica no proporciona un mecanismo normalizado para medir la interoperabilidad, si bien puede proporcionar un marco de referencia y un método para expresar y evaluar problemas de interoperabilidad. También indican que las métricas pueden ayudar a contestar cuestiones tales como: ¿Cuál es el estado general de la interoperabilidad?, ¿Cumple el sistema todos los objetivos y funcionará?, ¿El resultado de hacer algún cambio, será mejor o peor?, ¿Cuales son las áreas problemáticas y cuán graves son dichos problemas?, ¿Está la confianza bien definida?, ¿Se han realizado todas las verificaciones que debían de hacerse? o ¿Qué confianza debería asignarse al resultado?

Hamilton *et al.* (2004) han tratado el desarrollo de métricas de interoperabilidad y declaran que ésta es notoriamente difícil de medir por lo que, generalmente se propone el uso de modelos simples. Los modelos para realizar las medidas precisan disponer de métricas, que permitan medir el grado de interoperabilidad alcanzado. Las métricas pueden evaluar la interoperabilidad, desde un punto de vista cuantitativo y cualitativo.

Kasunic y Anderson (2004), en el documento “*Midiendo la interoperabilidad de los sistemas: retos y oportunidades*”, afirman que el desarrollo y uso de medidas precisas en un ámbito tan complejo y multidimensional como la interoperabilidad, es difícil de conseguir, y trata las múltiples facetas de la misma, asociadas a un

dominio; propone cuatro juegos de medidas que aborden los siguientes aspectos de este difícil problema: (a) conformidad técnica relativa al cumplimiento de normas, (b) medidas de interoperabilidad de sistemas, centradas en el flujo de la información, (c) interoperabilidad de explotación, centrada en medidas que verifican si se satisfacen los requisitos específicos del flujo de información nodo a nodo, y (d) medidas organizacionales y culturales.

Daclin *et al.* (2006), en el documento “*Medida de la interoperabilidad empresarial - Conceptos básicos*”, también afirman que la medida de la interoperabilidad tiene por objeto definir métricas para calificar el grado de interoperabilidad. La aplicación de la métrica con el objeto de medir el grado de interoperabilidad está relacionada con dos principios: (1) la identificación de los parámetros relativos a la interoperabilidad y (2) la caracterización de estos parámetros por medio de una métrica. El grado de interoperabilidad de un sistema dado, puede definirse por un vector compuesto por tres tipos de medidas: (a) medida de interoperabilidad potencialidad, (b) medida de compatibilidad de interoperabilidad y (c) medida de interoperabilidad de rendimiento, en el actual momento de la investigación.

La medida de la potencial interoperabilidad está relacionada con la identificación de un conjunto de propiedades del sistema, que tengan impacto en el desarrollo de la interoperabilidad. El problema surge cuando los sistemas involucrados evolucionan dinámicamente y la medida se debe acomodar con el nuevo sistema. La medida de compatibilidad se efectúa durante la etapa de ingeniería, es decir, cuando se hace la re-ingeniería del sistema, con el objeto de establecer la interoperabilidad. La medida de la ejecución deberá llevarse a cabo durante la fase operativa, es decir, en tiempo de ejecución, para evaluar la capacidad de funcionamiento conjunto de los dos sistemas.

Chen y Daclin (2007), en el documento “*la metodología para la interoperabilidad empresarial impulsada por los obstáculos*”, han propuesto un método similar al ciclo de desarrollo de aplicaciones informáticas, compuesto por: (a) definición de los objetivos y necesidades, (b) análisis del sistema actual, (c) selección y combinación de soluciones y, finalmente, (d) aplicación y prueba. En paralelo,

propone la participación de cuatro grupos de actores, como se define en la metodología de ingeniería “Gráfico con Resultados y Actividades Interrelacionadas” (GRAI): equipo de proyecto, grupo de síntesis, grupo de especialistas y grupo de entrevistadores.

Recientemente, Daclin *et al.* (2008), en el documento “*Metodología para la interoperabilidad empresarial*”, proporcionan algunas soluciones o mejoras para la interoperabilidad, y un método para medir y evaluar la capacidad de interoperar (grado de interoperabilidad y rendimiento). Esta metodología considera sólo a dos interlocutores y los resultados operativos se limitan a los aspectos técnicos (rendimiento de la comunicación y el intercambio de información) y, por último, otra limitación de esta metodología es el tipo de interoperabilidad con el que está relacionada, el intercambio de información entre las personas pertenecientes a diferentes empresas.

Janowicz *et al.* (2008) indican, en el documento *Medidas de similitud semántica*, que dichas medidas son propias de las ciencias cognitivas y que históricamente se ha intentado modelar la similitud y se han desarrollado modelos de razonamiento. La investigación reciente en informática se ha aplicado a las teorías de similitud computacional, como soporte de razonamiento para los procesos de recuperación y organización de la información. Los autores citan que las características de las medidas en el dominio geoespacial son: las teorías desarrolladas indican que la similitud es una relación asimétrica; los contextos tales como la edad, el conocimiento y vivencias culturales, la motivación y la propia aplicación son aspectos claves para medir la similitud; las teorías de similitud están ligadas al lenguaje utilizado para representarla; para que una teoría sea usable hay que garantizar que la representación computacional y las descripciones usadas por los participantes son comparables. Finalmente, los autores indican cuáles son los futuros campos de investigación en la medida de la similitud: explicación y aproximación de los valores de similitud, el estudio de la influencia de los contextos, la similitud entre objetos que no perduran indefinidamente, las diferencias entre similitud y analogía semántica y la similitud semántica dentro de los sistemas de referencia semánticos.

En esta tesis se propone un nuevo modo para medir la interoperabilidad basado en el uso de los metadatos. Los metadatos, en el contexto de las IDE, son componentes imprescindibles para describir los conjuntos de datos y los servicios. Como se ha indicado en el capítulo 2, los metadatos desempeñan las funciones de localización, evaluación, extracción y uso en este contexto gracias a la existencia de estándares que definen los ítems, su semántica y, en algunos casos, restringe el rango de valores. Desde nuestra perspectiva, los metadatos son elementos facilitadores de la interoperabilidad: describen la sintaxis, la semántica, definen el modo de acceder a los datos y las interfaces, catalogan los recursos, describen los aspectos legales y organizacionales relativos a los datos y servicios y su explotación, etc. y, por estas razones, alineados con la propuesta de Tolk, Diallo y Turnitsa (2007), se propone el uso de los metadatos como herramienta que permite medir la interoperabilidad en el contexto de las IDE.

### **3.7 Sumario**

En este capítulo se ha descrito la importancia de la interoperabilidad en el contexto de la información geográfica y en especial en el de las infraestructuras de datos espaciales. Se han revisados las definiciones del término interoperabilidad y se ha seleccionado una como marco de trabajo. Se han revisado los modelos de interoperabilidad en el contexto de la IG, de los *SoS* y organizacionales y se han revisado las definiciones o el alcance que los distintos autores otorgan a los niveles de interoperabilidad, para finalizar con la medida de la interoperabilidad como mecanismo de validación.

*La interoperabilidad como estandarización.* Uno de los objetivos de las IDE es facilitar el acceso, el uso, así como la reutilización de los datos. También desde el punto de vista de los servicios se pretende que los usuarios y las aplicaciones puedan explotarlos motivo por el que se precisa la estandarización. Desde el punto de vista tecnológico una IDE se ve como un conjunto de sistemas y como una II, al mismo tiempo que puede entenderse como un sistema en el que conviven muchos subsistemas o *SoS*. Para las IDE, las II y los *SoS* la interoperabilidad es un objetivo en sí misma.

*Revisión del concepto de interoperabilidad.* Se ha analizado el concepto interoperabilidad y se han interpretado las distintas definiciones. Tras el análisis se puede afirmar en que la interoperabilidad puede percibirse desde distintos puntos de vista. Por ejemplo, desde el punto de vista de las organizaciones de normalización, se presta especial atención a los aspectos técnicos, consistentes en el intercambio de recursos entre sistemas. Sin embargo, para los organismos y las instituciones políticas las definiciones propuestas para el término interoperabilidad deben poner el énfasis en la reutilización de la información.

*Revisión de los modelos.* Se han revisado los modelos de interoperabilidad en el contexto de la IG y de los SoS. El reto de la investigación en torno a los modelos de interoperabilidad consiste en examinar un modelo unificado, basado en las fortalezas de los diferentes modelos o pasarelas entre modelos de interoperabilidad (transformaciones) que permitan su integración.

La interoperabilidad es un problema complejo que debe de analizarse por partes y tanto las propias definiciones como los modelos de interoperabilidad, denominan a estas partes “*niveles*”. Se ha realizado una extensa revisión bibliográfica sobre los niveles de interoperabilidad en el contexto de los sistemas y de la IG. Esta revisión ha permitido detectar 15 niveles diferentes de interoperabilidad, con distinta frecuencia de aparición y relevancia al contexto geográfico y de las IDE.

*Medidas de la interoperabilidad.* Dado que la interoperabilidad es un objetivo para las IDE, también lo es el grado alcanzado entre los componentes y aplicaciones que operan sobre ellas. Por esta razón se ha revisado la literatura existente relacionada con la forma en la que puede medirse. En unos casos se trata de detectar los obstáculos y en otros de garantizar un determinado nivel. Finalmente se puede indicar que los principales esfuerzos se centran en torno a la interoperabilidad semántica y las métricas que permiten medir el grado alcanzado.

Una vez revisada la literatura en torno al concepto de interoperabilidad y a la taxonomía de los niveles de la misma, se procede a analizar algunos tipos de interoperabilidad en el contexto de las IDE para, posteriormente, proponer un modelo de interoperabilidad consistente en 7 niveles: técnica, sintáctica,

semántica, pragmática, dinámica, conceptual y organizacional. Esto se ha hecho tras analizar las definiciones propuestas por los autores y realizar un conjunto de reflexiones encaminadas a simplificar el modelo, eliminando aquellos niveles que no son de aplicación en este contexto, y mediante la reclasificación de otros por afinidad semántica.

De los tipos de interoperabilidad mostrados en este capítulo, se debe tener en cuenta que:

- Las interoperabilidades *física y empírica* propuestas por Assche (2006) atienden aspectos de la interacción hombre-máquina o de la cantidad de información expuesta al estudiante, en el contexto de la enseñanza electrónica (e-Learning). Dado que este tipo de interoperabilidad tiene poca aplicación en el contexto de las IDE como un caso de los SoS, se ha decidido no considerarlo en el modelo.
- Los niveles de interoperabilidad *internacionales, políticos/humanos, sociales y/o culturales, intracomunitarios, legales y de políticas* pueden ser englobados como aspectos que afectan a usuarios, instituciones y/u organizaciones. Los aspectos legales escapan, en muchas ocasiones, a las competencias de las instituciones, al ser éstos impuestos por normativas que emanan de estamentos superiores, como pueden ser las leyes nacionales para las comunidades autónomas o regiones estatales o las internacionales para las naciones. La colaboración y cooperación son aspectos que afectan tanto a las instituciones como a las organizaciones, permitiéndoles definir las interrelaciones a niveles regionales, nacionales, intracomunitarios e internacionales. Algunos de los autores citados (Harvey *et al.* 1999, Kuhn 2005, Assche 2006) hablan de las carencias de interoperabilidad a nivel social o la relacionan con los aspectos políticos, culturales y los valores. Paralelamente, el grupo de expertos (Goodchild *et al.* 1997) consideran la interoperabilidad organizacional como la más difícil de alcanzar.
- Los aspectos *esquemáticos o estructurales* citados por Nowak y Nogueras (2005), Goh (1997) y Shekhar (2004) pueden verse incluidos en otros tipos de interoperabilidades. Así, por ejemplo, Nowak y Nogueras (2005) identifican como falta de interoperabilidad esquemática o estructural a las diferencias en los modelos de datos. Esta falta de interoperabilidad también puede ser

considerada como carencia *conceptual*, dado que en muchos casos la definición de los propios modelos de datos no se describe mediante lenguajes de modelado que independicen el modelo de su implementación. Goh (1997) y Shekhar (2004) también tratan estos aspectos, si bien pueden ser considerados como aspectos sintácticos y conceptuales, al considerar los modelos de datos y los metadatos como mecanismos que permiten compartir los esquemas de los datos.

Para finalizar esta justificación, en el resumen del capítulo, resaltar que también se han considerado para definir el modelo compuesto por siete niveles, los siguientes modelos de interoperabilidad: *Integrated Interoperability in GIS* (Shanzhen Yi *et al.* 1999; Goodchild *et al.* 1997), *Coalition Model* (Tolk 2003) e *Integrated Interoperability Model* (Turnitsa y Tolk 2006).

## 4 METODOLOGÍA

### 4.1 Introducción

En este capítulo se presenta la metodología propuesta para la creación automática de metadatos, se describe el modelo de interoperabilidad integrado que se ha diseñado para las IDE y, finalmente, se presenta la metodología adoptada para analizar la interoperabilidad que aportan los ítems de la norma internacional de metadatos ISO19115.

El primer lugar se trata de responder a la tercera cuestión de investigación planteada en el proyecto de tesis: ¿Se pueden crear metadatos útiles de la información geográfica de un modo automático y eficaz? Se describe la metodología que permite extraer los metadatos almacenados en los datos ya sea de forma implícita o explícita, se describe las distintas fases de análisis y tratamiento que se realiza con la información obtenida y las fases de inferencia que permiten proponer otro conjunto de ítems que los catalogan o describen su modelo de datos.

En segundo lugar se presenta el modelo integrado de interoperabilidad propuesto para su aplicación en el contexto IDE. Se justifica la elección de los niveles del modelo y se describen de un modo general los objetivos de dichos niveles.

En tercer lugar se trata de responder a la segunda cuestión de investigación planteada en el proyecto de tesis: ¿Cuál es la aportación en términos de interoperabilidad de la información contenida en los metadatos? Para responder a esta cuestión se ha realizado un análisis exhaustivo de los ítems que conforman el núcleo fundamental de la norma por un lado y del conjunto de ítems de la misma por otro. Además de analizar a nivel agregado cómo proporcionan interoperabilidad a distintos niveles los ítems se ha analizado cómo favorecen simultáneamente varios niveles.

Finalmente se presentan las conclusiones del capítulo.

## 4.2 Metodología de creación automática de metadatos

Como se ha evidenciado y justificado en los anteriores apartados, la creación automática de metadatos (Greenberg 2004) es necesaria y útil, debe de integrarse en el flujo de trabajo de los datos y puede requerir grandes desarrollos informáticos. A continuación se presenta una metodología desarrollada en forma de procedimiento, que es susceptible de ser automatizada mediante un sistema informático para la creación de metadatos sobre repositorios o almacenes de datos espaciales. Se describen las etapas de que consta y los resultados que se pueden obtener. La figura 4.1 muestra un diagrama de bloques con las distintas etapas de la metodología y sus relaciones. Desde un punto de vista global, el método consiste en extraer la mayor cantidad de información posible del repositorio (3) de IG para cumplimentar un metadato en el modo más completo y detallado posible.

La metodología prevé tanto el caso de que el repositorio de IG no almacene información que identifique el tipo de coordenadas utilizado, como que se disponga de una plantilla de metadatos que completar. Además, se prevé que el usuario o aplicación que explote esta metodología pueda seleccionar el tipo de resultado deseado: un archivo *XML* (17) con el metadato, o un archivo comprimido *ZIP* con formato *MEF* (18)<sup>1</sup>, conteniendo tanto el metadato como los archivos adicionales generados: vista previa (12), modelo de datos *UML* (13) y esquema de aplicación *XSD* (15).

La figura comienza con la entrada (1) en la que, con una tonalidad más clara de texto, se muestran aquellos elementos que son opcionales, esto es, que pueden ser proporcionados o no serlo. El único elemento necesario y, por tanto obligatorio, es el *URI*, que describe dónde y cómo acceder al repositorio de IG. Se ha definido de este modo para permitir conectar con distintos protocolos (*http*, *ftp*, *file*) de transferencia y con distintos almacenes y bases de datos (*oracle*, *postgres*, *mysql*, *db2*, *informix*, *sqlite*, etc.). Estos últimos pueden requerir datos adicionales relativos al sistema de identificación basado en usuario y contraseña, además del identificador de la propia base de datos en el gestor.

---

<sup>1</sup> MEF (Metadata Exchange Format) definido en el proyecto GeoNetwork. Enlace: <http://geonetwork3.fao.org/workshops/workshop2007/GeoNetwork%20MEF%206-11-2007.pdf>

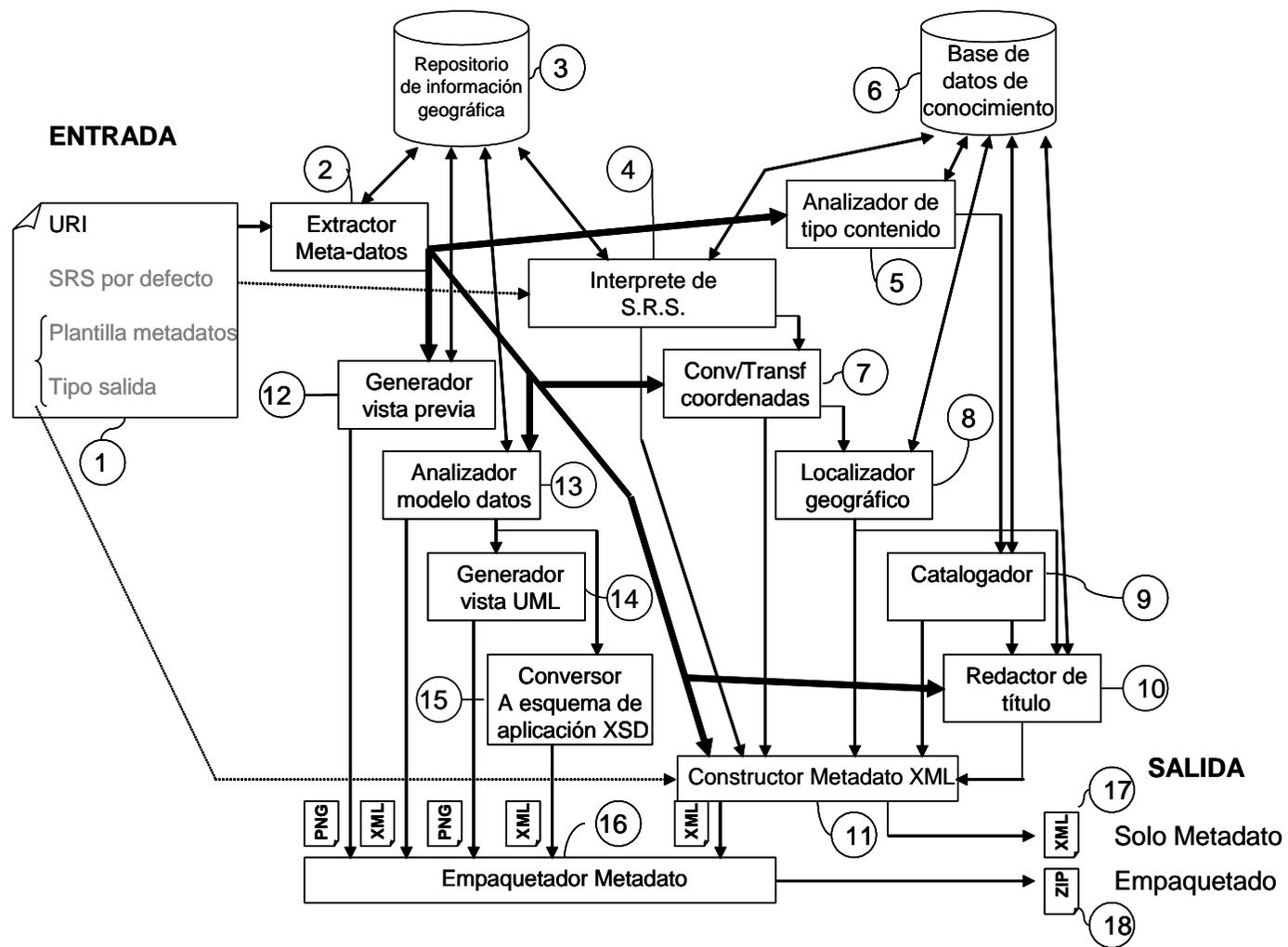


Figura 4. 1: Método de creación automática de metadatos: etapas y relaciones

La primera etapa de la metodología consiste en la extracción de metadatos; el *Extractor de Metadatos* (2) utiliza la información contenida en el *URI* (1) para acceder al *Repositorio de IG* (3) y, de este modo, obtener la mayor cantidad de información implícita del mismo. La información recuperada por el *Extractor de Metadatos* (2) dependerá del formato de almacenamiento usado por el repositorio de IG (Manso *et al.*, 2004). Se han podido identificar más de 90 formatos de almacenamiento para imágenes y datos matriciales (ráster), más de 25 formatos de almacenamiento de datos vectoriales o de dibujo y más de 10 sistemas de gestión de bases de datos (SGBD), con capacidad de almacenar información espacial. A continuación se particulariza, para cada formato, el procedimiento de extracción de información almacenada en el archivo o BBDD cuando son útiles para construir el metadato:

- El procedimiento general para las bases de datos consiste en acceder de forma recursiva a todas las tablas de la misma para: leer los nombres y tipos de columnas, la columna que contiene las geometrías, el tipo de geometría/s almacenadas, el número de registros de la tabla y los valores máximos y mínimos de las coordenadas.
- El procedimiento general para los archivos vectoriales es similar a las bases de datos, buscando las mismas propiedades, aunque el acceso es más complejo, ya que se mezclan en una misma capa distintos tipos de datos (equivalente a tabla) con distintas geometrías. Un caso particular de los formatos vectoriales son los formatos CAD, aún más complejos por las distintas soluciones adoptadas para codificar los elementos, hecho que dificulta la obtención de información fiable sobre el tipo de fenómenos geográficos contenidos en el mismo.
- El procedimiento general para las imágenes consiste en recuperar la información relativa al nº de bandas, el tipo de dato o nº de bits utilizados para cada punto de imagen (píxel), las dimensiones de la imagen (ancho y alto), las coordenadas máximas y mínimas, la resolución del píxel y los estadísticos media y desviación típica de cada banda.
- El procedimiento general de las mallas es similar al de las imágenes, ya que podrían considerarse un caso particular de ellas; sólo hay una banda y suele contener valores numéricos que representan atributos tales como altimetría o identificador de la clase a la que se ha asignado, y representan una superficie

rectangular. Para las mallas es interesante obtener los valores máximos y mínimos de la banda, además de los ya enumerados para las imágenes.

Esta información será utilizada por el analizador de tipo de contenido (5) para tratar de determinar qué tipo de contenido almacenan las imágenes y, consecuentemente, utilizarlo para catalogar la IG.

La complejidad de esta etapa de extracción se encuentra en el elevado número de formatos de repositorios de datos geográficos existentes y la falta de uniformidad en relación con el tipo, cantidad y forma de almacenar la información implícita (que puede ser tratada con un procedimiento) y la explícita (que ha de ser tratada de forma individualizada), que describe los datos. Se puede afirmar que hay un número significativo de formatos para los que no se puede aplicar un procedimiento común para recuperar la información (explícita) adicional que contienen.

La segunda etapa en importancia es el *Intérprete de Sistema de Referencia por Coordenadas* (SRC) (4). Al extraer del *Repositorio de IG* (3) la información que identifica el sistema de coordenadas utilizado, se puede encontrar (Manso y Bernabé, 2005), que existen distintos tipos de codificaciones para representar dicha información:

- Como información textual, con una estructura predefinida (p.e. "European\_Datum\_1950\_UTM\_Zone\_30N");
- Como estructuras de texto bien conocidas (WKT) (p.e. PROJCS[ED50 / UTM zone 30N,GEOGCS["ED50", DATUM ["European\_Datum\_1950", SPHEROID ["Internacional 1924",6378388,297, AUTHORITY["EPSG","7022"]], AUTHORITY ["EPSG","6230"]], PRIMEM["Greenwich",0,AUTHORITY["EPSG","8901"]], UNIT ["degree", 0.017453292519943 28,AUTHORITY[ (..)"]);
- Como mnemotécnicos (p.e. Ermapper datum ED50 y proyección NUTM30);
- Como códigos numéricos particulares (p.e. Intergraph datum: 4, projection: 7);
- También es usual que algunos utilicen secuencias de pares atributo-valor (p.e. Proj4 +proj=tmerc +lat\_0=0 +lon\_0=93 +k=1.000000 +x\_0=16500000 +y\_0=0 +a=6378140 +b=6356755.288157528 +units=m +no\_defs);

- Finalmente y coincidiendo con la tendencia actual, varios repositorios utilizan códigos numéricos estandarizados (p.e. *EPSG:23030*).

El objetivo de esta etapa consiste en identificar, en el formato usado por el repositorio de información espacial, el tipo de codificación utilizado, obtener los valores almacenados y relacionar dichos valores con los códigos numéricos estandarizados por el *European Petroleum Surveyor Group (EPSG)*, hoy integrado en la Asociación internacional de productores de gas y petróleo (OGP). La última operación de este proceso requiere explotar la información almacenada en la base de datos de conocimiento (6), lugar en el que se han almacenado las relaciones entre las codificaciones particulares de cada formato y los códigos *EPSG*. El resultado será incorporado en el metadato y usado para realizar los cálculos de conversión/transformación de coordenadas que se requieran.

La complejidad del intérprete de Sistema de Referencia por coordenadas reside en la identificación correcta del tipo de coordenadas y el datum utilizado en el repositorio para asignarle el código *EPSG*.

El *Generador de Vista Previa* (12) tiene como objetivo generar una imagen en formato *PNG*, con la visualización de la toda la información contenida en el repositorio de IG de una capa o conjunto de bandas. Esta imagen formará parte de la información adicional que acompaña al metadato y que puede ayudar a un usuario a decidir si el conjunto de datos localizado satisface las necesidades previstas. Al igual que le ocurre con el extractor de metadatos, donde la complejidad de este bloque reside en la gran variedad y heterogeneidad de los formatos de almacenamiento de IG existentes.

La etapa de *Conversión/Transformación de Coordenadas* (7) tiene por objeto obtener las coordenadas geográficas, utilizando como *datum* de referencia el sistema geodésico global de 1984 (*WGS84*) asociadas a la envolvente geográfica (norte, sur, este y oeste) de la capa o banda de información a partir de las coordenadas expresadas en el SRC (4) almacenado en el repositorio de datos. Como puede observarse en la figura 4.1, se requiere conocer las coordenadas almacenadas y el sistema de referencia utilizado para realizar el cálculo apropiado: transformación (con cambio de *datum*) o conversión (manteniendo el

*datum*). La característica destacable de este bloque es que debe ser capaz de interpretar muchos tipos de SRC definidos por su código *EPSG*, y saber realizar las transformaciones o conversiones necesarias para pasar de unos a otros. Las coordenadas obtenidas tras la conversión o transformación, además de formar parte del metadato, serán utilizadas como entrada del *Localizador Geográfico* (8).

El *Analizador de Modelo de Datos* (13) interpreta el tipo de repositorio de IG para determinar la tipología de los datos. Cuando la IG tiene asociada información alfanumérica en forma tabular, se trata de determinar las posibles relaciones entre los atributos de las tablas, para sugerir un modelo de datos basado en entidades y relaciones. Este trabajo es más sencillo cuando el repositorio de IG es una base de datos, ya que la mayoría de los SGBD almacenan metadatos correspondientes a las tablas, atributos y las relaciones (claves primarias, claves foráneas, restricciones, etc.). En el caso de archivos con estructura SIG y tablas o bases de datos alfanuméricas con tablas, esta fase del proceso se vuelve más compleja debiendo inferir las relaciones, ya que éstas no están definidas explícitamente. El modelo de datos deducido se representa en Lenguaje Unificado de Modelado (*UML*) y se almacena en un formato *XML* de intercambio de metadatos (*XMI*), para que pueda ser tratado de forma automática por otras aplicaciones. Además, el modelo de datos deducido será la fuente de información que utilizarán las etapas que generan una vista gráfica del diagrama de clases *UML* (13) y la que genera el conversor a formato de esquema de aplicaciones *XSD* (15).

El *Localizador Geográfico* (8) utilizará las latitudes y longitudes que delimitan la envolvente geográfica, para determinar los identificadores geográficos más relevantes de la zona e incluirlos en el metadato. Esta identificación se realizará en base a una cobertura que contiene los polígonos, con las delimitaciones geográficas de los países y regiones más importantes a nivel mundial, almacenados en la *Base de Datos de Conocimiento* (6). Se trata de una consulta espacial cuyos resultados se ordenan por relevancia de la región y población, realizando la función de un *Geocoder* inverso que recupera un topónimo en base a las coordenadas de la extensión geográfica. La complejidad del localizador geográfico radica en la necesidad de acceder o explotar tesauros especializados en

nombres de lugares. El identificador geográfico así calculado será incluido en el metadato y será utilizado por el *Redactor de Título* (10).

El *Generador de Vista UML* (14) construye una imagen representando el diagrama de clases contenido en el modelo de datos, y lo almacena en formato gráfico *PNG*. Dicho gráfico facilita la interpretación del modelo de datos por parte de los usuarios, ya que es más sencilla la interpretación del un modelo *UML* de forma gráfica que en lenguaje *XML*.

El *Catalogador* (9) utiliza la información recuperada por métodos de clasificación, minería de datos e inferencia, junto con las listas de términos pertenecientes a tesauros multilingües, para proponer conjuntos de palabras clave (*keywords*) pertenecientes a Tesauros multilingües que cataloguen el repositorio de IG. Además de sugerir un conjunto de palabras clave, esta fase del procedimiento selecciona uno o varios términos de la lista controlada que identifica la categoría temática (*topicCategory*) del recurso. Los resultados de esta fase del procedimiento se incluyen en el metadato y también son usados posteriormente por el *Redactor de Título* (10).

El *Convertor a Esquema de Aplicación XSD* (15) realiza las transformaciones necesarias para representar el modelo de datos *UML* en forma de esquema de aplicación *GML*, almacenándolo en formato *XML Schema (XSD)*. Este esquema de aplicación modela el repositorio o capa de información en el lenguaje estandarizado por el *Open Geospatial Consortium (OGC)* y la organización internacional de estandarización (ISO), para describir los modelos de datos geográficos en lenguaje de marcas (*GML*). El resultado de esta fase del procedimiento genera un archivo *XML* que acompañará al metadato.

El *Redactor de Título* (10) compone una frase que intenta sintetizar, en la medida de lo posible, la descripción de los datos. El título debe de contener información suficiente para responder a un buen número de preguntas tales como: ¿Qué?, ¿Dónde?, ¿Cuándo?, ¿De quién?, ¿Qué escala?, etc. Para componer el título, esta fase del proceso utilizará la información obtenida por el localizador geográfico (8)

(¿Dónde?), la clasificación realizada por el *Catalogador* (9) (¿Qué?) y la información almacenada en la *Base de Datos de Conocimiento* (6).

El *Constructor de Metadato XML* (11) ensambla toda la información proporcionada, extraída, calculada, inferida y elaborada que se ha descrito en las anteriores etapas. Si, como entrada del procedimiento, se dispone de un metadato o plantilla de metadato, esta etapa aportará sobre él dicha información. En caso contrario, construirá un nuevo metadato e insertará en él la información. El resultado de esta etapa es un archivo *XML*, conforme con la norma internacional de metadatos ISO19115-19139.

El *Empaquetador de Metadato* (16) toma el archivo *XML* con el metadato ensamblado por el constructor de metadatos, los archivos *XML* conteniendo el diagrama de clases *UML*, y el esquema de aplicación *GML* y los archivos gráficos, conteniendo la vista previa de los datos y del diagrama de clases y construye una estructura de archivos y directorios en los que se integra toda la información. Finalmente comprime en un documento con formato *Zip* (18) toda la estructura de archivos y directorios. La estructura interna de la información contenida en el archivo es conforme a la especificación de formato de intercambio de metadatos (*MEF*) propuesta por el proyecto *GeoNetwork*.

### **4.3 Modelo de interoperabilidad aplicado en las IDE**

Nuestro modelo de interoperabilidad consta de 7 niveles: Técnico, Sintáctico, Semántico, Pragmático, Dinámico, Conceptual y Organizacional.

En cuanto a las relaciones existentes entre los distintos niveles de interoperabilidad, nuestro modelo propone que dichas relaciones no tienen por qué ser jerárquicas, como se propone en la bibliografía revisada. Nuestro modelo apunta a que existen distintos tipos de relaciones entre los distintos niveles de interoperabilidad (Figura 4.2). Esta asunción se fundamenta en algunas evidencias o reflexiones, como por ejemplo que para alcanzar la interoperabilidad conceptual, que se fundamenta en los modelos de datos y los esquemas de

aplicación, son importantes las interoperabilidades sintáctica y semántica y, por el contrario, las interoperabilidades pragmática y dinámica no lo son. Para expresar los conceptos se necesita la sintaxis y la semántica, sin embargo un concepto no encuentra necesidad de los niveles pragmáticos y dinámicos, salvo que exista un repositorio de conceptos o modelo de datos y dichas interoperabilidades sean necesarias para acceder al mismo. Otro ejemplo similar versa sobre los aspectos legales, tales como la propiedad intelectual o las restricciones de acceso y/o uso de los recursos. Parece lógico pensar que para expresar o concretar estos aspectos legales se precisan las interoperabilidades sintáctica y semántica: el concepto y su concreción y, no son relevantes las interoperabilidades conceptual, dinámica o pragmática. Estos dos ejemplos revelan que las relaciones de dependencia entre los niveles de interoperabilidad no tienen una naturaleza jerárquica.

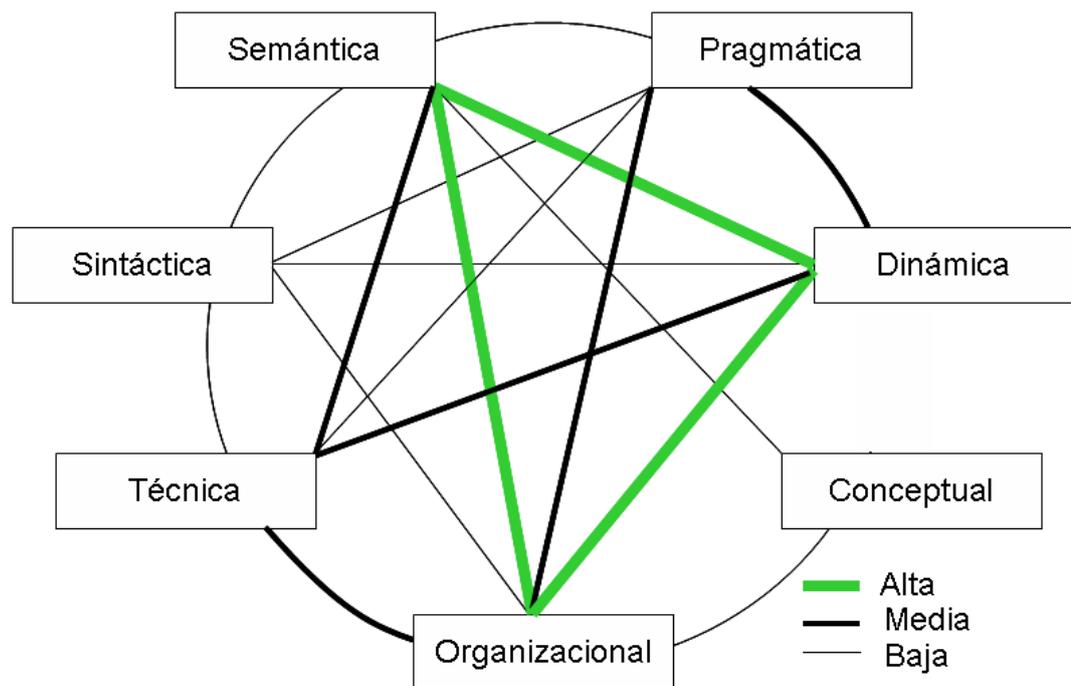


Figura 4. 2: Propuesta sobre modelo integrado de interoperabilidad

Fundamentado en la propuesta de modelo integrado de interoperabilidad se pretende analizar los niveles de interoperabilidad que facilitan los ítems de metadatos que describen un recurso, de la misma forma que propuso Shanzhen (1999) en el modelo Intermodel5 con las funciones de los metadatos. Por esta razón, asumimos como premisa que los ítems de metadatos puedan ser usados para construir y describir las relaciones entre los niveles de interoperabilidad en el modelo integrado, siendo ésta nuestra hipótesis.

Algunos autores como Tolk, Diallo y Turnitsa (2007) destacan la utilidad de los metadatos para alcanzar la interoperabilidad entre sistemas, en el contexto del modelo LCIM. Se cita, por ejemplo, la utilidad de los metadatos en las comunicaciones entre agentes de *software* inteligentes para ‘comunicar acerca de las situaciones’, ‘permitirles la selección de distintos componentes y evaluar la composición de los mismos’, ‘apoyos en la toma de decisiones’ y finalmente ‘apoyar en la composición y orquestación de componentes de un modo ágil, al menos hasta el nivel dinámico’.

Tanto la referencia de los autores Tolk, Diallo y Turnitsa (2007) como la hipótesis del modelo integrado de interoperabilidad, han motivado que clasifiquemos de los metadatos de acuerdo a los niveles que sus ítems proporcionan, de modo que se pueda inferir el tipo de relaciones existentes entre los niveles del modelo de interoperabilidad para las IDE, fundamentado en los metadatos.

A continuación se definen los objetivos o alcances de los niveles del modelo.

#### **4.3.1 Interoperabilidad Técnica**

Como Turnitsa y Tolk (2006) y por similitud de modelos, definimos interoperabilidad técnica como aquella que posibilita la interconexión de los sistemas a nivel de protocolos, y el intercambio de información en su nivel más básico: bits.

Se han identificado algunos ejemplos relacionados con las IDE como pueden ser: los juegos de caracteres y la codificación de los caracteres usados en los datos, los identificadores de archivos, la descripción del entorno de procesamiento, los nombres de los archivos, tipos de servicios y sus versiones, el tamaño de transferencia de los datos, los formatos y versiones de los archivos, medios de almacenamiento, así como los protocolos y los puntos de acceso a los servicios.

#### **4.3.2 Interoperabilidad Sintáctica**

Es aquella que posibilita el intercambio de información en un formato común, incluyéndose en este tipo de interoperabilidad los relacionados con los estándares de datos (formatos), que posteriormente han de intercambiar los sistemas.

Se entiende por “aspectos de la interoperabilidad sintáctica” los formatos estandarizados de intercambio de información. Este es el caso del formato XML y las reglas que definen la estructura de los datos en forma de esquemas (XSD), para todo tipo de información alfanumérica. También lo son los formatos gráficos de imágenes (JPEG, PNG, GTIFF) en otros casos. En el contexto de las IDE, serán aspectos de la interoperabilidad sintáctica, todos los esquemas XML definidos por el *OpenGeospatial Consortium* (OGC) para aplicaciones y servicios Web (WMS, WFS, WCS, CS-W, WPS, SOS), además de los formatos para la codificación de los datos (GML, O&M, SensorML, TML), así como la definición de los estilos de presentación de los objetos (SLD y SE), o la sintaxis para definir los filtros (FE). Todos ellos han sido identificados como iniciativas que incrementan la interoperabilidad sintáctica.

### **4.3.3 Interoperabilidad Semántica**

Es aquella que posibilita el intercambio de información, utilizando un vocabulario común y compartido, que evite las inexactitudes en la interpretación del significado de los términos.

Se entiende por “aspectos de interoperabilidad semántica” a los estándares y/o especificaciones que definen los esquemas de intercambio de información y el significado de cada uno de los ítems sin ambigüedades. Por ejemplo *Web Service Description Language* (WSDL) y *Simple Object Access Protocol* (SOAP), a nivel de interconexión de los servicios, *Geographic Mark-Up Language* (GML) en el transporte de IG vectorial, *Symbology Encoding* (SE), que define como codificar la simbología de los objetos, *Style Layer Description* (SLD), que define como usar la simbología conjuntamente con el servicio WMS, *Common Query Language* o *Filter Encoding* (CQL - FE) para las consultas y los filtros, etc. De modo similar, las normas de la familia ISO-19100 incluyen un apartado con la definición de los términos y listas controladas de términos, que ayudan a compartir un significado común. Un caso particular es la norma de metadatos ISO19115 (2003), que contiene al menos 24 listas controladas de términos enumerados. Otro caso es la directiva Europea INSPIRE, que también está adoptando listas de términos controladas para clasificar los tipos de datos (*Data Themes*), descritos en los anexos en las categorías definidas por ISO9115 (2003), para la clasificación de los recursos.

#### **4.3.4 Interoperabilidad pragmática**

Es aquella que posibilita que los sistemas conozcan y exploten los métodos y procedimientos que proporcionan los demás sistemas.

Se entiende por aspectos de interoperabilidad pragmática a los estándares y especificaciones que definen las taxonomías de servicios y sus interfaces de explotación. Por ejemplo, el estándar ISO19128 (2005), las especificaciones de servicios OGC (*WFS: Web Feature Service*, *WCS: Web Coverage Service*, *CS-W: Catalog Service Web*, *SOS: Sensor Observation Service*, *WNS: Web Notification Service*, *WAS: Web Alert Service*, *LBS: Location Based Services*, etc.), para los que la norma o las especificaciones respectivamente, definen las interfaces que posibilitan la explotación y los parámetros que son capaces de manejar. Un aspecto característico del contexto IDE (y OGC) relativo a los servicios es la operación obligatoria *getCapabilities*, que ofrecen todos. Esta operación permite consultar a cualquier servicio por sus capacidades, obteniendo una descripción de los mismos, de las operaciones implementadas y aportan el punto de acceso para su explotación. Esto también ocurre, de una forma similar aunque menos enriquecedora, en los Servicios Web (WS) y el lenguaje de descripción de servicios WSDL.

#### **4.3.5 Interoperabilidad dinámica**

Es aquella que permite a los sistemas autocorregir su funcionamiento ante los cambios en la transferencia de información, y obtener beneficio de ello.

Se entiende por “aspectos de interoperabilidad dinámica” aquellos que tienen la capacidad de sustituir dinámicamente un servicio por otro, si es inaccesible o si la calidad del servicio no cubre las necesidades. En este sentido, los sistemas deben de disponer de mecanismos que les permitan descubrir dinámicamente la existencia de servicios que cumplan los requisitos solicitados. A priori, este nivel de interoperabilidad requiere una explotación de la componente semántica que posibilite el descubrimiento de los servicios, en base a información que les describe (metadatos). La definición de taxonomías de servicios, como hace la norma ISO19119 con el concepto *Service Organizer Folder* (SOF) o las reglas de implementación de los metadatos promovidas por la directiva INSPIRE (2008), establece el uso de un conjunto de identificadores para los servicios que pueden fomentar este tipo de interoperabilidad. Así, la directiva INSPIRE en las reglas de implementación, promueve la clasificación de los servicios por su función en: descubrimiento, visualización, descarga y transformación. La norma ISO19119 clasifica

los servicios como: interacción humana, gestión del modelo, gestión de tareas, de procesamiento, de comunicaciones y de sistemas. Adicionalmente, la directiva INSPIRE (2008) establece una clasificación de los tipos de datos en los que se clasifican en las categorías definidas por ISO19115 (2003) (*topicCategory*), facilitando el intercambio de servicios de un modo dinámico, al menos desde un punto de vista teórico.

#### **4.3.6 Interoperabilidad conceptual**

Es aquella que permite conocer y reproducir el funcionamiento de un sistema en base a la documentación expresada en un formato usado en ingeniería.

Se entiende por “aspectos de interoperabilidad conceptual” aquellos que describen los modelos de datos y sistemas en forma de documentación estandarizada e intercambiable, desde un punto de vista de la ingeniería, sin depender del modelo utilizado para describirlo. La descripción mediante *UML* del modelo de datos, ya sea de un almacén de datos o el proporcionado por un servicio Web, posibilita este tipo de interoperabilidad. Algunas de las especificaciones de servicios del OGC como WFS (ISO19142), proporcionan una descripción conceptual de las entidades ofrecidas por el mismo, como respuesta a las operaciones *describeFeatureType*. En estos casos se está recuperando una descripción conceptual de la entidad en forma de esquema de aplicación GML. Actualmente, en las reglas de implementación de la directiva INSPIRE para los tipos de datos, se están definiendo los modelos utilizando herramientas CASE, que permiten intercambiar dichos modelos de clases y restricciones en lenguaje UML, utilizando formatos estandarizados de intercambio como XMI (por ejemplo *Addresses*).

#### **4.3.7 Interoperabilidad organizacional**

Es aquella que permite conocer los objetivos de negocio, los modelos de procesos, las leyes y políticas de acceso, así como el uso de los datos y los servicios.

Se entienden por “aspectos de interoperabilidad organizacional” aquellos que posibilitan conocer y entender las políticas de acceso y uso de datos y/o servicios, las responsabilidades personales o institucionales, así como los objetivos y los fines definidos por la organización, al crear un determinado tipo de dato o proporcionar un tipo de servicio sobre los mismos. Gran parte de la información relacionada con el propósito, las responsabilidades, las políticas de uso y acceso es considerada de identificación, o que

describe las restricciones y se trata de información útil para evaluar el uso de los recursos descritos por los metadatos.

#### **4.4 Clasificación de los ítems de metadatos según el modelo de interoperabilidad**

Se trata de analizar el papel que pueden desempeñar los metadatos en los distintos niveles de interoperabilidad del modelo propuesto. Para realizar este análisis se ha propuesto la clasificación de los distintos ítems que componen un metadato sobre información geográfica ISO19115/19139, atendiendo a los niveles de interoperabilidad que pueden facilitar.

Dado que la norma de metadatos ISO19115:2003 puede contener, potencialmente, más de 400 ítems (ANZLIC 2005), dependiendo del tipo de información geográfica que se describa, se ha optado por realizar el análisis en dos etapas. En la primera, el análisis sólo se aplica a los ítems del núcleo fundamental de la norma (*core*) y en la segunda se extiende el análisis a toda la norma.

Así, para el primer caso se han agrupado los ítems que componen el núcleo de la norma en una hoja de cálculo, y sobre la misma se ha analizado la interoperabilidad que facilita cada ítem. Para el segundo caso, los ítems de la norma de metadatos han sido tratados de acuerdo con los paquetes/clases que define la norma: identificación, calidad, restricciones, mantenimiento, distribución, sistema de referencia, esquema de aplicación, representación espacial e información del contenido.

Una vez diseñado y realizado el análisis, se han analizado a nivel descriptivo los resultados de la clasificación con dos alcances: a) las intensidades de los niveles de interoperabilidad asignados a los ítems, b) las intensidades de las relaciones entre parejas de niveles de interoperabilidad del modelo.

El primer análisis ha consistido en medir las intensidades/frecuencias de los niveles de interoperabilidad proporcionados por los niveles del modelo en el ámbito del estudio (ítems del núcleo fundamental o todos los ítems de la norma). El segundo ha consistido en medir las intensidades de las relaciones entre niveles

del modelo o las frecuencias de los ítems de los metadatos que proporcionan simultáneamente dos niveles de interoperabilidad, también en el ámbito del estudio.

A continuación, (1) se identifican los niveles de interoperabilidad de los ítems de metadatos, (2) se analizan los niveles proporcionados por los ítems de metadatos pertenecientes al núcleo fundamental y a toda la norma, y (3) se comparan los resultados. Después, se analizan las intensidades de las relaciones entre los niveles de interoperabilidad, repitiéndose el procedimiento; en primer lugar, los ítems del núcleo fundamental y después toda la norma para, posteriormente, comparar los resultados. Para finalizar este apartado se interpretan las relaciones entre los niveles del modelo integrado de interoperabilidad en el contexto IDE.

#### **4.4.1 Identificación de los niveles de interoperabilidad de los ítems de metadatos**

La identificación de los niveles de interoperabilidad que proporcionan los ítems de metadatos de la ISO19115 se ha realizado en una hoja de cálculo por cada paquetes/clase de ítems existente en la norma de metadatos, de modo que el tratamiento y su posterior análisis sea sencillo y completo. En estas hojas se han incorporado los nombres de los 52 ítems, una descripción resumida y siete columnas, en las que se identifican los niveles del modelo, como puede observarse en la figura 4.3. La primera columna incluye el nombre del ítem; la segunda, una descripción del mismo; y las siete siguientes sirven para identificar los niveles de interoperabilidad proporcionados por el ítem. Finalmente, en una columna que no se muestra en la figura 4.3, se presenta una descripción que justifica la elección de los niveles de interoperabilidad. El color rojo del fondo de una celda indica la obligatoriedad del ítem y el naranja la condicionalidad (obligatorio en ciertas condiciones).

Paquete/Clase	MD_Metadata	Tipos de Interoperabilidad						
Ítem	Descripción	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
fileIdentifier	Identificador del fichero	x						
language	Idioma		x					x
characterSet	Conjunto de caracteres	x	x					
parentIdentifier	Identificador del padre	x						
hierarchyLevel	Nivel jerárquico		x				x	x
hierarchyLevelName	Nombre del nivel jerárquico							x
contact (CI_ResponsableParty)	Contacto							
dateStamp	Fecha de Creación			x		x		x
metadataStandardName	Norma de Metadatos		x	x				x
metadataStandardVersion	Versión de la Norma de Metadatos		x	x				x
dataSetURI	Uri del Conjunto de Datos	x		x	x			
locale (PT_Locale) (ISO19139)	Información sobre alternativa lingüística							
spatialRepresentationInfo (MD_SpatialRepres	Información sobre la Representación Espacial							
referenceSystemInfo (MD_ReferenceSystem)	Información del Sistema de Referencia							
metadataExtensionInf (EX_Extent)	Información de Extensiones de Metadatos							
identificationInfo (MD_identification)	Información de identificación							
contentInfo (MD_ContentInformation)	Información del Contenido							
distributionInfo (MD_Distribution)	Información de Distribución							
dataQualityInfo (DQ_DataQuality)	Información sobre Calidad de los datos							
portrayalCatalogueInf (MD_PortrayalCatalogu	Información del Catálogo de Representación							
metadataConstraints (MD_Constraints)	Constricciones de los metadatos							
applicationSchemaInf (MD_ApplicationScher	Información del Modelo de Aplicación							
metadataMaintenance (MD_MaintenanceInfo)	Mantenimiento de los metadatos							
<b>PT_Locale</b>	<b>Identificador de idioma, país y tipo de codificación</b>							
language	Idioma			x				x
country	País			x				x
characterEncoding	Juego de caracteres de la codificación	x	x					

Figura 4. 3: Esquema de datos del documento de análisis de interoperabilidad

A continuación se presentan los resultados y la interpretación de los mismos.

La tabla 4.1 muestra las intensidades de los niveles de interoperabilidad de los ítems del núcleo fundamental de la norma y la figura 4.4 muestra dichos valores en forma de histograma.

Nivel de interoperabilidad	Total de ítems 52	%
Técnica	8	15,38
Sintáctica	3	5,7
Semántica	40	77
Pragmática	3	5,7
Dinámica	31	59,6
Conceptual	1	1,9
Organizacional	43	82,7

Tabla 4. 1 Frecuencias de los niveles de interoperabilidad en los ítems del núcleo fundamental

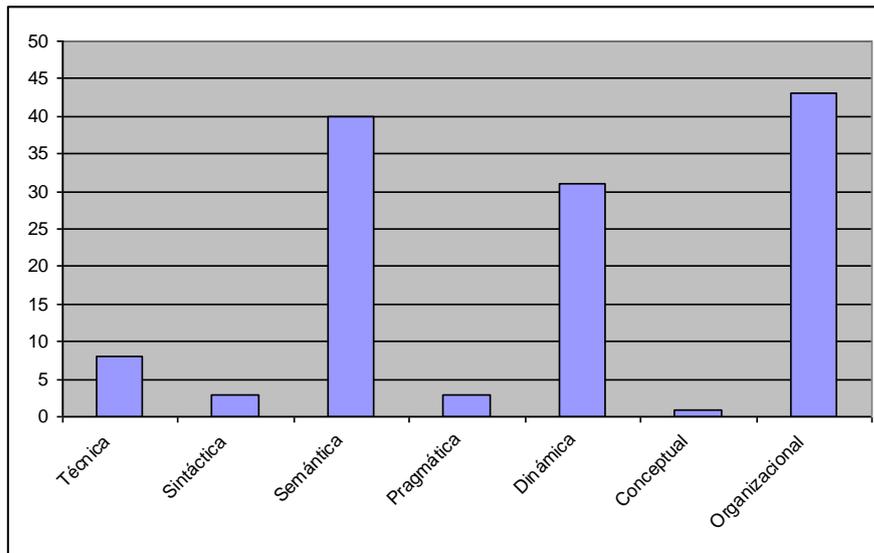


Figura 4. 4: Histograma de los niveles de interoperabilidad en los ítems del núcleo ISO19115

En la figura 4.4 se puede observar que los niveles organizacional, semántico y dinámico son claramente los más favorecidos por los ítems del núcleo de la norma. Aproximadamente el 80% de los ítems dotan de interoperabilidad organizacional y semántica, el 60% dinámica y el resto de niveles de interoperabilidad se ven poco favorecidos con porcentajes entre 15% y el 2%. Ante estos resultados surge la incertidumbre sobre los resultados que arrojará el análisis del conjunto de ítems que define la norma de metadatos: ¿Serán similares los resultados?. La tabla 4.2 y la figura 4.5 muestran los resultados de dicha clasificación.

Nivel de interoperabilidad	Total de ítems 235	%
Técnica	37	16
Sintáctica	6	2,6
Semántica	128	54
Pragmática	15	8,7
Dinámica	99	42
Conceptual	7	3
Organizacional	217	92

Tabla 4.

2:

Frecuencias de los niveles de interoperabilidad en los ítems de la norma

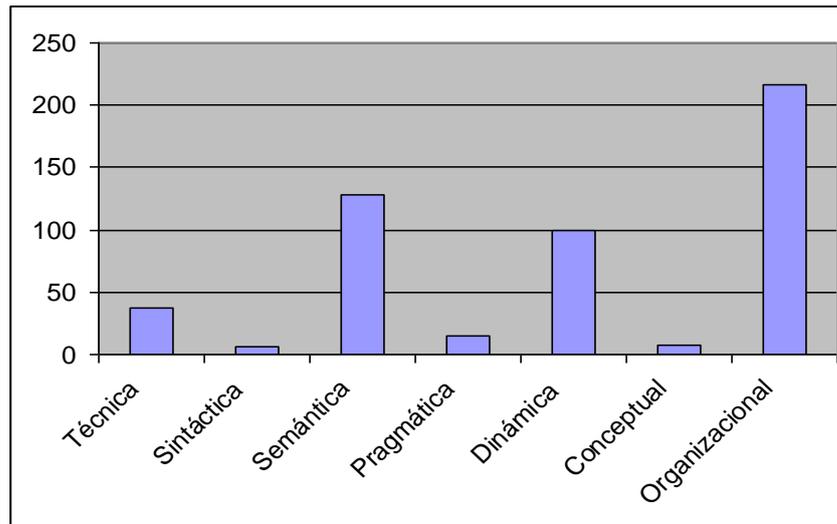


Figura 4. 5: Histograma de los niveles de interoperabilidad en los ítems de la norma ISO19115

Las principales conclusiones que se pueden extraer son:

- Hay un alto grado de paralelismo en los resultados de los dos análisis. Los niveles más favorecidos en ambos casos son organizacional, dinámico y semántico. Los niveles menos favorecidos por los ítems de los metadatos son conceptual y pragmático.
- El porcentaje de ítems que proporcionan los niveles organizacional, pragmático y conceptual se mantiene en los dos alcances del estudio. El resto de niveles de interoperabilidad se ven afectados disminuyendo su representación al estudiarse el conjunto de ítems de la norma frente a los ítems del núcleo fundamental.
- Se observa que persiste en los dos ámbitos la carencia de ítems que proporcionan los niveles de interoperabilidad pragmático y conceptual, por ser muy bajos los porcentajes de ítems que la facilitan: 6 y 2% respectivamente.
- Atendiendo a las anteriores conclusiones, se puede afirmar que la norma de metadatos ISO19115/19139 proporciona principalmente la interoperabilidad organizacional, aportando mucha información que describe las políticas de acceso y el uso de los datos, las responsabilidades, los objetivos y los propósitos; todo ello aporta información útil para evaluar el uso de los datos y los servicios.
- También se puede afirmar que dicha norma no propicia la interoperabilidad conceptual de los datos. La norma de metadatos sólo incluye 7 elementos útiles para describir el modelo de los datos, desde un punto de vista de la ingeniería.

#### 4.4.2 Relaciones entre los niveles de interoperabilidad del modelo

Se trata en primer lugar de analizar cómo proporcionan simultáneamente los metadatos a los niveles de interoperabilidad y, en segundo lugar, se trata de describir las relaciones entre los niveles y sus intensidades. El método empleado para el análisis ha consistido en contabilizar el número de ítems de la norma de metadatos ISO19115 que proporcionan simultáneamente dos o más niveles del modelo, definiendo dos alcances para el estudio: (1) los ítems del núcleo y (2) todos los ítems.

Atendiendo a los ítems del núcleo fundamental de la norma, la tabla 4.3 muestra las cantidades de ítems que simultáneamente facilitan dos niveles de interoperabilidad del modelo, de tal modo que en la intersección de una fila y una columna se observan los valores absolutos de las mismas, y los relativos al nivel identificado en la columna. Así por ejemplo, los valores 28/70% que aparecen en la fila Dinámica y columna Semántica, indican que existen 28 ítems que, simultáneamente, proporcionan los dos niveles y representan el 70% de los (40) ítems que proporcionan la interoperabilidad Semántica.

	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
<b>Técnica</b>	8	0	4	2	3	0	3
	100%	0%	10%	67%	10%	0%	7%
<b>Sintáctica</b>	0	3	3	0	1	0	2
	0%	100%	8%	0%	3%	0%	5%
<b>Semántica</b>	4	3	40	0	28	1	29
	50%	100%	100%	0%	90%	100%	67%
<b>Pragmática</b>	2	0	0	3	2	0	1
	25%	0%	0%	100%	6%	0%	2%
<b>Dinámica</b>	3	1	28	2	31	0	23
	38%	33%	70%	67%	100%	0%	53%
<b>Conceptual</b>	0	0	1	0	0	1	1
	0%	0%	3%	0%	0%	100%	2%
<b>Organizacional</b>	3	2	29	1	23	1	43
	38%	66%	73%	33%	74%	100%	100%

Tabla 4. 3: Relaciones entre los niveles del modelo de interoperabilidad según los ítems del núcleo

Al analizar los valores de la tabla 4.3, se observa que los valores absolutos de la diagonal principal son y coinciden con las frecuencias analizadas en el epígrafe anterior. También se ve que dichas cantidades aparecen de forma simétrica en la tabla 4.3 y los resultados son consistentes con el método utilizado para obtener los datos que consistió en contabilizar la cantidad de ítems que proporcionan dos niveles de interoperabilidad, tratando los mismos de un modo independiente.

En la tabla 4.3 se puede observar que el número de ítems de metadatos que proporcionan los niveles Organizacional, Semántico y Dinámico de interoperabilidad, tomados dos a dos es elevado y, en término medio, el 70% de los ítems que favorece uno de estos niveles proporciona los otros dos.

En la tabla 4.3 sólo se pueden mostrar las cantidades de ítems que proporcionan dos niveles del modelo a la vez. Existen ítems que proporciona 3 o más niveles de interoperabilidad simultáneamente. La siguiente tabla 4.4 contabiliza el número de ítems que sólo proporcionan uno, dos, tres, etc., niveles del modelo simultáneamente y la figura 4.6 muestra la distribución de forma gráfica.

<b>nº niveles simultáneos</b>	1	2	3	4	5	6	7
<b>Nº de ítems que los favorecen</b>	11	17	27	1	0	0	0

Tabla 4. 4: Número de ítems del núcleo de la norma de metadatos que sólo proporcionan uno, dos, etc., niveles en el modelo

Cabe destacar que los 11 ítems que aparecen en la segunda columna de tabla 4.4, que indican que sólo proporcionan un nivel de interoperabilidad, lo hacen a nivel organizacional.

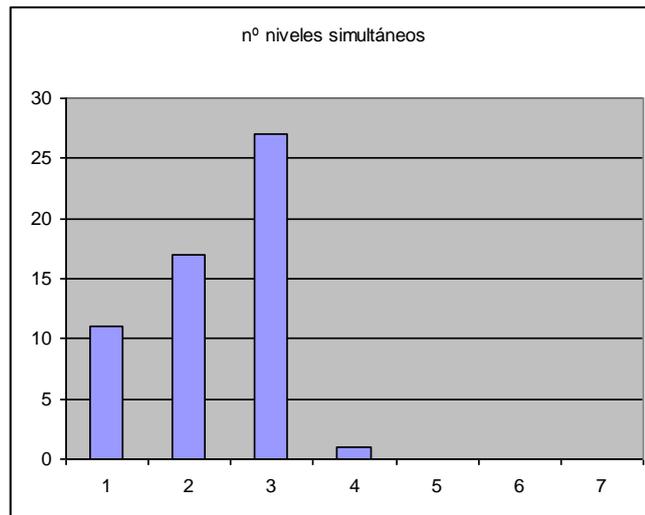


Figura 4. 6: Distribución de las cantidades de ítems del núcleo de la norma de metadatos que sólo proporcionan un nº de niveles del modelo

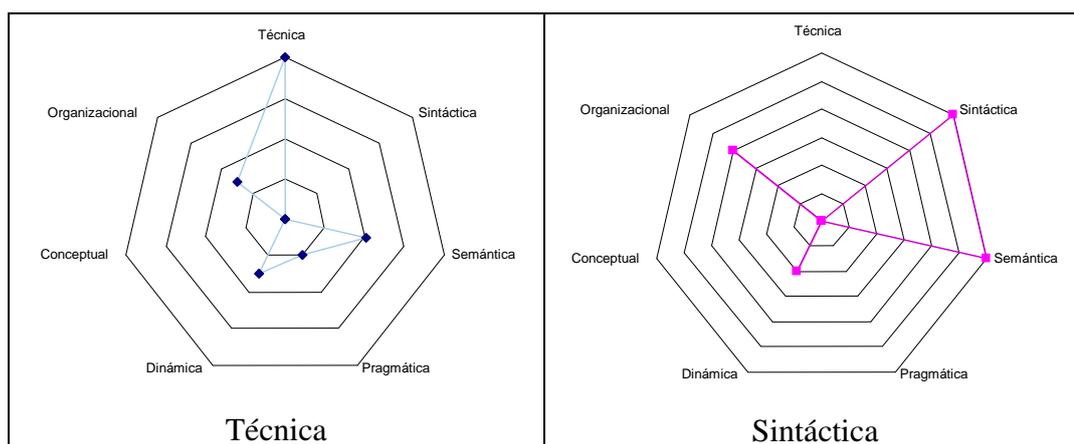
La tabla 4.3 sólo muestra las cantidades de ítems que proporcionan dos niveles de interoperabilidad simultáneamente. Por esta razón se ha confeccionado la tabla 4.5, en la que se muestran el resto de combinaciones de niveles de interoperabilidad habilitados a la vez por los ítems del núcleo fundamental de la norma de metadatos. Se puede observar que hay un número importante de ítems que proporcionan simultáneamente los niveles Semántico, Dinámico y Organizacional.

<b>Niveles de interoperabilidad proporcionados por los ítems</b>	<b>Nº ítems</b>
Técnico, Semántico y Dinámico	1
Técnico, Semántico y Organizacional	1
Técnico, Dinámico y Organizacional	3
Sintáctico, Semántico y Dinámico	2
Sintáctico, Semántico y Organizacional	2
Semántico, Dinámico y Organizacional	20
Semántico, Conceptual y Organizacional	1
Pragmático, Dinámico y Organizacional	1
Técnico, Semántico, Dinámico y Organizacional	1

Tabla 4. 5: Niveles de interoperabilidad proporcionados simultáneamente

Las principales diferencias que pueden encontrarse entre las tablas 4.3, 4.4 y 4.5 en cuanto al número de ítems que proporcionan determinados niveles del modelo, tienen su origen en la forma de realizar el conteo. Mientras que la tabla 4.4 muestra las cantidades de ítems que sólo proporcionan un número de niveles, las otras dos muestran todas las posibles combinaciones de niveles. Por ejemplo, si un determinado ítem facilita 3 niveles en el modelo (1, 2 y 3), en la tabla 4.4 solo se computa como “ítem que proporciona 3 niveles”, mientras que en la tabla 4.3 se cuentan las combinaciones de estos 3 elementos tomados de dos en dos (1,2 + 1,3 + 2,3).

La disparidad de magnitudes mostradas en la tabla 4.3 dificulta notoriamente la interpretación y descripción de las interrelaciones. Este hecho ha motivado la búsqueda de alguna técnica de visualización que permita representar las relaciones de un modo gráfico, fácil de leer e interpretar. Se realizaron diversas pruebas hasta decidir el uso de los diagramas radiales como técnica de representación. En una primera instancia se representaron todos los valores de la tabla en un único diagrama y el resultado no mostraba más información que la propia tabla. Finalmente, se seleccionaron este mismo tipo de gráficos mostrando por cada nivel del modelo, uno con las relaciones de éste con los demás. Los resultados de estas representaciones pueden observarse en la tabla 4.6.



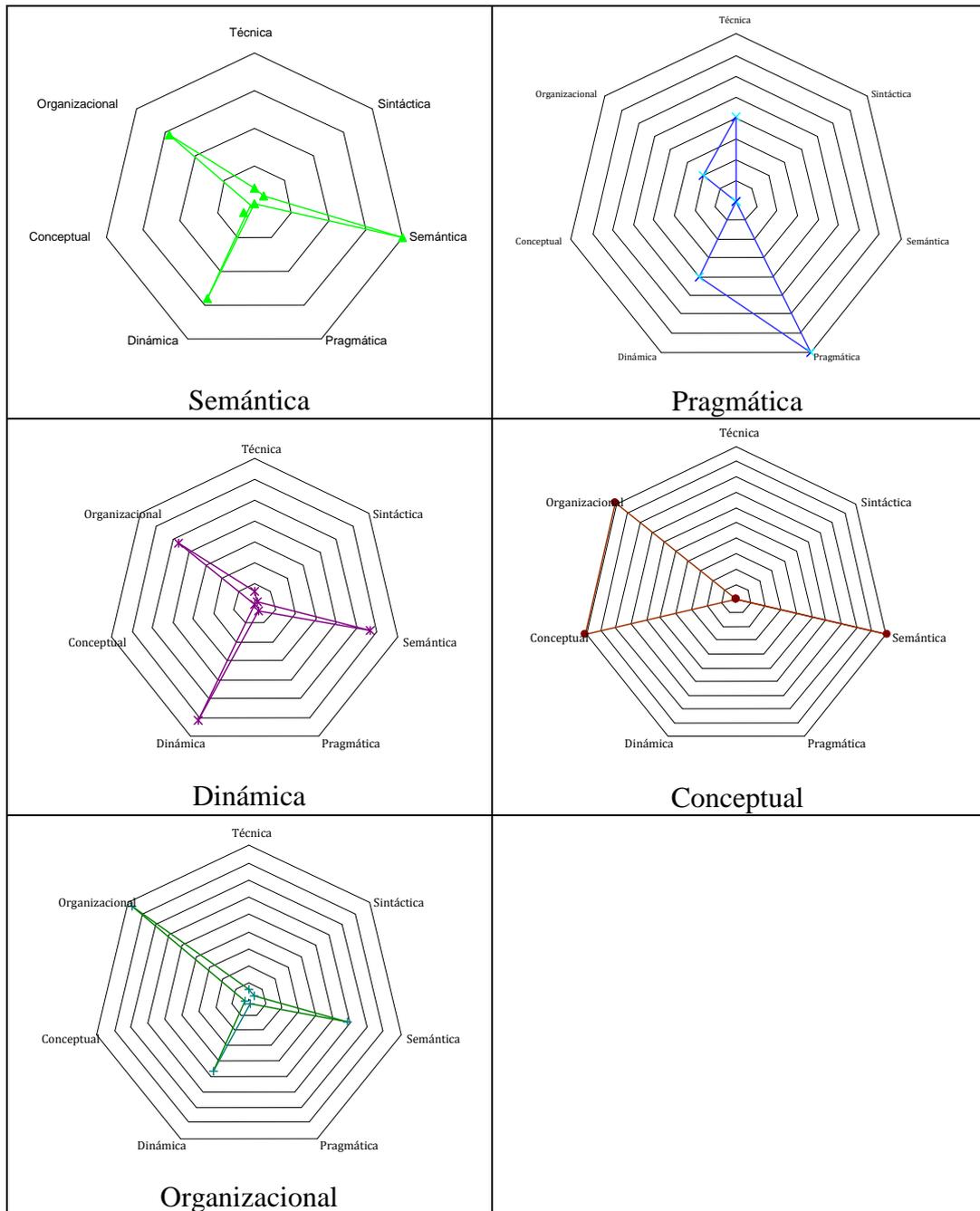


Tabla 4. 6: Representación gráfica del número de ítems que proporcionan dos niveles del modelo

Cada gráfico radial representa, en términos porcentuales, el número de ítems que además del nivel de interoperabilidad que aparece en el pie de cada figura proporciona el resto de niveles. Se observa la similitud de los gráficos correspondientes a la interoperabilidad semántica, dinámica y organizacional denotando que los mismos ítems proporcionan los tres niveles simultáneamente. En el resto de gráficos se aprecia las relaciones que se establecen entre pares de niveles de interoperabilidad sin encontrarse reciprocidad.

Se pueden extraer las siguientes conclusiones sobre los gráficos de la tabla 4.6:

- Los tres gráficos asociados a la interoperabilidad semántica, dinámica y organizacional son similares, porque las magnitudes asociadas al número de éstos que proporcionan estos ítems respecto al resto de niveles son muy altas.
- Se identifican 7 pares de niveles de interoperabilidad que no se ven proporcionados por los ítems del núcleo de la norma de metadatos. En unos casos se puede justificar argumentando la dificultad para identificar un ítem de metadato que pueda facilitar aspectos de la interoperabilidad tan diferentes como son el técnico y el conceptual o el sintáctico y el conceptual.
- Se ha identificado un número muy reducido de ítems en el núcleo de la norma que faciliten la interoperabilidad técnica, sintáctica y pragmática. Resulta difícil obtener conclusiones a partir de estos datos, porque el núcleo fundamental de la norma de metadatos ha sido definido atendiendo a un tipo de criterios próximos a las funciones que desempeñan los datos, y no relacionados con la interoperabilidad.

Se ha aplicado la misma metodología para analizar todos los ítems de la norma de metadatos, y los resultados del análisis se muestran en la tabla 4.7. Como puede observarse, en los elementos de la diagonal principal de la tabla 4.7, el número de ítems analizados para modelar las relaciones entre los niveles de interoperabilidad discrepan de los mostrados antes en la tabla 4.2. En el caso anterior se han tenido en cuenta los 400 posibles ítems (ANZLIC 2005); sin embargo, en este estudio sólo se ha computado una vez los ítems pertenecientes a los paquetes/clases *CI\_Citation* y *CI\_ResponsibleParty*.

	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
<b>Técnica</b>	37	3	12	8	11	0	26
	100%	50%	9,3%	53,3%	11,1%	0%	12%
<b>Sintáctica</b>	3	6	3	3	4	0	3
	8,1%	100%	2,3%	20%	4%	0%	1,3%
<b>Semántica</b>	12	3	128	4	83	6	118
	32,4%	50%	100%	26,6%	83,8%	85,7%	54,3%
<b>Pragmática</b>	8	3	4	15	12	0	10

	21,6%	50%	3,1%	100%	12,12%	0%	4,6%
<b>Dinámica</b>	11	4	83	12	<b>99</b>	0	90
	29,7%	67%	64,8%	80%	100%	0%	41,4%
<b>Conceptual</b>	0	0	6	0	0	<b>7</b>	7
	0%	0%	4,6%	0%	0%	100%	3,2%
<b>Organizacional</b>	26	3	118	10	90	7	<b>217</b>
	70,27%	50%	92,2%	66,6%	90,9%	100%	100%

Tabla 4. 7: Relaciones entre los niveles del modelo de según los ítems de la norma

En la tabla 4.7 se observa:

- Hay 4 parejas de niveles de interoperabilidad (conceptual y técnico, o sintáctico, o pragmático o dinámico) en el modelo, que no han sido proporcionadas por los ítems de la norma de metadatos.
- Los 7 ítems que facilitan el nivel conceptual también proporcionan los niveles organizacional (7) y semántico (6). Se puede interpretar, por tanto, que existe una relación estrecha entre el nivel conceptual y los niveles organizacional y semántico. Esta relación no es simétrica ya que tan solo 7 de los 217 ítems que proporcionan la interoperabilidad organizacional facilitan la conceptual.
- Persiste, como en el análisis del núcleo fundamental, el bajo número de ítems (7) que proporcionan la interoperabilidad conceptual.
- Se detecta un número representativo de ítems que proporcionan la interoperabilidad pragmática (15) además de la dinámica (12). Por tanto, se puede interpretar como una estrecha e intensa relación entre ambos niveles del modelo, aunque esta relación también es asimétrica al observarse que solo el 12% de los ítems que proporcionan la interoperabilidad dinámica facilitan la pragmática.
- Como puede observarse en la última fila de tabla 4.7, los porcentajes de ítems que proporcionan a la vez la interoperabilidad organizacional para cada uno de los niveles del modelo es alto: 50%, 66%, 70%, 90% y 100%. Una posible interpretación de estos datos es que la norma de metadatos ISO19115 ha sido definida, principalmente, para satisfacer las necesidades de interoperabilidad a nivel organizacional.
- La correlación matemática calculada entre las cantidades de ítems de metadatos que proporcionan cada nivel en los ámbitos del estudio (tabla 4.3 y

4.7) adopta el valor 0,9436 indicando que existe un alto grado de asociación entre los resultados de ambos análisis.

También se ha analizado el número de ítems que sólo proporcionan simultáneamente uno, dos, tres, etc., niveles del modelo, mostrándose los resultados en la tabla 4.8 y en la figura 4.7 respectivamente.

Nº niveles simultáneos	1	2	3	4	5	6	7
Nº de ítems que los favorecen	80	46	101	6	2	0	0

Tabla 4. 8: Número de ítems de la norma de metadatos que sólo proporcionan uno, dos, etc., niveles en el modelo

Otro hecho que refuerza la interpretación expuesta sobre el nivel organizacional como centro de gravedad en la definición de la norma de metadatos, es que 74 de los 80 ítems (el 92%) que proporcionan un solo nivel de interoperabilidad, mostrado en la tabla 4.8) facilitan dicho nivel.

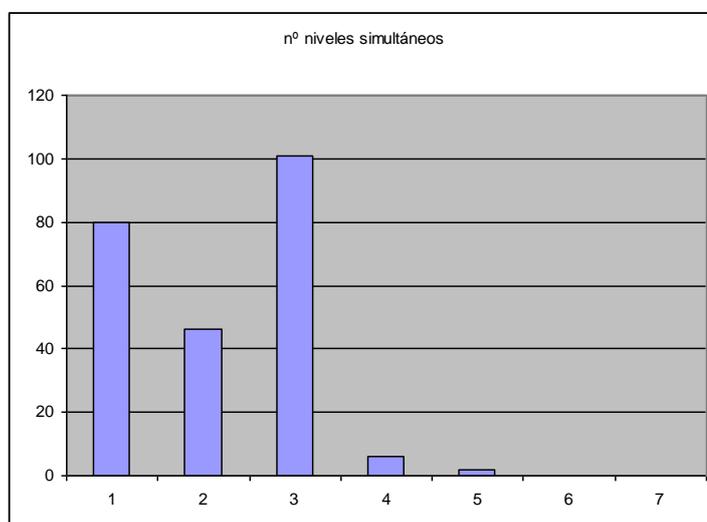


Figura 4. 7: Distribución de las cantidades de ítems de la norma de metadatos que sólo proporcionan un nº de niveles del modelo

A continuación se interpretan las relaciones entre los niveles del modelo basándose en los ítems de metadatos que lo facilitan, para describir los beneficios que proporcionan en el otro nivel.

*Interoperabilidad técnica y pragmática.* Los ítems: punto de acceso a los datos o el servicio, el formato, la versión de archivo, el juego de caracteres y el protocolo de comunicación a utilizar, posibilitan la interoperabilidad técnica y facilitan el uso de los datos/servicios de un modo práctico.

*Interoperabilidad técnica y dinámica.* Algunos ítems tales como: la técnica de compresión de datos, el tamaño del archivo, el tipo de servicio o la versión del mismo, además de posibilitar la interoperabilidad técnica, permiten evaluar si los datos o el servicio pueden usarse en un contexto dinámico por otros sistemas.

*Interoperabilidad sintáctica y pragmática.* La interoperabilidad pragmática se beneficia de los aspectos sintácticos que permiten describir atributos tan importantes como los nombres de los parámetros de un servicio, la tipología de datos, la obligatoriedad y su cardinalidad.

*Interoperabilidad semántica y técnica.* La interoperabilidad técnica se beneficia de los ítems de metadatos que aportan semántica, por ejemplo, las listas controladas de términos que identifican el formato de archivo (*MD\_MediumFormatCode*), el soporte de almacenamiento (*MD\_MediumNameCode*) o el juego de caracteres usados en los datos (*MD\_CharacterSetCode*).

*Interoperabilidad semántica y pragmática.* Los elementos que dotan interoperabilidad semántica también proporcionan el nivel pragmático. Por ejemplo: los ítems de metadatos que utilizan listas controladas para definir las categorías de los Geoservicios, como ocurre con las reglas de implementación de los metadatos de INSPIRE (2008), el tipo de representación espacial (*MD\_SpatialRepresentationTypeCode*), el estado de los datos (*MD\_ProgressCode*) o las restricciones de seguridad (*MD\_SecurityConstraints*) son ítems que, además de fortalecer la interoperabilidad semántica, facilitan la pragmática.

*Interoperabilidad semántica y dinámica.* La interoperabilidad dinámica se ve favorecida por los ítems de metadatos que dotan de interoperabilidad semántica.

Estos ítems son, por ejemplo, las listas de términos que permiten clasificar los recursos en tópicos o palabras clave pertenecientes a diccionarios multilingües de los dominios (tema, lugar, estrato). Estos ítems proporcionan el nivel semántico y mejoran la calidad de los resultados de las búsquedas de datos o servicios que pueden ser usados en un entorno dinámico por los sistemas.

*Interoperabilidad semántica y conceptual.* El ítem más representativo que facilita ambos niveles del modelo es el que define el alcance de los metadatos (*MD\_ScopeCode*).

*Interoperabilidad semántica frente a organizacional.* La interoperabilidad organizacional se beneficia de algunos aspectos de la semántica, por ejemplo, aquellos en los que se define el rol que desempeñan los responsables de los datos, los tipos de restricciones legales de acceso y uso de los datos o los servicios y las restricciones de seguridad (*CI\_RoleCode*, *MD\_RestrictionCode*, *MD\_ClassificationCode*). Como puede apreciarse, todos los ejemplos mencionados también son listas controladas de términos.

*Interoperabilidad organizacional y dinámica.* La interoperabilidad dinámica se apoya en aspectos relacionados con la organización, tales como las fuentes de datos, las restricciones y/o limitaciones de acceso o de uso de los datos. Los ítems de metadatos que describen los aspectos organizacionales citados, facilitan la reutilización de los datos y la explotación de los servicios.

*Interoperabilidad pragmática y dinámica.* La interoperabilidad dinámica se ayuda de los aspectos pragmáticos para poder explotar los datos o servicios, una vez que han sido identificados y se ha evaluado la idoneidad de los mismos para los objetivos perseguidos. En particular, se puede afirmar que esta relación entre los niveles pragmático y dinámico no es simétrica.

## 4.5 Conclusiones

A continuación se presentan las conclusiones del capítulo:

- Se ha propuesto una nueva metodología de creación automática de metadatos. Esta metodología puede ser utilizada para crear los metadatos de un modo automático y sin intervención humana. También puede utilizarse como primer paso de un refinamiento posterior por parte de los usuarios expertos en el dominio del conocimiento o por los expertos en catalogación de la información y finalmente ser utilizada para actualizar los metadatos cuando cambian los datos.
- La nueva metodología:
  - Automatiza la extracción de información tanto implícita como explícita de los almacenes de datos,
  - Realiza razonamientos y traducciones para identificar sistemas de referencia de las coordenadas,
  - Realiza cálculos de conversión/transformación de coordenadas,
  - Localiza identificadores geográficos relevantes para la envolvente geográfica,
  - Representa gráficamente los datos,
  - Infiere el contenido de los datos en base a datos estadísticos y reglas almacenadas que permiten catalogar los datos,
  - Infiere el modelo de datos y lo representa en formatos usados en ingeniería, para finalmente
  - Empaquetar toda la información obtenida por los diversos métodos en un metadato en formato XML incluyéndolo junto al resto de la información en un archivo con formato de intercambio (*MEF*).
- Desde el punto de vista de la interoperabilidad, se ha definido un modelo integrado para las IDE fundamentado en siete niveles: técnico, sintáctico, semántico, pragmático, dinámico, conceptual y organizacional. La elección de los mismos se ha justificado teniendo en cuenta que las IDE son un caso particular de los sistemas de sistemas y en este contexto existen modelos de interoperabilidad. Dado que para las IDE son importantes los aspectos relacionados con la organización, en este modelo se ha incluido, además de los niveles pertinentes para los *SoS*, el organizacional.

- Se ha analizado la influencia de los ítems pertenecientes a la norma internacional de metadatos en los niveles del modelo. El alcance del estudio ha sido doble: los ítems del núcleo fundamental y toda la norma. Los resultados a nivel de interoperabilidad de ambos casos son similares, e indican que los niveles del modelo más favorecidos son el organizacional, dinámico y semántico.
- Al analizar cómo proporcionan interoperabilidad simultáneamente en varios niveles los ítems de la norma de metadatos, se ha observado que un porcentaje importante de ítems que proporcionan un determinado nivel, también proporcionan el organizacional. También el 92% de los ítems que sólo proporcionan un nivel del modelo lo hacen a nivel organizacional. Interpretamos estos hechos del siguiente modo: la norma de metadatos ISO19115 satisface, en gran medida, las necesidades de interoperabilidad a niveles de organización y no ha sido diseñada para proporcionar otros niveles de interoperabilidad pertinentes para las IDE.
- Los gráficos que representan las cantidades de ítems que proporcionan simultáneamente uno, dos, tres, etc. niveles de interoperabilidad (4.6 y 4.7) indican que los ítems de la norma proporcionan pocos niveles de interoperabilidad simultáneamente. En el conjunto de la norma, hay 80 ítems (37%) que solo proporcionan un nivel y 101 ítems (46%) que proporcionan 3 simultáneamente. El 80% de estos últimos se corresponden con los niveles semántico, dinámico y organizacional, predominante en este estudio.
- Bajo nuestro punto de vista, la interoperabilidad sintáctica la garantiza la propia norma, al definir las reglas de codificación de los ítems que la forman.
- La interoperabilidad técnica y pragmática se han de garantizar por otras normas de la misma familia, definiendo los protocolos de comunicaciones y transferencias de datos o definiendo las interfaces de los servicios que posibilitan el acceso, el tratamiento, la conversión o la visualización de los datos.
- Se detecta una carencia importante de ítems que proporcionan la interoperabilidad conceptual en la norma ya que sólo dispone de 7 ítems que pueden describir el modelo de datos de la información y su definición es vaga.
- El modelo de interoperabilidad propuesto, basado en 7 niveles, puede ser utilizado para definir perfiles de metadatos que satisfagan los distintos niveles

de interoperabilidad del modelo de un modo equilibrado y maximizando el ratio de niveles de interoperabilidad por ítem de metadato. Esto sólo se podría maximizar si se diseñase una nueva norma de metadatos que responda a las necesidades y requisitos de interoperabilidad.

## 5 IMPLEMENTACIÓN Y RESULTADOS

### 5.1 Introducción

En este capítulo se presentan en primer lugar los metadatos implícitos contenidos en la IG como fuentes de ítems para la creación automática de metadatos y la heterogeneidad de formatos de representación de los sistemas de referencia espacial. El objetivo de este epígrafe es conocer detalladamente la tipología y el número de ítems de metadatos que pueden obtenerse de los datos, mediante utilidades y librerías que extraigan la información almacenada en los formatos, o mediante tablas que identifiquen las codificaciones utilizadas por los mismos, con el fin de expresar los sistemas de referencia espaciales para las coordenadas.

En segundo lugar, se analiza la metodología propuesta para crear metadatos automáticamente, considerando los resultados de los estudios previos e identificando el resto de ítems que pueden crearse mediante cálculo e inferencia o aquellos que pueden deducirse del contexto. En este apartado se describen individualmente los ítems que pueden obtenerse con la metodología, identificando la tipología de datos a los que aplican los mismos.

En tercer lugar se trata de dar respuesta la quinta cuestión de investigación planteada en el proyecto de tesis: ¿Cuáles son las fortalezas y las debilidades de los metadatos generados manualmente y automáticamente desde el punto de vista de la interoperabilidad de los sistemas que los explotarán (infraestructuras de datos espaciales)? Con este propósito se analizan los ítems que puede crear la metodología de creación automática de metadatos, desde dos puntos de vista: el de la interoperabilidad que facilitan y el de las funciones que desempeñan. Para realizar este análisis, se han clasificado los datos según su naturaleza: ráster, modelos digitales de elevaciones y datos vectoriales. Los resultados conseguidos con la metodología se comparan con los obtenidos en el análisis de los niveles de interoperabilidad de los ítems, tanto del núcleo fundamental como de la norma de metadatos completa para, finalmente, presentar un conjunto de reflexiones sobre este apartado.

En cuarto lugar, se trata de dar respuesta a la cuarta cuestión de investigación planteada en el proyecto de tesis: ¿Qué propuesta de validación es más adecuada para validar un modelo de interoperabilidad de sistemas en el contexto de las infraestructuras de datos espaciales? Se plantea la validación del modelo de interoperabilidad, hecho para el que se ha diseñado una encuesta dirigida a un conjunto de expertos en metadatos, a los que se ha proporcionado la información necesaria para identificar los niveles de interoperabilidad proporcionados por los ítems del núcleo fundamental de la norma de metadatos. El análisis de los resultados, a nivel individual y agregado, se utiliza para detectar discrepancias frente al modelo propuesto y, finalmente, para realizar la validación.

Por último, se presentan las conclusiones de este capítulo relativo a la implementación y los resultados.

## **5.2 Formatos de almacenamiento usados en las IDE**

La información geográfica, en función de su naturaleza, puede ser almacenada de diversas maneras de modo que se puede encontrar en formatos ráster (imágenes), formatos vectoriales, modelos digitales de elevaciones, datos en forma de tablas o bases de datos como soportes de almacenamiento.

En cada país, bien sea por la cultura o por la tecnología disponible, se han adoptado, como estándares de hecho (facto) para el intercambio de la información geográfica, un conjunto de formatos. En muchos casos, estos formatos son propiedad de empresas que desarrollan tecnología en este sector y no existe información pública que describa la estructura del formato; en otros los formatos son fruto de los esfuerzos de consenso y estandarización de organizaciones que colaboran a nivel internacional.

El objetivo último de este epígrafe es identificar los ítems de los metadatos que pueden extraerse de los distintos formatos de almacenamiento. Se han clasificando los formatos de acuerdo con la naturaleza de los datos que almacenan, para finalizar con la descripción de los metadatos que pueden ser recuperados de los mismos.

- Formatos ráster (matriciales) de propósito general: hace tiempo, se utilizaron los formatos gráficos desarrollados en el contexto de las herramientas de ofimática para almacenar imágenes aéreas. La principal ventaja de estos formatos reside en la disponibilidad de herramientas y librerías informáticas que permiten manipular dichos gráficos. Frente a las ventajas aparecen los múltiples inconvenientes: número de bandas, resolución radio-métrica o número de bits por píxel limitados o no poder almacenar metadatos. La aparición de nuevos formatos estandarizados y propietario ha motivado que estos formatos estén cada vez más en desuso en este contexto. Son formatos ráster de propósito general: BMP, PNG, RAS, TIFF, JPEG, GIF, IFF y PCX.
- Formatos ráster utilizados para almacenar fotografías aéreas digitales, ortofotografías y cartografía rásterizada: estos formatos permiten almacenar grandes volúmenes de datos y soportar mayor número de bandas, disponen de más tipos de datos para los píxeles, proporcionan mayor capacidad de compresión y multi-resolución. También, en algunos casos permiten almacenar metadatos relativos a los sistemas de referencia espacial. Algunos formatos ráster con éstas características son: GeoTIFF, MrSID, ECW, JPEG2000, GeoJP2, INGR y NITF.
- Formatos ráster utilizados en teledetección: en esta técnica, para almacenar imágenes de satélite, se precisan formatos que permitan almacenar imágenes hiper-espectrales (muchas bandas), con radiometrías almacenadas como números reales, además de almacenar una gran cantidad de metadatos adicionales. Algunos formatos ráster utilizados en teledetección son: IMG (o *HFA*), PIX (o *PCIDISK*), ERS, IMG (*Idrisi*), NOAAL1B y EOSAT.
- Formatos vectoriales de dibujo asistido por ordenador (CAD): el CAD es una herramienta informática utilizada para editar cartografía digital, cuya finalidad es la impresión en forma de mapa o plano. Se utiliza en el dominio de la ingeniería y la arquitectura y, aunque dispone de herramientas para almacenar información geográfica, se suele utilizar para almacenar las geometrías y los estilos de visualización. De la misma forma que les ocurre a los formatos ráster de propósito general, los formatos CAD no suelen almacenar los metadatos: tipos de entidades, sistemas de referencia espacial, etc. Algunos

formatos de CAD son: DGN (ISFF), DWG y DXF, FHX (*Macromedia*), AI (*Adobe*) y BIN (*DIGI21*).

- Formatos vectoriales usados por los SIG: los SIG suelen independizar la geometría, la información alfanumérica y la visualización de las entidades. Son formatos utilizados por los SIG: ADF y E00 (*ArcInfo*), SHP, MIF y DAT (*MapInfo*), VEC (*Idrisi*), a los que hay que sumar otros formatos SIG ráster.
- Modelos digitales de terreno almacenados en forma vectorial: se trata de formatos que almacenan redes de triángulos irregulares (*TIN*) vectorialmente. Algunos formatos son: TIN (*Intergraph*), MDT (*MDTop*).
- Modelos digitales del terreno almacenados matricialmente: son formatos que soportan datos radio-métricos y otros muchos casos de tipo real, para almacenar las elevaciones del terreno por medio de puntos. La principal carencia de estos formatos es que no suelen almacenar metadatos y por lo tanto, no almacenan ni el sistema de referencia espacial ni otras posibles informaciones. Algunos formatos matriciales para MDT son: ADF, GRD (*ESRI*), GRD (*GoldenSoft*), DEM (USGS), DEM (*MicroDEM*), DTE (*Socet Set*), DT0 (DoD), HGT (*Shuttle Radar Topography Mission*), BIL, BIP, BSQ (*MapInfo*).
- Información geográfica almacenada en Bases de Datos espaciales: los sistemas de gestión de las bases de datos (SGDB) han evolucionado y, además de permitir almacenar objetos alfanuméricos, son capaces de almacenar objetos geométricos y de realizar con ellos operaciones topológicas e indexar el contenido desde un punto de vista espacial. En otros casos, algunos SIG implementan funciones intermedias (*middleware*) entre los SGDB relacionales y los SIG para complementar a las bases de datos con éstas capacidades. En ambos casos, los SGDB almacenan metadatos en forma de tablas auxiliares. Algunos ejemplos de SGBD con extensiones espaciales nativas son: Oracle Spatial, PostGreSQL+PostGIS, MySQL, DB2 (IBM) y SQL2008 (*Microsoft*). Otras aplicaciones que son capaces de almacenar información geográfica en SGBD relacionales son: MGE, Geomedia, ArcSDE, MapInfo.

## 5.2.1 Formatos matriciales de propósito general

Tras revisar la documentación técnica de los formatos ráster: BMP, PNG, RAS, TIFF, JPEG, GIF, IFF, PCX y PSD, se ha constatado que se puede leer de los archivos el siguiente conjunto de características:

- Las dimensiones de la imagen (ancho y alto expresadas en píxeles),
- El número de bandas o componentes,
- El número de bits utilizados para representar cada píxel,
- El tipo de compresión utilizada.

La información relativa a la georreferenciación se gestiona mediante un archivo de texto auxiliar que contiene los parámetros necesarios para realizar una transformación afín (giro, traslación y cambio de escala) denominado “*World file*” (World file, 2007).

El inconveniente de estos formatos es el desconocimiento del sistema de referencia espacial al que hacen referencia las coordenadas que aporta el fichero de georreferenciación. Algunos formatos como PNG y TIFF, pueden contener metadatos adicionales como son: autor, descripción del contenido, fecha de creación y cita a fuentes de información. La tabla 5.1 recoge los ítems de metadatos que pueden extraerse de este tipo de archivos.

Formato	Ancho/Alto	Nº Bits por píxel	Nº de Bandas	Tipo de compresión	Extensión geográfica	Resolución del píxel	Otros
BMP	X	X	X	X	X <sup>(1)</sup>	X <sup>(1)</sup>	Resolución de impresión
GIF	X	X	X	X	X <sup>(1)</sup>	X <sup>(1)</sup>	
IFF	X	X	X	X	X <sup>(1)</sup>	X <sup>(1)</sup>	
JPEG	X	X	X	X	X <sup>(1)</sup>	X <sup>(1)</sup>	
PCX	X	X	X	X	X <sup>(1)</sup>	X <sup>(1)</sup>	
PNG	X	X	X	X	X <sup>(1)</sup>	X <sup>(1)</sup>	Autor, descripción, fecha de creación, fuentes, pasos del proceso y restricciones legales
PSD	X	X	X	X	X <sup>(1)</sup>	X <sup>(1)</sup>	
TIFF	X	X	X	X	X <sup>(1)</sup>	X <sup>(1)</sup>	Autor, descripción, fecha de creación y fuentes
RAS	X	X	X	X	X <sup>(1)</sup>	X <sup>(1)</sup>	

Tabla 5.1: Formatos ráster genéricos

<sup>(1)</sup> Condicionado a la existencia del archivo de georreferenciación (*World file*)

## 5.2.2 Formatos ráster de alta compresión

La principal dificultad para realizar este estudio es la ausencia de documentos técnicos públicos que describan los formatos, en especial con los privativos MrSID y ECW. Estas dificultades se han paliado parcialmente analizando la información que muestra herramientas desarrolladas por las mismas firmas comerciales, e identificando aquella que tenga ítems de metadatos.

La información que puede obtenerse para los formatos de este tipo analizados (GTIFF, MrSID, ECW, JPEG2000, GeoJP2, INGR y NITF) es, además de la común obtenida para los formatos ráster de propósito general:

- La resolución del píxel en cada eje,
- Las unidades de medida,
- El sistema de referencia espacial.

Algunos formatos como el GeoTIFF codifican el sistema de referencia espacial mediante identificadores numéricos definidos por el EPSG. Otros, identifican sistemas de referencia mediante codificaciones textuales, nemotécnicos o códigos particulares. Además, algunos formatos contienen otros metadatos como: fecha de la creación, calidad de la compresión, cita a la fuente de datos, restricciones de uso y/o acceso a la información, así como parámetros más detallados de las proyecciones cartográficas. Destaca la capacidad del formato JPEG2000 para el almacenamiento de metadatos, al incluir un bloque de información en la cabecera que pueden almacenar los mismos. La tabla 5.2 recoge los ítems de metadatos de los formatos usados para almacenar este tipo de información.

Formato	Ancho y alto	Nº de bits por píxel	Nº de bandas	Tipo de compresión	Extensión geográfica	Resolución del píxel	Sistema de referencia espacial	Unidades de medida	Otros
ECW	X	X	X	X	X	X	X	X	
GeoJP2	X	X	X	X	X	X	X	X	Autor, descripción, fecha de creación, fuentes y otros parámetros de la proyección
GTIFF	X	X	X	X	X	X	X	X	Autor, descripción, fecha de creación, fuentes y otros parámetros de la proyección
INGR	X	X	X	X	X	X	X	X	
JPEG2000	X	X	X	X	X	X	X	X	Abierto. Especificación GML JP2K
MrSID	X	X	X	X	X	X	X	X	Calidad compresión, fecha creación
NITF	X	X	X	X	X	X	X	X	Fecha de creación, título, autor, fuente, restricciones legales

Tabla 5.2: Formatos ráster para grandes volúmenes de datos

### 5.2.3 Formatos ráster utilizados en teledetección

Tras realizar la revisión de los formatos, se dispone de información suficiente para poder identificar los ítems de metadatos almacenados en los formatos de almacenamiento. Este análisis se ha realizado para los formatos LAN e IMG (ERDAS), PCIDISK, ERS, IDRISI, NOAAL1B y F-EOSAT. Todos ellos tienen en común que almacenan abundantes metadatos, muchos de los cuales no tienen cabida en la norma ISO19115, si bien podrá ser incluida parcialmente en la ISO19115-2. Esta norma pretende almacenar los metadatos necesarios para describir tanto las mallas regulares de información como las imágenes de satélite. Además de los metadatos comunes para los formatos anteriores, estos tipos de archivos almacenan:

- El tipo de dato utilizado para almacenar los píxeles,
- La fecha de la adquisición,
- La plataforma del satélite,
- La estación de recepción que aplica el preprocesado,
- El nivel de preprocesado,
- Los parámetros estadísticos de los valores digitales de los píxeles,
- Los parámetros del sistema de adquisición (posición y ángulos).

y otros que pueden ser difíciles de ubicar en otros ítems de un metadato. La tabla 5.3 recoge los ítems de metadatos de los formatos usados para almacenar este tipo de información.

Formato	Ancho y alto	Tipo dato	Nº de bandas	Extensión geográfica	Resolución del píxel.	Estadísticos Max, min	Otros estadísticos	Sistema de referencia espacial	Otros
LAN (Erdas)	X	X	X	X	X			X	
IMG (Erdas)	X	X	X	X	X	X	X	X	Tipo de contenido
PIX	X	X	X	X	X			X	Fecha de creación, fecha proceso, proceso
ERS	X	X	X	X	X	X		X	Sensor, resolución espectral bandas.
DOC (Idrisi)	X	X	X	X	X	X		X	Título, fuente de información, categorías.
NOAAL1B	X	X	X	X	X			X	Plataforma, puntos de control, receptor y formato
F-EOSAT	X	X	X	X	X			X	Fecha de adquisición, sensor, satélite, estación de recepción

Tabla 5.3: Formatos ráster utilizados en teledetección

## 5.2.4 Modelos digitales del terreno almacenados matricialmente

Los formatos identificados para el almacenamiento de modelos digitales del terreno matriciales son: BIL, BIP y BSQ (MapInfo), Gtopo30, Ráster de exportación (Erdas), HGT (SRTM), ADF y GRD (Esri), Grid (Surfer), DEM (USGS), DEM (MicroDEM), DTED (DoD), DOQ2 y DTE (Socet Set). Se ha constatado que los ítems de metadatos que pueden obtenerse son similares para todos ellos. Se diferencian principalmente en lo relativo al almacenamiento del sistema de referencia espacial (Datum, elipsoide, proyección, zona, parámetros). Por ejemplo, para los formatos GTOPO30 y HGT es implícito; los formatos DTE (Socet Set) y Grid (Surfer) no almacenan la extensión espacial. La información que se puede extraer de los formatos es básicamente la misma que en los anteriores, incluyendo la cota máxima y mínima. La tabla 5.4 recoge los ítems de metadatos de los formatos usados para almacenar este tipo de información.

Formato	Ancho y alto	Extensión geográfica	Resolución del píxel	Nº de bits por píxel	Nº de bandas	Estadísticos máx, mín	Otros estadísticos	Unidades horizontales	Sistema de referencia espacial	Otros
DOQ2 (USGS)	X	X	X	X	X			X	X	
DTED	X	X	X	X	X			X	X	
Gtopo30	X	X	X	X	X	X <sup>(2)</sup>		X	X	
HGT (SRTM)	X	X	X	X	X	X <sup>(2)</sup>		X	X	
BIL, BIP, BSQ (Mapinfo)	X	X	X	X	X	X <sup>(2)</sup>		X	X	
DTE (Socet set)	X	X	X	X	X			X		
GRD (Esri)	X	X	X			X <sup>(2)</sup>		X	X	
GRD (Surfer)	X	X	X			X				
DEM (USGS)	X	X	X	X		X		X	X	
DEM (MicroDEM)	X	X	X	X	X	X		X	X	
E00 grd	X	X	X <sup>(3)</sup>		X	X <sup>(3)</sup>	X <sup>(3)</sup>	X <sup>(3)</sup>	X <sup>(3)</sup>	
ADF grd	X	X	X <sup>(3)</sup>		X	X <sup>(3)</sup>				

Tabla 5.4: Formatos ráster utilizados en modelos digitales del terreno

<sup>(2)</sup> Se puede leer todo el archivo y calcular dichos valores

<sup>(3)</sup> Si todas las secciones del archivo están presentes

## 5.2.5 Formatos vectoriales tipo CAD

Se han identificado los siguientes formatos de CAD: DGN (ISFF), DWG y DXF, BNA (Atlas BNA) y BIN (DIGI21). Los principales ítems de metadatos que pueden leerse o calcularse de este tipo de formatos son:

- Las coordenadas máximas y mínimas del rectángulo envolvente,
- La altimetría (máxima y mínima),
- El número y tipo de geometrías existentes.

Además, algunos formatos agrupan la información por capas, en cuyo caso los nombres de las mismas pueden ser útiles para identificar los tipos de fenómenos en el catálogo. La tabla 5.5 recoge los ítems de metadatos de los formatos usados para almacenar este tipo de información.

Formato	Extensión geográfica	Nº de capas	Nombre de las capas	Nº de tipos de fenómenos	Nombre y nº de los fenómenos	Otros
DGN	X	X		X	X	Puede ir acompañado del archivo CSF, conteniendo información del sistema de referencia espacial
DXF	X	X	X	X	X	
DWG	X	X	X	X	X	Puede existir información relativa al sistema de referencia espacial, si se trata de Autocad Map
BIN	X	X <sup>(4)</sup>	X <sup>(4)</sup>	X	X	
BNA	X	X	X	X	X	

Tabla 5.5: Formatos vectoriales tipo CAD

<sup>(4)</sup> Semántica equivalente a nombre de capa

## 5.2.6 Formatos vectoriales tipo SIG

A diferencia de los formatos CAD, los formatos SIG pueden vincular información tabular sobre las geometrías e independizar la visualización. Una primera aproximación a esta solución consistió en asociar mediante un identificador las geometrías con una fila de una tabla. Este es el caso de los proyectos MGE (*Intergraph*) o el formato *Shapefile*. En el primer caso, la tabla está almacenada en algún gestor de bases de datos y, en el segundo, se trata de un archivo DBF.

Otros formatos almacenan las geometrías y los atributos asociados a las mismas conjuntamente. También lo hacen así las bases de datos espaciales almacenando en la misma tabla geometrías y atributos. El conjunto de formatos de ficheros analizados es: E00, SHP y ADF (*Esri*), MIF y TAB (*MapInfo*), GML y KML

(*OpenGIS*), GMT, GRASS, SDTS, UK-NTF, Tiger-Line, Interlist, GeoConcept, GeoJSON y SDF.

Los principales ítems que pueden obtenerse de este tipo de formatos son:

- Las coordenadas máximas y mínimas del rectángulo envolvente,
- El número de entidades existente para cada tipo de geometría (puntos, arcos, polilíneas, polígonos, textos, etc.),
- El nombre de las capas en que están organizadas las entidades (nombre de las entidades),
- El sistema de referencia espacial

Algunos formatos contienen también información relativa a los procesos aplicados sobre la información, las fechas de los mismos y las fuentes de información utilizadas. La tabla 5.6 recoge los ítems de metadatos de los formatos usados para almacenar este tipo de información.

Formato	Extensión geográfica	Nº de capas	Nombre de las capas	Nº de tipos de fenómenos	Nombre y nº de los fenómenos	Sistema de referencia espacial	Nombre y tipo de atributos	Otros
E00 arc	X					X	X	
ADF arc	X			X	X	X	X	
SHP	X	X	X	X	X	X	X	
MIF	X			X	X	X	X	
TAB	X			X	X	X	X	
VEC (idrisi)	X			X	X	X		
GML	X	X	X	x	X	X	X	
KML	X	X	X	X	X	X fijo	X	
GMT	X	X	X	X	X	X		
GRASS	X	X	X	X	X	X		
SDTS	X	X	X	X	X	X		
UK-NTF	X	X	X	X	X	X		
TIGER	X	X	X	X	X	X		
GeoConcept	X	X	X	X	X	X		
GeoJSON	X	X	X	X	X	X		
SDF (mapguide)	X	X	X	X	X	X	X	

Tabla 5.6: Formatos vectoriales tipo SIG

### 5.2.7 Bases de datos con extensiones espaciales

Como se ha expuesto, la tendencia actual consiste en que los SGBD más extendidos y usados dispongan de tipos de datos capaces de almacenar datos geométricos (puntos, líneas, polígonos) y procesarlos mediante operadores

espaciales. Estos operadores son considerablemente más complejos que los alfanuméricos (Yeung y Brent, 2007). Los objetos espaciales que comúnmente se conocen como “geometrías”, son el mecanismo que permite representar datos espaciales. Desde un punto de vista matemático, el concepto de geometría está relacionado con las propiedades y las relaciones entre puntos, líneas, ángulos, superficies y sólidos en espacios de una o más dimensiones. Desde un punto de vista de estandarización, la especificación definida por el OGC (1999) para objetos espaciales simples (*Simple Feature*), junto a versión especializada para SQL (OpenGIS Simple Feature Specification for SQL), identifica los formatos de texto y binarios para la representación de los objetos y define las estructuras de las tablas que deben dar soporte a los metadatos mínimos. Se habla comúnmente del modelo *GeoDatabase*, que aprovecha las capacidades de las bases de datos para manejar índices, definir restricciones y mantener la integridad de los datos espaciales, a la par de dotar de mecanismos para gestionar las transacciones. La tabla 5.7 recoge los ítems de metadatos de los formatos usados para almacenar este tipo de información.

SGBD	Extensión geográfica	Nº de capas	Nombre de las capas	Nº de tipos de fenómenos	Nombre y nº de los fenómenos	Sistema de referencia espacial	Nombre y tipo de atributos	Otros
PostGis	X	X	X	X	X	X	X	
MySQL	X	X	X	X	X		X	
Oracle	X	X	X	X	X	X	X	
DB2	X	X	X	X	X	X	X	
Informix DataBlade	X	X	X	X	X	X		
Microsoft SQL 2008	X	X	X	X	X	X	X	
SQLite (spatialLite)	X	X	X	X	X	X	X	
INGRES	X	X	X	x	X	X		
ArcSDE (MsAccess, Oracle, SQL Server, DB2 e Informix)	X	X	X	X	X	X	X	Catálogo de entidades
MapInfo (spatialware) SQL Server	X	X	X	X	X	X	X	
SQLite + SpatiaLite	X	X	X	X	X	X	X	
SQLite + OGR-FDO	X	X	X	X	X	X	X	
H2 + Spatial DB in a Box	X	X	X	X	X	X	X	
HSQldb + Spatial DB	X	X	X	X	X	X	X	
Derbi + Spatial DB	X	X	X	X	X	X	X	

Tabla 5.7: Bases de datos con extensiones espaciales

### 5.2.8 Heterogeneidad en el almacenamiento de los Sistemas de Referencia Espacial

Como se ha venido indicando en capítulos anteriores existen diferentes modos para representar e identificar el Sistema de Referencia Espacial (SRE) al que están referidas las coordenadas. Esta variedad es consecuencia de la cantidad de formatos de almacenamiento de IG existentes; se trata de un problema de heterogeneidad, tanto sintáctica como semántica, relacionado con los parámetros que identifican unívocamente un SRE. Estos son (a) el datum geodésico de referencia, definido por el elipsoide y el origen de coordenadas, y (b) la proyección cartográfica, junto a los parámetros que la caracterizan. Para ilustrar la mencionada heterogeneidad, se enumeran y describen a continuación las distintas formas usadas por los diferentes formatos para expresar el SRE. Asociados a dichas formas se muestran en las tablas 5.8 – 5.13 algunos casos concretos de SRE. A continuación se describen en líneas generales el formato de almacenamiento de los archivos:

- Los formatos GeoTIFF, GeoJP2 y MrSID almacenan valores numéricos estandarizados y definidos en un diccionario público (p.e. EPSG), como se describe en la tabla 5.8.
- Otros formatos utilizan valores numéricos definidos en diccionarios particulares; éste es el caso de los formatos: MIF, DAT, DEM, CSF, DGN, IMG, RAS y LAN, que pueden observarse en la tabla 5.9.
- Un tercer tipo de formatos utilizan nemotécnicos definidos en diccionarios particulares como: ECW, ERS, F-EOSAT, NITF y PIX. Véase la tabla 5.10.
- Un cuarto grupo de formatos, entre los que están ADF, E00, GRD y Proj4, utilizan formato de texto en una representación cuasi-estructurada, como puede observarse en la tabla 5.11.
- Un quinto grupo de formatos, entre los que están SHP y GRD ASCII utilizan la sintaxis estructurada *Well-Known Text* (WKT) definida por el OpenGIS. Véase la tabla 5.12.
- Y, finalmente, hay un sexto grupo, entre el que se encuentran las bases de datos con extensiones espaciales: PostGIS, Oracle, DB2, SQL Server 2008, *MapInfo SpatialWare* (como puede observarse en la tabla 5.13), que utilizan valores numéricos estandarizados y, junto a codificaciones también numéricas

o de texto particulares, intentan paliar la heterogeneidad semántica y facilitar su interpretación.

Sistema de Referencia por Coordenadas	EPSG	TIFF tag
Datum Europeo de 1950, Proyección UTM Hemisferio Norte Huso 30	23030	3072
Coordenadas Geográficas 2D, Datum ETRS89	4258	3072
Datum ETRS89, Proyección UTM Hemisferio Norte Huso 30	25830	3072

Tabla 5.8: Ejemplos de descripción numérica estandarizada: GeoTIFF, MrSID

Sistema de coordenadas	Formato	Datum	Elipsoide	Proyección
Datum Europeo 1950 UTM	Intergraph	4	5	7
Datum Europeo 1950 UTM	Erdas	Texto	5	1
Datum Europeo 1950 UTM	Mentor	Texto	Texto	46
Datum Europeo 1950 UTM	Mapinfo	28	Texto	8

Tabla 5.9: Ejemplos de descripción numérica no estandarizada

Proyección	Proj4	PCI	FME	Ermapper
OBLIQUE MERCATOR	OMER	OM	HOM10V	obmerc_b
LAMBERT AZ EQUAL AREA	LAEA	LAEA	AZMEA	lambazea
EQUIDISTANTE CYLINDRICAL	EQC	ER	EDCYL	-

Tabla 5.10: Ejemplos de descripción nemotécnica

DB2 & Esri	ECW	MapInfo & Oracle
GCS_European_1950	European_datum_1950	Longitude /latitude (ED50)
GCS_North_American_1927	North_American_1927	Longitude /latitude (NAD27)
GCS_WGS_1984	WGS_1984	Longitude /latitude (WGS 84)

Tabla 5.11: Ejemplos de descripción mediante texto cuasi-estructurados

WKT
<pre> PROJCS["ED50 / UTM zone 30N",   GEOGCS["ED50",     DATUM["European_Datum_1950",       SPHEROID["International 1924",6378388,297]],     PRIMEM["Greenwich",0],     UNIT["degree",0.01745329251994328]   ],   PROJECTION["Transverse_Mercator"],   PARAMETER["latitude_of_origin",0],   PARAMETER["central_meridian",-3],   PARAMETER["scale_factor",0.9996],   PARAMETER["false_easting",500000],   PARAMETER["false_northing",0],   UNIT["metre",1] ]</pre>

Tabla 5.12: Ejemplos de descripción mediante texto estructurado (WKT)

Base de datos	Esquema de tabla y ejemplos valores					
POSTGIS	<b>srid</b>	<b>auth_name</b>	<b>auth_srid</b>	<b>srtext</b>	<b>proj4text</b>	
	32636	EPSG	32636	PROJCS["WGS 84 / UTM zone 36N", ...	+proj=utm +zone=36 +ellps=WGS84 +datum=WGS84 +units=m +no_defs	
Oracle	<b>Cs_name</b>	<b>Srid</b>	<b>Auth_srid</b>	<b>Auth_name</b>	<b>Wktext</b>	
	WGS 84 / UTM zone 36N	32636	32636		PROJCS["WGS 84 / UTM zone 36N", ...	
DB2	<b>Coordsys_name</b>	<b>Organization_coordsys_id</b>	<b>Organization</b>	<b>Definition</b>		
	WGS_1984_UTM_ZONE_36N	32636	EPSG	PROJCS[WGS_1984_UTM_Zone_36N", ...		
SQL Server 2008	<b>spatial_reference_id</b>	<b>authority_name</b>	<b>authorized_spatial_reference_id</b>	<b>well_known_text</b>	<b>unit_of_measure</b>	<b>unit_conversion_factor</b>
	4326	EPSG	4326	GEOGCS["WGS 84", ..	metre	1
MapInfo SpatialWare	<b>cs_name</b>	<b>srid</b>	<b>Auth_srid</b>	<b>Auth_name</b>	<b>srtext</b>	
	UTM Zone 36, Northern Hemisphere (WGS 84)	82356	82356	MapInfo	PROJCS[UTM Zone 36, Northern Hemisphere (WGS 84)", GEOGCS..	

Tabla 5.13: Ejemplos de descripciones numéricas en bases de datos espaciales

Una vez descrito el problema de heterogeneidad en la representación del SRE en los formatos de almacenamiento de la IG se muestran, a continuación, algunos datos cuantitativos que ilustran la magnitud del problema. En la tabla 5.14 se resume la información de los formatos, el tipo de codificación usada en el SRE y la cantidad de elementos (diferentes datum, elipsoides, proyecciones, parámetros, etc.) o definiciones completas de SRE existentes en cada caso.

Forma de identificar SRE	Fuente	Cantidad
Numérica	Base de datos EPSG (v6.18.2)	~ 4.362
WKT + numérica + proj4	PostGIS (v1.3.5)	~ 3.162
Numérico + WKT	Microsoft (SQL Server 2008)	~390
Numérica + WKT	Oracle (v10g)	4.384
Numérica + WKT	IBM-DB2	2.360
Numérica + WKT	MapInfo SpatialWare (for SQL Server)	950
Nemotécnicos + WKT	Esri	2.400
WKT	Ermapper	875
Nemotécnicos	Ermapper	165
Numéricas	Intergraph	190
Numéricas	MapInfo	270
Nemotécnicos	PCI Geomatics	290
Nemotécnicos	FME	338
Nemotécnicos	Proj4	193
Nemotécnicos	Erdas Imagine	254
Numéricos	Idrisi	430
Texto estructurado	Mentor	~1890
Numéricas + Nemotécnicos para proyecciones	GCTP (General Cartographic Transformation Package)	54

Texto Datum + proyecciones	TouratechQV	~280
Texto Datum + proyecciones	OziExplorer	~150
Texto Datum + proyecciones	CompeGPS	~150

Tabla 5. 14: Tipos y cantidades de identificadores diferentes en los formatos

Para enmarcar el problema de los SRE se han identificado las organizaciones de estandarización y de consenso que existen en este ámbito, siendo el organismo internacional de estandarización ISO con la norma ISO19111:2007 *Spatial Referencing by Coordinates* y el OGC en la especificación *Spatial Referente by Coordinates* (SRbC) quienes abordan este tema. Ambas organizaciones también tratan esta información en otras normas y estándares, entre los que cabe destacar: el esquema de datos *MD\_ReferenceSystem* definido en la norma ISO19115:2003 *Geographic Information: Metadata* y el esquema de datos *coordinateReferenceSystems* definido en el estándar *Geographic Mark-Up Language* (GML) de OGC (Versión 3.1.1).

También se ha identificado la existencia de diccionarios públicos que definan y describan de un modo preciso los SRE, pudiéndose citar como trabajo pionero el realizado por el *European Petroleum Surveyor Group* (EPSG), con la base de datos de parámetros geodésicos. El EPSG actualmente está integrado en la asociación internacional de productores de gas y petróleo (*International Assosiation of Oil & Gas Producers*: OGP), que mantiene y proporciona libremente una base de datos que contiene este diccionario, a la vez que facilita el acceso a las definiciones de los SRE mediante un catálogo accesible en Internet como Servicio Web.

Desde un punto de vista práctico, OGC e ISO, en las normas relacionadas con los GeoServicios para los que se precisa definir de un modo conciso los SRE, proponen el uso de los valores numéricos del diccionario o codeSpace EPSG (p.e. EPSG:23030).

### 5.3 Creación automática de metadatos

A continuación se identifican los ítems de metadatos que pueden crearse de forma automática, en base a los contenidos de forma implícita en la IG, a los calculados, los que se pueden inferir y los que se pueden obtener del contexto.

### *MD\_Metadata*

*---fileIdentifier* Es habitual que el Identificador Único Universal (UUID) se construya siguiendo una reglas de codificación que permitan identificar a la organización y, dentro de ella, al producto y elemento. Por esta razón, puede generarse automáticamente en base a información del contexto que identifique la organización (parámetro configurable del método, consulta inversa al servicio de nombres de dominio, etc.).

*---language* Se trata del idioma en el que se expresa el metadato. De la misma forma que en el anterior ítem, puede obtenerse de un parámetro de configuración del sistema de generación de metadatos o asumirse un idioma concreto (spa, eng). Si se dota de capacidad de traducción entre idiomas al sistema de generación, se puede crear el metadato en varios idiomas, rompiendo otra barrera adicional.

*---characterSet* Se puede establecer, como parámetro de configuración del sistema de generación de metadatos, el tipo de codificación de los caracteres a utilizar para materializar el resultado en un archivo. Se propone el uso del utf8 como valor por defecto, ya que se trata de la codificación más extendida y usada.

*---hierarchyLevel* El valor de este ítem depende del contenido del URI proporcionado como entrada del método. Si hace referencia a un archivo, su valor será un conjunto de datos (*dataSet*). Si proporciona los datos de acceso a una base de datos o hace referencia a un directorio de datos vectoriales, el generador de metadatos trata cada tabla o fichero del directorio como una entidad (*featureType*) y, además de generar un metadato por cada tabla o archivo, se puede generar uno más que describa el conjunto de entidades y sus relaciones, materializando el modelo de datos obtenido, deducido o inferido para el URI.

*---contact (CI\_ResponsibleParty)* En raras ocasiones la información de contacto relativa a los metadatos es implícita o puede calcularse, aunque también es cierto que suele ser común a toda la IG de una organización y, por tanto, puede obtenerse como parámetro configurable del creador de metadatos. Además, estos valores se utilizarán si se crea un nuevo metadato y no se proporciona ninguna plantilla.

*-----individualName* idem.

-----*role (CI\_RoleCode)* se propone usar el valor “*author*”

---*dateStamp* Se establecerá o actualizará el valor del ítem con la fecha y hora del sistema en el que se esté usando el generador.

---*metadataStandardName* Nombre del estándar: ISO19115/19139

---*metadataStandardVersion* Versión del estándar: ISO-19139:2007

---*datasetURI* Se puede incorporar el valor del URI de acceso a los datos, ya que es el mecanismo que utiliza el propio creador para acceder a ellos. Dependiendo de los niveles de seguridad y de las políticas de acceso/uso de los datos, cierta información del URI (usuario y contraseña p.e.) debe omitirse.

---*distributionInfo (MD\_Distribution)*

-----*transferOptions*

-----*unitsOfDistribution* No es un ítem de metadato implícito y resulta, en ocasiones, difícil de calcular o inferir. Esto ocurre porque ciertos formatos se utilizan para almacenar conjuntos de datos que pueden representar una capa completa, un elemento de una tesela o simplemente, un área geográfica. Como primera aproximación se propone utilizar los valores usuales. Ejemplo: si hay varios archivos con imágenes y su tamaño es regular, se trataría de un elemento de una tesela (“*tile*”); es igualmente válido el razonamiento si los archivos son de tipo CAD (p.e. *dgn*) y tienen un tamaño regular. Los archivos de tipo *shapefile* contienen generalmente coberturas temáticas completas, por tanto se trata de capas (“*layer*”); las bases de datos no utilizan teselas y las unidades de distribución son áreas geográficas (“*geographical area*”) o coberturas temáticas (“*layer*”).

-----*transferSize* Cuando se trata de archivos, el creador de metadatos puede acceder al sistema de ficheros para solicitar su tamaño y transformar las unidades a MBytes. En las bases de datos se pueden realizar algunas consultas dependientes del SGBD, con mayor o menor complejidad, que permiten conocer el tamaño físico de disco que ocupa cada tabla, aunque no se pueda conocer el tamaño del archivo para su transferencia.

-----*distributionFormat*

-----*name* Se puede indicar el nombre del formato cuando se trata de archivo y, por el contrario, el nombre del formato de distribución no se conoce para las bases de datos.

-----*version* Algunos formatos de archivo incluyen ítems de información que permiten identificar la versión usada del formato y, por tanto, del formato de distribución.

-----*fileDescompresionTechnique* Algunos formatos poseen ítems de información que identifican la técnica de compresión/descompresión usada. Éste es el caso de los formatos GTIFF, MrSID, JP2 o ECW en el contexto de las imágenes.

---*contentInfo (MD\_ContentInformation)*

-----*MD\_CoverageDescription* Este paquete de información descriptiva se aplica a los formatos de archivo que almacenan imágenes o bases de datos ráster.

-----*contetType* Algunos formatos como Erdas e Idrisi permiten diferenciar el tipo de contenido, al nombrar las categorías de información temática almacenada y, por tanto, permiten identificar su contenido como clasificaciones temáticas (“*thematicClassification*”). La mayoría de los formatos usados para las imágenes sólo soportan el tipo imagen (“*image*”). Los formatos de archivo que contienen información multiespectral, usualmente utilizados en teledetección, pueden contener datos como la reflectancia y la transmitancia, que pueden considerarse medidas físicas (“*physicalMeasurement*”).

-----*dimension (MD\_RangeDimension)* La cobertura está compuesta por una o más bandas. Para cada una de ellas se incorpora la siguiente información que la describe.

-----*sequeceIndetifier* Número de secuencia de la banda en la cobertura.

-----*descriptor* Nombre o número asignado a la banda/capa.

-----*MD\_Band*

-----*maxValue* Valor máximo de los píxeles. Si el formato no lo almacena, se puede realizar el cálculo.

-----*minValue* Valor mínimo de los píxeles. Si el formato no lo almacena, se puede realizar el cálculo.

-----*units* Unidades de medida para los valores máximo y mínimo. Es un valor difícil de obtener.

-----*bitsPerValue* Número de bits usados para codificar los valores de cada píxel.

---*MD\_FeatureCatalogueDescription* Catálogo de objetos, aplicado tanto a datos vectoriales como a las bases de datos o a las imágenes temáticas.

-----*complianceCode* Actualmente, los formatos no incluyen información que indique si los datos almacenados responden a especificaciones o estándares concretos. Las reglas de implementación de la directiva INSPIRE indican que este ítem es obligatorio. Cuando se trate de un directorio o de una base de datos, una vez deducido el modelo de datos, se puede establecer una comparación con las descripciones de los objetos en un catálogo de tipos de fenómenos.

-----*Language* Se puede deducir el idioma de los datos a partir del análisis de los mismos, utilizando diccionarios.

-----*includedWithDataSet* No es común que el catálogo de objetos se almacene junto a los datos, adoptando como valor “*false*”. Algunos formatos de imágenes como Erdas e Idrisi o las geodatabases sí pueden almacenarlo. Una vez comprobado que lo hacen, se asigna el valor “*true*” al ítem. Los formatos usados en CAD (Dgn, Dwg, Dxf, MIF) suelen denominar a las capas de los archivos con el nombre o nemotécnico asociado al tipo de entidades almacenadas. En estos casos, también se puede asignar el valor *true* a este ítem.

-----*featureTypes* Si se pueden determinar los nombres de los tipos de entidades contenidos en el conjunto de datos, se incluirán como tipos de entidades del mismo.

---*spatialRepresentationInfo (MD\_SpatialRepresentation)*

-----*MD\_GridSpatialRepresentation* Aplicado a imágenes y datos matriciales.

-----*numberOfDimensions* Número de dimensiones en espacio y tiempo. En las imágenes y capas matriciales (bandas) suele coincidir con el número de bandas.

-----*cellGeometry* Tipo de geometría: superficie “*area*” o punto “*point*”

-----*transformationParametersAvailability* Algunos formatos, además de almacenar la imagen rectificadas, almacena también la información que permite

realizar dicha transformación. Este es el caso de los formatos de Erdas, o algunos otros como L1B; en estos casos se puede indicar que si están disponibles los parámetros de transformación “*true*” y en el resto de casos lo no están “*false*”.

-----*axisDimensionProperties* (*MD\_Dimension*)

-----*dimensionName* Nombre de la dimensión: fila (“*row*”), columna (“*column*”), vertical (“*vertical*”) o duración temporal (“*time*”). Los casos más comunes serán los dos primeros y, excepcionalmente, se dispondrá de información para completar los otros.

-----*dimensionSize* Se ha de indicar el conteo de filas y columnas de las imágenes, así como el número de dimensiones adicionales: vertical o temporal que puedan existir.

-----*resolution*

-----*resolution* Tamaño del píxel según la dimensión: fila, columna, etc.

-----*unitsOfResolution* Unidades de la medida resolución.

-----*MD\_VectorSpatialRepresentation* Se usa en el caso de datos vectoriales

-----*topologicLevel* Nivel topológico de los datos. En los formatos vectoriales usados en CAD no suele tratarse la topología, en cuyo caso se indicaría que sólo se consideran las geometrías “*geometryOnly*”. El nivel topológico puede adoptar otros valores como: “*topology1D, planarGraph fullPlanarGraph, surfaceGraph, fullSurfaceGraph, topology3D, fullTopology3D or abstract*”, en función del tipo de contenido. Este tipo de información no se encuentra descrita en los metadatos implícitos de los datos.

-----*geometricObjects*

-----*geometricObjectType* Se identifican, mediante los códigos que se enumeran, los tipos de geometrías contenidas en los datos: “*complex, composite, curve, surface, point or solid*”.

-----*geometricObjectCount* Se indica el número de objetos de cada tipo de geometrías.

---*identificationInfo* (*Md\_Identification*)

-----*citation* (*CI\_Citation*) Cita que permite indicar el título del recurso.

-----*title* Título propuesto para el recurso. El generador de metadatos debe de proponer un título para el recurso, basándose en la información implícita obtenida, los datos calculados y los inferidos.

-----*date* Fecha de creación del título

-----*dateType (CI\_DateTypeCode)* Código que identifica el tipo de fecha. Si el metadato no existía será creación “*creation*”, y si ya existía se revisa “*revision*”

-----*otherCitationDetails* Aunque este ítem no es obligatorio puede usarse para indicar cómo se ha generado el título del metadato, es decir, describir los elementos que se han considerado para generar el título. También se puede referenciar al creador de metadatos.

-----*pointOfContact (CI\_ResponsibleParty)*

-----*individualName* Se incluirá la identificación del responsable o autor de los datos si se dispone o se puede obtener información relativa. Algunos formatos contienen dicha información, y en otros casos ésta podría ser completada con información parametrizada por defecto que adopte el creador de metadatos.

-----*role (CI\_RoleCode)* Generalmente se dispondrá de información sobre el autor de los datos y el código a utilizar será “*author*”.

-----*graphicOverview (MD\_BrowseGraphic)*

-----*fileName* El generador de metadato calculará una vista previa de los datos, que será almacenada en un archivo. Se trata del nombre del archivo que se genera.

-----*fileDescription* De la misma forma que asociado al título del recurso se pueden incluir otros detalles de la cita y se puede describir cómo se ha generado el título del metadato, en este caso se puede describir el gráfico y el modo en el que se ha generado el mismo.

-----*fileType* El tipo de fichero será el que el creador de metadatos utilice para almacenar la vista previa. Se utilizarán formatos soportados por la Web y cuyo tratamiento esté extendido; por ejemplo: PNG o JPEG.

-----*resourceFormat (MD\_Format)*

-----*name* Se puede indicar el nombre del formato cuando se trata de archivo y *desconocido* si se trata de las bases de datos.

-----*version* Algunos formatos de archivo incluyen ítems de información que permiten identificar la versión usada del mismo y, por tanto, del formato de distribución.

-----*fileDescompresionTechnique* Algunos formatos poseen ítems de información que identifican la técnica de compresión/descompresión usada. Éste es el caso de los formatos GTIFF, MrSID, JP2 o ECW en el contexto de las imágenes.

-----*resourceSpecificUsage (MD\_Usage)*

-----*userDeterminedLimitations* Algunos formatos almacenan información, notificando de las limitaciones de uso, por tanto, se incluye dicha información.

-----*resourceConstraints (MD\_Constraints)*

-----*useLimitation* Se trata de un ítem que puede parametrizar en la configuración del creador de metadatos para responder a la política de uso de los datos de la organización. De este modo, se establecería un valor predeterminado para el ítem.

-----*accessConstraints (MD\_RestrictionCode)* Se trata de conocer el código que identifica las restricciones de acceso a los datos. En algunos formatos se almacena información que define la política de acceso. Los posibles valores de este ítem son “*copyright, patent, patentPending, trademark, license, intellectualPropertyRights, restricted, otherRestrictions*”.

-----*useConstraints (MD\_RestrictionCode)* Algunos formatos como PNG permiten almacenar restricciones de uso.

-----*otherConstraints* Como ocurre con el ítem *useLimitation* su valor puede obtenerse de la parametrización del creador de metadatos, en sintonía con la política sobre los datos de la organización. En este caso puede ser un mensaje de tipo cláusula de exención de responsabilidad (*disclaimer*) sobre el uso inadecuado de los datos.

-----*classification (MD\_ClassificationCode)* Código que clasifica la confidencialidad de los datos. En algunos casos, los formatos contienen este tipo de información almacenada. Se han de usar los valores “*unclassified, restricted, confidential, secret, topSecret*”.

-----*descriptiveKeywords (MD\_Keywords)*

-----*keyword* Una vez que el módulo del creador de metadatos ha identificado el tipo de contenido de los datos, el módulo catalogador puede seleccionar un conjunto de palabras clave “*keywords*”, que permitan identificar más fácilmente el contenido.

-----*type* Se ha de indicar el tipo de la palabra clave propuesta, pudiendo pertenecer a las siguientes clases: “*discipline, place, stratum, temporal, theme*”. El módulo catalogador, además de seleccionar las palabras clave, identifica el tipo y el tesoro al que pertenece.

-----*thesaurusName (CI\_Citation)*

-----*title* Título del tesoro que contiene la palabra clave usada para catalogar el recurso.

-----*date* Fecha del tesoro, según sea creación, revisión, etc.

-----*dateType (CI\_DateTypeCode)* Código que identifica el tipo de fecha anterior: “*creation, publication, revision*”

-----*spatialRepresentationType* En función del formato y el tipo contenido, el tipo de representación espacial de los datos puede clasificarse en: “*vector, grid, textTable, TIN, stereoModel, video*”.

-----*spatialResolution (MD\_Resolution)* Escala de los datos.

-----*equivalentScale (MD\_RepresentativeFraction)*

-----*denominator* Basado en el tamaño del píxel para las imágenes y los datos matriciales, se puede calcular el denominador de la escala equivalente.

-----*distance (Distance)* Tamaño del píxel en las imágenes y datos matriciales.

-----*language* El idioma en el que están expresados los datos puede inferirse al utilizar diccionarios, con los que comparar valores de los atributos de los datos.

-----*characterSet* En las bases de datos y algunos formatos de archivo, se puede obtener información que identifique el juego de caracteres usado en los datos.

-----*topicCategory* Cuando el creador de metadatos ha determinado el tipo de contenido de los datos, el módulo catalogador puede seleccionar uno o varios temas de la lista de categorías en las que la norma de metadatos clasifica los datos.

-----*environmentDescription* El creador de metadatos puede acceder a la información que le proporcione el Sistema Operativo en el que se ejecute, para describir el entorno de procesamiento en el que se encuentran los datos, el tipo de sistema operativo, sistema de ficheros, base de datos, etc.. Se trata de información de contexto.

-----*supplementalInformation* Algunos formatos de archivos usados en teledetección contienen información adicional relativa a la plataforma y los sensores que han captado los datos. Parte de esta información, puede almacenarse en la nueva norma de metadatos que, a lo largo del año 2009, se aprobará ISO19115-2. Otra información pueden incluirse en este ítem.

-----*extent (Ex\_Extent)*

-----*description* Texto descriptivo que explica cómo se han obtenido y/o calculado los valores almacenados para las extensión geográfica.

-----*geographicElement (EX\_GeographicExtent)*

-----*EX\_GeographicBoundingBox*

-----*extentTypeCode* Se asigna el valor *true* para indicar que la extensión geográfica es el interior del rectángulo que se indica a continuación.

-----*westBoundLongitude* Los valores de las coordenadas norte, sur, este y oeste se obtienen del módulo conversión/transformación de coordenadas, una vez identificado el sistema de referencia espacial de las mismas.

-----*eastBoundLongitude* Idem

-----*northBoundLatitude* Idem

-----*southBoundLatitude* Idem

-----*EX\_GeographicDescription*

-----*extentTypeCode* Se asigna el valor *true* para indicar que la extensión geográfica se corresponde con el nombre del lugar que a continuación se relaciona.

-----*geographicIdentifier (RS\_Identifier)*

-----*authority (CI\_Citation)*

-----*title* Nombre o título del tesoro de lugar usado.

-----*date* Fecha del tesoro.

-----*dateType(CI\_DateTypeCode)* Código que identifica el tipo de fecha del tesoro. Los posibles valores son: “*creation, publication, revision*”

-----*code* Valor del nombre de lugar. Se obtendrá utilizando un *geocoder* inverso, de modo que dadas las coordenadas del BBOX, se solicite el topónimo más relevante del lugar. Se buscará nombres de ciudades o núcleos de población.

-----*verticalElement (EX\_VerticalExtent)* Usualmente, para conjuntos de datos 3D y para mallas de datos regulares conteniendo modelos digitales del terreno o de elevaciones.

-----*minimumValue* Altura mínima contenida en el formato o calculada.

-----*maximumValue* Altura máxima contenida en el formato o calculada

-----*unitOfMeasure* Unidades de las medidas. Suele ser el metro, aunque dependerá de cada formato.

---*dataQualityInfo (DQ\_DataQuality)*

-----*scope (DQ\_Scope)*

-----*level (MD\_ScopeCode)* Generalmente, los formatos que incluyen este tipo de información suelen describir los pasos del proceso, indicando los conjuntos de datos afectados. Por esta razón, el nivel al que se aplican es “*dataSet*”, aunque está previsto que se puedan detallar los siguientes niveles de detalle: “*attribute, series, features*”.

-----*linage (LI\_Lineage)*

-----*source (LI\_Source)*

-----*description* Algunos formatos incluyen una descripción de las fuentes de datos utilizadas, por ejemplo IDRISI.

-----*processStep (LI\_ProcessStep)*

-----*description* Algunos formatos almacenan los pasos realizados sobre los datos.

-----*dateTime (DateTime)* Algunos formatos también lo incluyen.

---*applicationSchemaInfo (MD\_ApplicationSchemaInformation)*

-----*name (CI\_Citation)*

-----*title* Descripción del método usado para crear o inferir el esquema de aplicación.

-----*date* Fecha de creación del esquema de aplicación.

-----*dateTimeCode (CI\_DateTypeCode)* Tipo de fecha del esquema de aplicación. Si no existía el metadato y se ha generado el esquema de aplicación, será “*creation*”, en otros casos será “*revision*”

-----*schemaLanguage* Los lenguajes que utiliza el generador de metadatos para definir los esquemas son (*UML* y *GML-Schema*).

-----*constraintLanguage* El generador de metadatos utiliza como lenguajes de restricciones el lenguaje de modelado unificado y el lenguaje de marcas para datos geográficos (*GML*).

-----*schemaAscii* El generador de metadatos, con la ayuda del módulo que determina el modelo de datos, almacenará el esquema de aplicación de los datos en formato *ASCII*.

-----*graphicsFile* El generador de metadatos creará una vista previa del esquema de aplicación y lo almacenará en un formato gráfico.

-----*softwareDevelopmentFile* Se hace referencia al fichero XML que contiene el esquema de aplicación, ya sea el modelo UML, en cuyo caso se almacena en formato XMI, o sea el esquema de aplicación GML, en cuyo caso se almacena en formato XSD.

-----*softwareDevelopmentFileFormat* En un caso será XMI (UML) y en otro XSD (GML Schema).

---*referenceSystemInfo (MD\_ReferenceSystem)*

-----*referenceSystemIdentifier (RS\_Identifier)*

-----*codeSpace* *EPSG* Se establece como diccionario de códigos que identifican el sistema de referencia espacial.

-----*version* Versión de la base de datos EPSG usada.

-----*authority* (*CI\_Citation*)

-----*title* Se hace referencia a la asociación internacional de productores de gas y petróleo (OGP) y, en especial, al *European Petroleum Surveyor Group* (EPSG), quien ha definido una base de dato de parámetros Geodésicos.

-----*date* (*CI\_Date*)

-----*date* Fecha de la base de datos EPSG utilizada.

-----*dateType* (*CI\_DateTypeCode*) “*revision*”

-----*code* El módulo que identifica el sistema de referencia espacial para las coordenadas debe de obtener el código EPSG, que identifica las mismas.

#### **5.4 Funciones desempeñadas y niveles de interoperabilidad favorecidos por los ítems de metadatos creados automáticamente**

En el apartado anterior se ha realizado una descripción detallada de los posibles ítems de metadatos que se pueden extraer, calcular o inferir para un determinado conjunto de datos en una IDE. Es difícil determinar el número exacto de ítems que pueden generarse automáticamente para un repositorio de información geográfica, ya que cada tipo (imágenes aéreas, imágenes multiespectrales, DTM, archivos de dibujo y capas vectoriales) tiene asociados un conjunto diferente de ítems de metadatos. A este hecho se ha de añadir que la información que puede extraerse de los archivos y almacenes de datos, depende estrechamente del formato de almacenamiento usado.

En la tabla 5.15 se han identificado, de manera optimista, los ítems de metadatos que podrían crearse automáticamente con la metodología propuesta. En la primera columna de la tabla se muestra la clase a la que pertenece el ítem y la segunda identifica el ítem. La tercera columna clasifica los ítems como “C” creado (extraído, calculado o inferido), “N” con cardinalidad, dependiente del dataset y, en último lugar, “F” si se puede establecer un valor fijo por convenio o dependiente del contexto. La cuarta columna identifica aquellos ítems que sólo se

aplican a determinado tipo de IG (“R” datos ráster, “D” modelos digitales del terreno y “V” datos vectoriales).

La quinta columna de la tabla identifica la función que desempeña el ítem de metadato (“L” localización, “E” evaluación, “A” acceso y “U” uso).

En las columnas 6 - 12 se describen los siete niveles de interoperabilidad: *técnica*, *sintáctica*, *semántica*, *pragmática*, *dinámica*, *conceptual* y *organizacional*, definidos en el Modelo Integrado de interoperabilidad para las IDE.

MD_Metadato: Paquete de metadatos	Item del metadato	(C) Creado. (N) depende de los datos (F) Fijado	(R: Ráster, V: Vectorial, D: DTM)	Función: (L: localización, E: evaluación, A: acceso, U: uso)	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
	language	C					X				X
	characterSet	F			X		X				
	hierarchyLevel	C					X			X	X
	metadataStandardName	F				X	X				X
	metadataStandardVersion	F				X	X				X
	dataSetURI	C			X		X	X			
	dateStamp	C					X		X		X
distributionInfo:transferOptions	unitsOfDistribution	C		A			X	X			X
distributionInfo:transferOptions	transferSize	C		A	X		X				X
distributionInfo:distributionFormat	name	C		A	X		X		X		X
distributionInfo:distributionFormat	version	C		A	X				X		X
distributionInfo:distributionFormat	fileDecompressionTechnique	C		A			X		X		X
contentInfo:MD_CoverageDescription	contentType	C	R	U			X		X		X
contentInfo:MD_CoverageDescription: dimension:MD_RangeDimension	sequenceIdentifier	N	R	U			X		X		X
contentInfo:MD_CoverageDescription: dimension:MD_RangeDimension	descriptor	N	R	U							X
contentInfo:MD_CoverageDescription: dimension:MD_Band	maxValue	N	R	U			X		X		X
contentInfo:MD_CoverageDescription: dimension:MD_Band	minValue	N	R	U			X		X		X
contentInfo:MD_CoverageDescription: dimension:MD_Band	units	N	R	U			X		X		X
contentInfo:MD_CoverageDescription: dimension:MD_Band	bitsPerValue	N	R	U			X		X		X
contentInfo:MD_ContentInformation: MD_FeatureCatalogueDescription	complianceCode	C		U			X				X
contentInfo:MD_ContentInformation: MD_FeatureCatalogueDescription	language	C		U							X
contentInfo:MD_ContentInformation: MD_FeatureCatalogueDescription	includedWithDataset	C		U			X		X		X
contentInfo:MD_ContentInformation	featuresTypes	C		U			X		X		X
spatialRepresentationInfo: MD_GridSpatialRepresentation	numberOfDimensions	N	R	U			X		X		X
spatialRepresentationInfo: MD_GridSpatialRepresentation	cellGeometry	N	R	U			X		X		X
spatialRepresentationInfo: MD_GridSpatialRepresentation	transformationParameterAvailability	N	R	U			X		X		X
spatialRepresentationInfo: MD_GridSpatialRepresentation: axisDimensionProperties:MD_Dimension	dimensionName	N	R	U			X		X		X
axisDimensionProperties:MD_Dimension	dimensionSize	N	R	U			X		X		X
spatialRepresentationInfo: MD_GridSpatialRepresentation: axisDimensionProperties:MD_Dimension: resolution	value	N	R	U		X	X		X		X
axisDimensionProperties:MD_Dimension: resolution	units	N	R	U		X	X		X		X
spatialRepresentationInfo: MD_VectorSpatialRepresentation	topologyLevel	N	V	U			X		X		X
spatialRepresentationInfo:	geometricObjectType	N	V	U			X		X		X

MD_VectorSpatialRepresentation: geometricObjects: MD_GeometricObjects									
spatialRepresentationInfo: MD_VectorSpatialRepresentation: geometricObjects: MD_GeometricObjects	geometricObjectCount	N	V	U			X	X	X
identification:MD_Identificacion: MD_DataIdentification: citation: CI_Citation	title	C		L			X	X	X
identification:MD_Identificacion: MD_DataIdentification:citation: CI_Citation: date: CI_Date	date	C		L			X	X	X
identification:MD_Identificacion: MD_DataIdentification:citation: CI_Citation: date: CI_Date	dateType	C		L			X	X	X
identification:MD_Identificacion: MD_DataIdentification	credit	C		L					X
identification:MD_Identificacion: MD_DataIdentification: graphicOverview: MD_BrowseGraphic	filename	C		L	X				X
identification:MD_Identificacion: MD_DataIdentification: graphicOverview: MD_BrowseGraphic	fileDescription	C		L					X
identification:MD_Identificacion: MD_DataIdentification: graphicOverview: MD_BrowseGraphic	fileType	C		L			X		X
identification:MD_Identificacion: MD_DataIdentification: resourceFormat: MD_Format	name	C		L	X		X	X	X
identification:MD_Identificacion: MD_DataIdentification: resourceFormat: MD_Format	version	C		L	X			X	X
identification:MD_Identificacion: MD_DataIdentification: resourceFormat: MD_Format	fileDecompressionTechnique	C		L			X	X	X
identification:MD_Identificacion: MD_DataIdentification: resourceSpecificUsage: MD_Usage	userDeterminedLimitations	C		E			X	X	X
identification:MD_Identificacion: MD_DataIdentification: resourceConstraints: MD_LegalConstraints	useConstraints	C		E			X	X	X
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords	keyword	C		L			X	X	X
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords	type	C		L			X	X	X
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords: thesaurusName:CI_Citation	title	C		L			X	X	X
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords: thesaurusName:CI_Citation	date	C		L			X	X	X
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords: thesaurusName:CI_Citation	dateType	C		L			X	X	X
identification: MD_Identificacion	spatialRepresentationType	C		L			X	X	X
identification: MD_Identificacion: spatialResolution: MD_Resolution	distance	N	R	L			X	X	X
identification: MD_Identificacion: spatialResolution: MD_Resolution: equivalentScale: MD_RepresentativeFraction	denominator	N	R	L			X	X	X
identification: MD_Identificacion	lenguaje	C		L			X		X
identification: MD_Identificacion	characterSet	C		L	X		X		
identification: MD_Identificacion	topicCategory	C		L			X		X
identification: MD_Identificacion	environmentDescription	C		L	X				X
identification: MD_Identificacion	supplementalInformation	N	R	L					X
identification: MD_Identificacion: extent: EX_Extent	description	C		L					X
identification: MD_Identificacion: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicBoundingBox	extentTypeCode	C		L			X	X	X
identification: MD_Identificacion: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicBoundingBox	westBoundLongitude	C		L			X	X	X
identification: MD_Identificacion: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicBoundingBox	eastBoundLongitude	C		L			X	X	X
identification: MD_Identificacion: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicBoundingBox	southBoundLatitude	C		L			X	X	X
identification: MD_Identificacion: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicBoundingBox	northBoundLatitude	C		L			X	X	X
identification: MD_Identificacion: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicDescription: geographicIdentifier: RS_Identifier: authority: CI_Citation	title	C		L			X	X	X
identification: MD_Identificacion: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicDescription: geographicIdentifier: RS_Identifier: authority: CI_Citation : date : CI_Date	date	C		L			X	X	X
identification: MD_Identificacion: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicDescription: geographicIdentifier: RS_Identifier: authority: CI_Citation : date : CI_Date	dateType	C		L			X	X	X
identification: MD_Identificacion: extent: EX_Extent: verticalElement:Ex_VerticalExtent	minimumValue	N	D	L			X	X	
identification: MD_Identificacion: extent: EX_Extent: verticalElement:Ex_VerticalExtent	maximumValue	N	D	L			X	X	
dataQualityInfo: DQ_DataQuality: scope: DQ_Scope	level	C		E			X		X
dataQualityInfo: DQ_DataQuality: lineage: LI_Lineage: source : LI_Source	description	C		E					X
dataQualityInfo: DQ_DataQuality: lineage: LI_Lineage: processStep: LI_ProcessStep	description	C		E					X



Atendiendo a la función que desempeñan los ítems de metadatos, 34 desempeñan la función de localización (42%), 6 de evaluación (8%), 5 de acceso (6%) y 36 de uso (44%), se puede concluir que las funciones más beneficiadas por la automatización son la localización y el uso.

Al analizar la interoperabilidad que facilitan los ítems de metadatos creados con la metodología, además de clasificarlos mediante los niveles de interoperabilidad que facilitan, se han analizado los valores obtenidos y se ha comparado, tanto con la norma de metadatos ISO19115 en toda su extensión, como con los ítems pertenecientes al núcleo de la norma. La tabla 5.16 recoge las cantidades de ítems de la norma, del núcleo y los que se pueden obtener por cada tipo de IG, clasificados por niveles de interoperabilidad.

	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
ISO-19115	37	6	128	15	99	7	217
Core ISO-19115	8	3	40	3	31	1	43
Automático Ráster	11	3	56	6	46	5	67
Automático DTM	11	3	59	6	49	5	67
Automático Vector	11	2	45	6	35	5	54
Automático ISO Core	6	2	23-25	2	20-22	0	26

Tabla 5.16: Número de ítems metadatos que la favorecen los niveles de interoperabilidad

Como puede observarse en la tabla 5.16, el número de ítems que facilitan la interoperabilidad organizacional es muy elevado, como también lo es el número de ítems de la norma de metadatos definidos para describir los datos desde el punto de vista organizacional y de su ciclo de vida. Cabe destacar que el número de ítems que facilitan los distintos niveles de interoperabilidad es similar entre la categoría de datos ráster y la categoría de datos DTM y se diferencian de la categoría datos vectoriales, en que hay menos ítems que facilitan las interoperabilidades semántica, dinámica y organizacional. Se verifica que el número de ítems de metadatos creados automáticamente superan los umbrales definidos por los ítems que conforman el núcleo de la norma, a excepción de la interoperabilidad sintáctica. Su relevancia es escasa para los almacenes de IG, ya

que el 50% de los ítems que facilitan este tipo de interoperabilidad están vinculados con los servicios.

Con objeto de analizar la eficiencia del método de generación automática de metadatos y facilitar su interpretación, se han normalizado las cantidades de ítems creados, respecto a los valores del conjunto de la norma ISO-19115. Así, se facilita la comparación de las intensidades de interoperabilidad que proporcionan los metadatos creados de este modo, como puede verse en la figura 5.1.

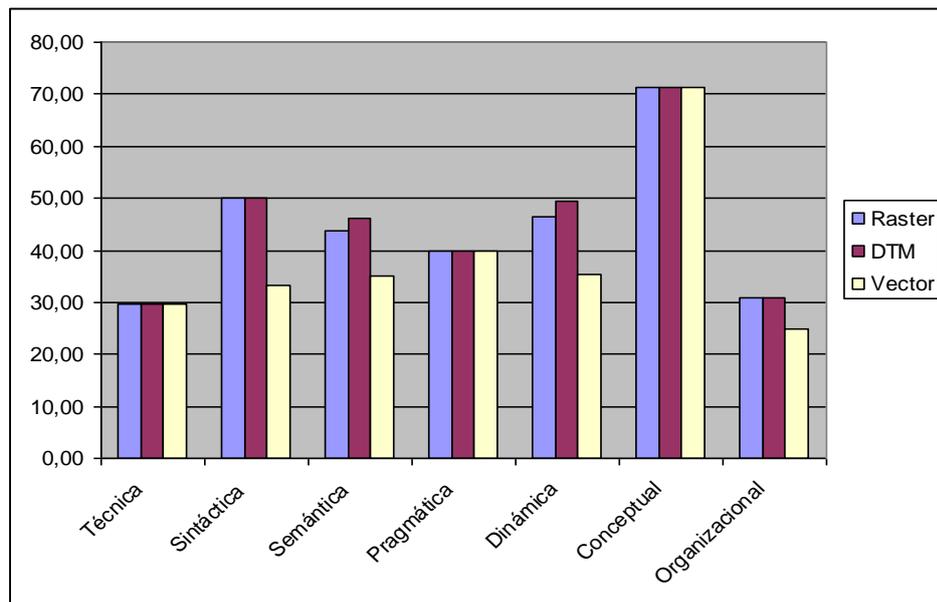


Figura 5.1: Porcentajes de ítems creados automáticamente por nivel de interoperabilidad

La figura 5.1 evidencia que:

- los porcentajes de ítems creados automáticamente por la metodología son suficientemente homogéneos para los distintos niveles de interoperabilidad;
- los valores medios de dichos porcentajes se encuentran comprendidos entre el 30 y el 40%, excepto la interoperabilidad organizacional para los datos vectoriales que es un poco inferior y la conceptual que es sensiblemente superior;
- se puede afirmar que el método de generación automática presta más atención a los niveles de interoperabilidad menos favorecidos por la norma de metadatos ISO 19115 (sintáctica y conceptual), aunque el número de ítems es muy escaso 3 y 5 ítems respectivamente.

## 5.5 Validación del modelo de interoperabilidad

Una vez definido el modelo de interoperabilidad aplicado a las IDE formado por siete niveles: técnico, sintáctico, semántico, pragmático, dinámico, conceptual y organizacional y, tras haber analizado la interoperabilidad que pueden proporcionar los ítems de la norma de metadatos ISO19115 (ver capítulo 4 epígrafe 4), en este capítulo se muestran las acciones realizadas para validar los resultados del análisis, a través de la consulta a un grupo expertos, procedentes de distintos dominios de conocimiento, concedores de la norma e involucrados directamente con las IDE. El objeto del trabajo de validación es cuantificar la subjetividad de la clasificación de los niveles de interoperabilidad proporcionados por los ítems de metadatos imputables al autor del modelo.

Para realizar esta validación se ha contado con 5 profesionales involucrados en las IDE y concedores de la norma de metadatos ISO19115, adscritos a dominios de conocimiento: Informática, Geodesia, Cartografía y Topografía, de 3 nacionalidades diferentes: argentina, colombiana y española. Se les ha solicitado que identifiquen y asignen los niveles de interoperabilidad que facilitan los ítems del núcleo fundamental de la norma, incluyendo o excluyendo niveles del modelo, a través de una hoja de cálculo. Además, se ha solicitado también que expongan narrativamente la justificación de su decisión.

La decisión de utilizar únicamente los ítems pertenecientes al núcleo fundamental de la norma para realizar la validación, se justifica en la cuarta conclusión del capítulo cuatro, en la que se indica que las relaciones entre los niveles de interoperabilidad son similares tanto al analizar los ítems del núcleo como al hacerlo con todos los ítems de la norma de metadatos ISO19115. De esta forma se simplifica el esfuerzo realizado por los colaboradores.

Una vez definida la extensión, el alcance y el objetivo de la validación, se describe la metodología usada, estructurando el resto del apartado de la siguiente forma: en primer lugar se describe cómo se ha diseñado el proceso de validación, la elección de los colaboradores y el modo de recoger sus interpretaciones; en segundo lugar se explica cómo se han procesado las encuestas; y, en tercer lugar, se analiza la validez de las relaciones entre los niveles del modelo y la

clasificación de los niveles atribuidos a los distintos ítems. Se han dejado los resultados y las conclusiones para el último apartado de este capítulo.

### 5.5.1 Definición de las encuestas

La encuesta utilizada para realizar la validación del modelo ha sido diseñada en una hoja de cálculo, de tal modo que la primera columna contiene los ítems del núcleo de la norma de metadatos ISO19115/19139. Las siguientes siete columnas identifican los niveles de interoperabilidad seleccionados en la definición del modelo. En una columna adicional se expresa el criterio que el autor ha utilizado para seleccionar los niveles al diseñar el modelo. Finalmente se han incluido dos columnas, la primera en la que el colaborador muestra su acuerdo o desacuerdo con los niveles del modelo y la última, en la que el colaborador ha de aportar sus comentarios y justificaciones. Para ilustrar gráficamente la tabla se presenta la figura 5.2, que contiene una captura de pantalla de la hoja de cálculo utilizada.

	A	B	C	D	E	F	G	H	I	J	K	L
	Metadata Attributes	Mandatory/conditional	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organisational	Results and Comments	¿Estas de acuerdo?	Comentarios: ¿Quitarías algún tipo de interoperabilidad? ¿Añadirías algún tipo no mandado?
1	fileIdentifier	X			X					Technical: identifier associated to the file and metadata file system. Intended to avoid maintenance and management problems in databases. Pragmatic: allows dynamic handling of metadata by identifiers: collection.		
2	Language			X					X	Semantic: term of a list identifying language of metadata. Organisational: enables language negotiation in business models.		
3	characterSet	X	X							Technical: allows correct handling of bytes from the metadata register. Semantic: identifies univocally encoding type on the base of a list of controlled		
4	parentIdentifier	X			X					Technical: identifies univocally the parent node in the hierarchical relation and makes access to metadata possible. Pragmatic: allows automatically browsing through metadata relations		
5												
6												
7												
8												
9												

Figura 5.2: Apariencia y estructura de la hoja de cálculo de la encuesta

Junto a la citada hoja de cálculo, se ha distribuido un documento en el que se definen los niveles de interoperabilidad del modelo.

El siguiente paso ha consistido en seleccionar a los colaboradores que han participado en la encuesta atendiendo a los siguientes requisitos: 1) conocer la norma de metadatos ISO19115, 2) estar adscritos a instituciones que gestionan Información Geográfica y 3) pertenecer a una diversidad geográfica y provenir de distintos dominios de conocimiento.

Las instituciones involucradas son:

- Instituto Geográfico Nacional de España (IGN-E),
- Instituto Geográfico Agustín Codazzi de Colombia (IGAC),

- Instituto de Desarrollo Regional de la Universidad de Castilla la Mancha (IDR-CM),
- Instituto Geográfico Militar de Argentina (IGM),
- Universidad Politécnica de Madrid (UPM).

Los dominios de conocimiento involucrados son:

- Ingeniería Geodésica y Cartográfica,
- Ingeniería Geográfica,
- Ingeniería de Agrimensura,
- Ingeniería Informática.

### 5.5.2 Procesamiento de las encuestas

Recibidos los documentos que recogen las respuestas de los encuestados, el siguiente tarea ha sido su procesamiento. Dado que un colaborador ha preferido responder en un documento de texto, el primer tratamiento de esta encuesta ha consistido en trasladar sus respuestas a la hoja de cálculo diseñada. La segunda etapa del tratamiento común de los datos ha consistido en agrupar todas las respuestas en un único documento, sobre el que realizar el resto de los tratamientos.

Una vez agrupados los datos, se ha reproducido a nivel individual el tratamiento dado en el proceso de definición del modelo de interoperabilidad al aplicarlo sobre los metadatos. El objetivo de este tratamiento individual de los resultados es analizar e interpretar los conteos de ítems que facilitan los niveles del modelo obtenidas para cada colaborador. De este modo, se han obtenido los valores que se presentan numérica y gráficamente, en la tabla 5.17 y figura 5.3 respectivamente. Para preservar el anonimato de los encuestados se han nombrado las aportaciones de los mismos como “encuestado n”.

	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
Modelo Validar	7	4	39	2	31	1	42
Encuestado 1	12	16	39	9	36	1	44
Encuestado 2	2	7	33	1	12	3	24
Encuestado 3	9	5	40	16	31	3	43
Encuestado 4	8	5	52	3	31	1	47
Encuestado 5	1	26	39	7	12	1	31

Tabla 5.17: Intensidades de los niveles de interoperabilidad por encuestado al analizar los ítems del núcleo fundamental de la norma ISO19115 de metadatos

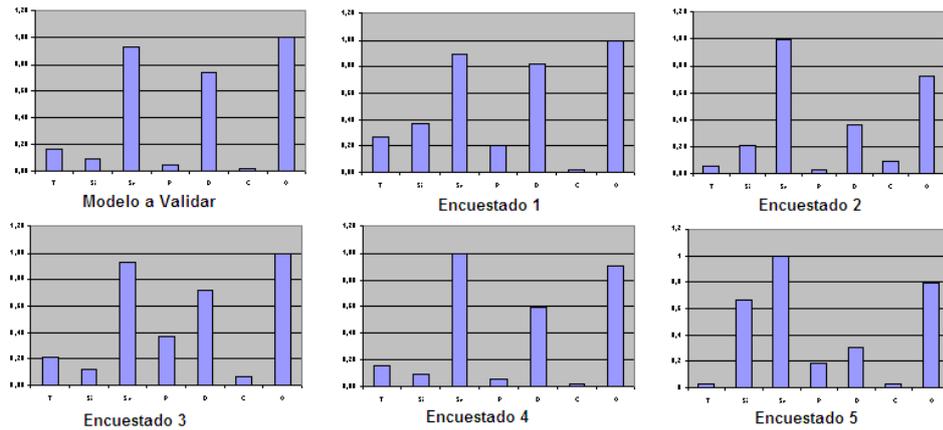


Figura 5.3: Histograma de las intensidades de interoperabilidad por encuestado

Los histogramas de la figura 5.3 permiten realizar las siguientes interpretaciones:

- Para los encuestados 1, 2, 3 y 4, se mantienen elevadas las intensidades de los niveles semántico, dinámico y organizacional, con valores bien diferenciados del resto de niveles.
- El encuestado 5 considera como aspectos sintácticos de la interoperabilidad muchos aspectos que el resto de encuestados consideran dinámicos.
- Los encuestados 2 y 5 identifican más ítems que promueven la interoperabilidad semántica que los que promueven la organizacional en discordancia con el resto.
- Todos los encuestados coinciden con el modelo propuesto, en el bajo número de ítems que proporcionan interoperabilidad conceptual.

Continuando con el método de análisis realizado en el modelo, se ha analizado la frecuencia con la que los ítems de metadatos promueven simultáneamente varios niveles de interoperabilidad. Las tablas de la figura 5.4 muestran las frecuencias, tanto para el modelo a validar, como para cada uno de los colaboradores.

Modelo validar	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
Técnica	8	0	4	2	3	0	3
Sintáctica	0	3	3	0	1	0	2
Semántica	4	3	40	0	28	1	29
Pragmática	2	0	0	3	2	0	1
Dinámica	3	1	28	2	31	0	23
Conceptual	0	0	1	0	0	1	1
Organizacional	3	2	29	1	23	1	43
Encuestado 1	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
Técnica	12	5	8	4	6	0	6
Sintáctica	5	16	14	3	11	0	11
Semántica	8	14	39	6	31	1	29
Pragmática	4	3	6	9	5	0	8
Dinámica	6	11	31	5	36	1	27
Conceptual	0	0	1	0	1	1	1
Organizacional	6	11	29	8	27	1	44
Encuestado 2	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
Técnica	2	0	0	0	1	0	0
Sintáctica	0	7	1	0	0	1	1
Semántica	0	1	33	0	10	1	14
Pragmática	0	0	0	1	0	0	0
Dinámica	1	0	10	0	12	0	3
Conceptual	0	1	1	0	0	3	2
Organizacional	0	1	14	0	3	2	24
Encuestado 3	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
Técnica	9	0	5	3	3	0	4
Sintáctica	0	5	5	2	2	0	3
Semántica	5	5	40	12	28	1	28
Pragmática	3	2	12	16	14	0	12
Dinámica	3	2	28	14	31	0	23
Conceptual	0	0	1	0	0	3	3
Organizacional	4	3	28	12	23	3	43
Encuestado 4	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
Técnica	8	1	7	2	3	0	7
Sintáctica	1	5	5	0	3	0	3
Semántica	7	5	52	2	31	1	22
Pragmática	2	0	2	3	1	0	3
Dinámica	3	3	31	1	31	0	23
Conceptual	0	0	1	0	0	1	1
Organizacional	7	3	22	3	23	1	47
Encuestado 5	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
Técnica	1	1	1	0	0	0	0
Sintáctica	1	26	26	0	4	0	9
Semántica	1	26	39	4	9	1	18
Pragmática	0	0	4	7	1	1	6
Dinámica	0	4	9	1	12	1	9
Conceptual	0	0	1	1	1	1	1
Organizacional	0	9	18	6	9	1	31

Figura 5.4: Tablas de las relaciones entre los niveles del modelo de interoperabilidad según los ítems del núcleo para cada encuestado

Las tablas de la figura 5.4 contienen el conteo de ítems que proporcionando el nivel de interoperabilidad que indica la columna, proporcionan simultáneamente los niveles identificados en las filas. Así por ejemplo para el encuestado 5, de los 26 ítems que proporcionan interoperabilidad sintáctica, 9 proporcionan organizacional, 4 dinámica, 1 técnica y todos semántica.

La figura 5.5 muestra de forma gráfica las intensidades de las relaciones entre los niveles del modelo para cada colaborador, resultantes de interpretar la interoperabilidad de los ítems del núcleo fundamental de la norma de metadatos. Debido a las grandes diferencias en las cantidades de ítems en las relaciones de los niveles del modelo, se ha adoptado el criterio de utilizar los valores máximo de cada combinación para definir las isolíneas y por tanto las figuras no son comparables entre si.

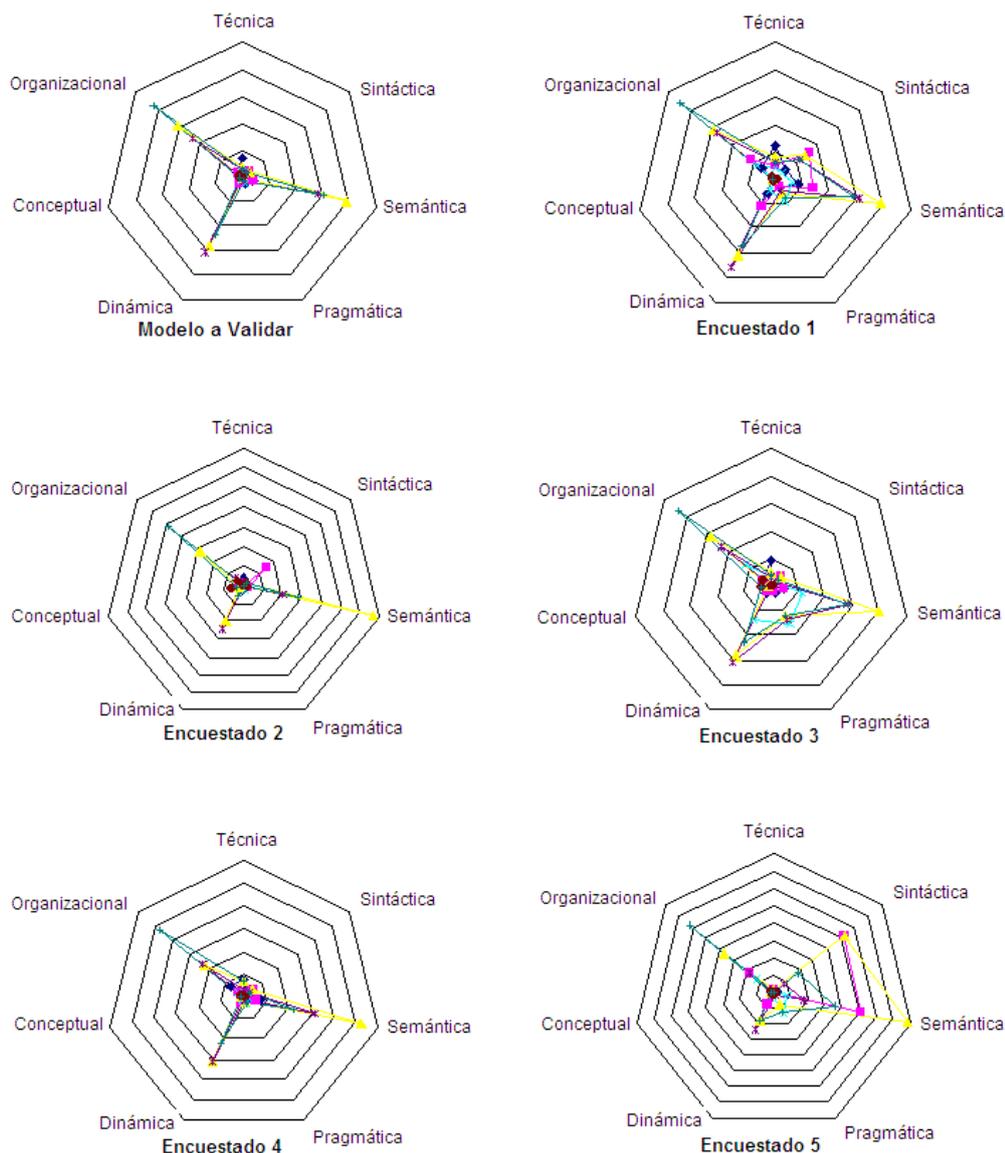


Figura 5.5: Diagramas radiales con las relaciones entre los niveles del modelo de interoperabilidad

Los gráficos de la figura 5.5 permiten realizar las siguientes interpretaciones:

- La distribución de las intensidades en las relaciones de los encuestados 1, 2 y 4 es similar a las observadas en el modelo a validar.
- Se observa mayor dispersión en las relaciones del encuestado 1. Esto se debe a que dicho colaborador ha considerado que los ítems de metadatos aportan simultáneamente varios niveles de interoperabilidad.
- Se observan menos relaciones y con menor intensidad, en el gráfico del encuestado 2. Al contrario del encuestado 1, éste ha considerado que los ítems no proporcionan interoperabilidad en determinados niveles del modelo.
- Para el encuestado 3 se observan mayores relaciones entre los niveles pragmático, dinámico y semántico. Se debe a que ha interpretado que un número significativo de ítems aportan simultáneamente interoperabilidades pragmática, dinámica y semántica.
- Se aprecian mayores intensidades entre los niveles semántico y sintáctico para el encuestado 5. Se debe a que ha interpretado que los ítems no proporcionan interoperabilidad en los niveles técnico ni dinámico; sin embargo, sugiere que muchos ítems proporcionan interoperabilidad en los niveles pragmático y sintáctico.

Una vez rehecho el análisis realizado con el modelo con las interpretaciones de los colaboradores, se procede a identificar las discrepancias con el modelo propuesto. Se pretende identificar:

- las permutaciones del nivel de interoperabilidad propuestos por los colaboradores para un ítem,
- las discrepancias (por exceso o por defecto) para cada ítem, relacionándolo con el modelo a validar.

La figura 5.6 muestra gráficamente el trabajo realizado con las encuestas. En las dos primeras columnas se muestran los nombres de los paquetes e ítems analizados. Los siguientes cinco bloques muestran las discrepancias encontradas con el modelo propuesto, de modo que los ítems que proporcionan un nivel más de interoperabilidad que el modelo, se resaltan con color verde y, en color naranja, los ítems que no proporcionan ningún nivel de interoperabilidad en el modelo.

	Metadata items	Encuestado 1					Encuestado 2					Encuestado 3					Encuestado 4					Encuestado 5						
		Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual
Md_Metadata	fileIdentifier																											
	Language																											
	characterSet																											
	parentIdentifier																											
	hierarchyLevel																											
	hierarchyLevelName																											
	dateStamp																											
	metadataStandardName																											
	metadataStandardVersion																											
PT_Locale	language																											
	characterEncoding																											
CI_ResponsibleParty	organizationName																											
	positionName																											
	role																											
RS_Identifier	code																											
MD_Identification	abstract																											
MD_DataIdentification	spatialRepresentationType																											
	language																											
	characterSet																											
	topicCategory																											
	distance																											
MD_RepresentativeFraction	denominator																											
CI_ResponsibleParty	organizationName																											
	positionName																											
CI_Citation	title																											
CI_Date	date																											
	dateType																											
Ex_Extent	description																											
Ex_BoundingPolygon	polygon																											
Ex_GeographicBoundingBox	westBoundLongitude																											
	eastBoundLongitude																											
	southBoundLatitude																											
	northBoundLatitude																											
MD_Identifier	code																											
Ex_TemporalExtent	extent																											
Ex_VerticalExtent	minimumValue																											
	maximumValue																											
	unitOfMeasure																											
MD_Format	name																											
	version																											
CI_OnlineResource	linkage																											
DQ_DataQuality_LI_Lineage	statement																											
LI_ProcessStep	description																											
LI_Source	description																											
Ex_Extent	description																											
Ex_GeographicExtent	polygon																											
	westBoundLongitude																											
	eastBoundLongitude																											
	southBoundLatitude																											
	northBoundLatitude																											
Ex_TemporalExtent	extent																											
Ex_VerticalExtent	minimumValue																											
	maximumValue																											
	unitOfMeasure																											
MD_Identifier	code																											

Figura 5.6: Gráfico con las concordancias y discrepancias entre encuestados y modelo a validar

Los gráficos de la figura 5.6 pueden interpretarse del siguiente modo:

- Los colaboradores 2 y 5 interpretan que la interoperabilidad proporcionada por los ítems de metadatos responden de un modo diferente a cómo se ha interpretado al definir el modelo.
- Los colaboradores 1, 3 y 4 indican que los ítems de metadatos proporcionan algún nivel de interoperabilidad adicional al propuesto en el modelo.
- Para el colaborador 2, los metadatos no proporcionan la interoperabilidad en los niveles dinámico ni organizacional.
- El colaborador 5 discrepa con el resto de colaboradores y con el modelo a validar al considerar que un conjunto de ítems proporcionan la interoperabilidad sintáctica mientras que para el resto proporcionan el nivel dinámico.
- El colaborador 5 interpreta que no hay tantos ítems que proporciona el nivel organizacional de interoperabilidad.

- Para el encuestado 2 no resulta fácil identificar cambios desde un nivel a otro en un ítem. En algunos casos, se propone no considerar más que un nivel en cada ítem; en otros, añadir un nivel y, tan sólo en 4 de ellos, sí se percibe que algunos aspectos de nivel organizacional en el modelo son considerados aspectos dinámicos (*language* y *topicCategory*).

El siguiente paso ha consistido en unir las propuestas de los colaboradores para, posteriormente filtrar los resultados, de modo que se descarten las discrepancias puntuales de un solo encuestado. El objetivo de éste análisis es obtener el modelo agregado y compararlo con el propuesto.

La figura 5.7 muestra los resultados de este análisis. Esta figura comparte con la anterior (figura 5.6) las dos primeras columnas. La tercera columna muestra el resultado de la agregación de las 5 encuestas y el modelo que se está validando. Mediante una leyenda de colores se indica el acuerdo entre los 5 colaboradores y, mediante marcas de texto (x), se muestran los niveles del modelo que se está validando. La cuarta columna repite el razonamiento, agregando las 3 encuestas más discrepantes o críticas que corresponden a los colaboradores 1, 2 y 5. La última muestra la agregación de los cinco colaboradores.

La leyenda de colores utilizada para presentar los resultados es la siguiente:

- Si todos los encuestados opinan igual, se colorea el fondo de la celda con color verde oscuro.
- Cuando sólo discrepa un colaborador, el fondo de la celda se rellena con color verde medio.
- Si discrepan dos, el color seleccionado es el verde más claro.
- Si discrepan 3 encuestados respecto al modelo propuesto, el fondo de la celda se rellena de color rojo.
- Las marcas de texto (x) de la tercera columna identifican los niveles de interoperabilidad definidos en el modelo a validar.
- Las marcas de texto (A) de la cuarta columna identifican los ítems y niveles de interoperabilidad en el que los colaboradores más críticos coinciden, discrepando respecto al modelo.

- La casilla roja marcada con el símbolo “!” es el ítem de metadatos en el que dos de los tres encuestados más críticos discrepan del modelo. Se analiza con más detalle posteriormente.

	Metadata items	Resultados agregados					Encuestados 1,2,5					Encuestados 1,2,3,4,5											
		Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional	
Md_Metadata	fileIdentifier	X		X																			
	Language			X			X																
	characterSet	X		X					A														
	parentIdentifier	X			X																		A
	hierarchyLevel			X			X	X															
	hierarchyLevelName							X															
	dateStamp				X		X	X															
	metadataStandardName			X	X			X															
	metadataStandardVersion			X	X			X															
	PT_Locale	language			X			X															
characterEncoding		X		X					A														
CI_ResponsibleParty	organizationName						X																
	positionName						X																
	role			X			X																
RS_Identifier	code			X		X	X																
MD_Identifier	abstract						X																
MD_DataIdentification	spatialRepresentationType			X		X	X																
	language			X			X																
	characterSet	X		X					A														
	topicCategory			X			X																
	distance			X	X		X																
MD_RepresentativeFraction	denominator			X	X		X																
CI_ResponsibleParty	organizationName						X																
	positionName						X																
	role			X			X																
CI_Citation	title			X	X		X																
CI_Date	date			X	X		X																
	dateType			X	X		X																
Ex_Extent	description						X																
Ex_BoundingPolygon	polygon		X	X		X	X																
Ex_GeographicBoundingBox	westBoundLongitude		X	X		X	X																
	eastBoundLongitude		X	X		X	X																
	southBoundLatitude		X	X		X	X																
	northBoundLatitude		X	X		X	X																
MD_Identifier	code			X		X	X																
Ex_TemporalExtent	extent			X	X		X																
	minimumValue			X	X		X																
Ex_VerticalExtent	maximumValue			X	X		X																
	unitOfMeasure			X	X		X																
	name	X		X			X																
MD_Format	version	X		X			X																
	linkage	X		X		X	X																
CI_OnlineResource	statement						X																
DQ_DataQuality_LI_Lineage	description						X																
LI_ProcessStep	description						X																
LI_Source	description						X																
Ex_Extent	description						X																
Ex_GeographicExtent	polygon		X	X		X	X																
	westBoundLongitude		X	X		X	X																
	eastBoundLongitude		X	X		X	X																
	southBoundLatitude		X	X		X	X																
	northBoundLatitude		X	X		X	X																
Ex_TemporalExtent	extent			X	X		X																
Ex_VerticalExtent	minimumValue			X	X		X																
	maximumValue			X	X		X																
	unitOfMeasure			X	X		X																
MD_Identifier	code			X		X	X																

Figura 5.7: Grado de acuerdo en los niveles de interoperabilidad por ítems y grupos de encuestados

Los gráficos de la figura 5.7 permiten realizar las siguientes interpretaciones:

- Se observa una correlación gráfica estrecha entre los resultados de las tres agregaciones. Esto significa que las tres agrupaciones presentan resultados similares.

- Atendiendo a la columna central, que recoge las interpretaciones de los tres colaboradores más críticos, se percibe que debería considerarse que algunos ítems de metadatos proporcionan algunos niveles de interoperabilidad más que los propuestos en el modelo. Por ejemplo: las latitudes y longitudes máximas y mínimas o los juegos de caracteres proporcionan interoperabilidad a nivel sintáctico, los ítems *topicCategory* y *herachicalLevel* proporcionarían interoperabilidad dinámica y el ítem *parentIdentifier* proporcionaría el nivel organizacional. Para facilitar la identificación de éstos ítems se les ha etiquetado con el texto “A”.
- También se han identificado ítems que, para los encuestados más críticos, no aportan interoperabilidad dinámica. Estos son los elementos del paquete *Ex\_Extent* (longitudes y latitudes máximas y mínimas, la extensión temporal y vertical), y se han identificado con la etiqueta “?”.

Identificadas las discrepancias se ha procedido a analizarlas. Como resultado del análisis se han adoptado decisiones que conducen a considerar las propuestas o discutirlos y rechazarlos. A continuación se muestran las decisiones adoptadas y su justificación.

- El encuestado 5 considera que los ítems que contienen las longitudes y latitudes máximas y mínimas proporcionan interoperabilidad sintáctica. En el modelo propuesto, la interoperabilidad sintáctica está encaminada a definir cómo se codifican los datos para facilitar su uso, mientras que en estos ítems se almacenan valores numéricos que definen el contexto geográfico. Por esta razón se discrepa con esta opinión.
- Efectivamente los juegos de caracteres utilizados para crear los metadatos o el usado en los datos facilitan la interoperabilidad sintáctica, ya que definen la sintaxis de los datos y posibilitan la interpretación de la información.
- Se considera que los ítems que contienen la extensión geográfica tanto horizontal como vertical y la temporal, facilitan el nivel dinámico, ya que el conocimiento de estos ítems permiten localizar y seleccionar un conjunto de datos o servicio que satisfaga ciertas necesidades. Por tanto, permiten sustituir la fuente de datos, si fuera preciso, por cambios de disponibilidad en los servicios.

- Se concuerda con los encuestados más críticos en que los ítems *topicCategory* y *herachicalLevel*, además de los niveles propuestos en el modelo, proporcionan interoperabilidad a nivel dinámico, al considerarse que su conocimiento permite seleccionar otro conjuntos de datos o servicio que satisfaga la necesidad derivada de eventuales cambios de disponibilidad.
- También se está de acuerdo con los colaboradores más críticos en relación al ítem *parentIdentifier*, al considerar que puede proporcionar interoperabilidad a nivel organizacional. Esto se debe a que los metadatos de los elementos de una serie (p.e. cartográfica) pueden apuntar al metadato padre que describe la serie y, por tanto, permite acceder a otro conjunto de información que describe la misma.

Finalmente, tras interpretar y discutir las discrepancias encontradas entre el modelo propuesto y las aportaciones de los colaboradores, razonando la inclusión o el descarte de niveles de interoperabilidad asignados a los ítems, se muestra el modelo final que incluye los cambios. En la figura 5.8 se exponen simultáneamente los dos modelos: el modelo a validar y el modelo final, etiquetando con el texto “X” los niveles de interoperabilidad que aportan los ítems. Sobre la columna de la derecha, que contiene el modelo final, se han resaltado, rellenando las celdas con un color crema, los niveles del modelo con la incorporación de algunos ítems de metadatos en el proceso de validación.

	Metadata ítems	Modelo a Validar						Modelo Final							
		Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
Md_Metadata	fileIdentifier	X			X				X						
	Language			X			X			X					X
	characterSet	X	X					X	X	X					
	parentIdentifier	X		X							X				X
	hierarchyLevel		X			X	X			X		X	X	X	X
	hierarchyLevelName						X								X
	dateStamp			X	X		X			X	X				X
	metadataStandardName		X	X			X		X	X					X
	metadataStandardVersion		X	X			X		X	X					X
PT_Locale	language			X			X			X					X
	characterEncoding	X	X					X	X	X					
CI_ResponsibleParty	organizationName						X								X
	positionName						X								X
	role			X			X			X					X
RS_Identifier	code			X	X		X			X	X				X
MD_Identification	abstract						X								X
MD_DataIdentification	spatialRepresentationType			X	X		X			X	X				X
	language			X			X			X					X
	characterSet	X	X					X	X	X					
	topicCategory			X			X			X		X			X
	distance		X	X			X			X	X				X
MD_RepresentativeFraction	denominator		X	X			X			X	X				X
CI_ResponsibleParty	organizationName						X								X
	positionName						X								X
	role		X				X			X					X
CI_Citation	title			X	X		X			X	X				X
CI_Date	date		X	X			X			X	X				X
	dateType		X	X			X			X	X				X
Ex_Extent	description						X								X
Ex_BoundingPolygon	polygon		X	X	X				X	X	X				X
Ex_GeographicBoundingBox	westBoundLongitude		X	X	X		X			X	X				X
	eastBoundLongitude		X	X	X		X			X	X				X
	southBoundLatitude		X	X	X		X			X	X				X
	northBoundLatitude		X	X	X		X			X	X				X
MD_Identifier	code		X	X	X		X			X	X				X
Ex_TemporalExtent	extent		X	X	X		X			X	X				X
Ex_VerticalExtent	minimumValue		X	X						X	X				
	maximumValue		X	X						X	X				
	unitOfMeasure		X	X						X	X				
MD_Format	name	X	X	X	X		X		X	X	X				X
	version	X			X	X	X		X		X				X
CI_OnlineResource	linkage	X			X	X	X		X		X				X
DQ_DataQuality.LI_Lineage	statement						X								X
LI_ProcessStep	description						X								X
LI_Source	description						X								X
Ex_Extent	description						X								X
Ex_GeographicExtent	polygon		X	X	X				X	X	X				
	westBoundLongitude		X	X	X		X			X	X				X
	eastBoundLongitude		X	X	X		X			X	X				X
	southBoundLatitude		X	X	X		X			X	X				X
	northBoundLatitude		X	X	X		X			X	X				X
Ex_TemporalExtent	extent		X	X	X		X			X	X				X
Ex_VerticalExtent	minimumValue		X	X						X	X				
	maximumValue		X	X						X	X				
	unitOfMeasure		X	X						X	X				
MD_Identifier	code		X	X	X		X			X	X				X

Figura 5.8: Tabla comparativa con los niveles de interoperabilidad de los ítems del núcleo de la norma de metadatos antes y después de la validación

Una vez incluidas las modificaciones sobre el modelo, se vuelve a analizar cómo afectan los cambios en las intensidades de las relaciones entre niveles.

Para hacer éste análisis se ha vuelto a calcular el número de ítems que aportan otros niveles de interoperabilidad para cada uno de los niveles del modelo, obteniéndose los datos mostrados en la tabla 5.18, en la que se presentan el

número de ítems que proporcionan interoperabilidad en cada nivel, y el número que proporcionar simultáneamente dos de ellos.

Modelo validado	Técnica	Sintáctica	Semántica	Pragmática	Dinámica	Conceptual	Organizacional
Técnica	7	3	4	1	3	0	3
Sintáctica	3	7	7	0	2	0	2
Semántica	4	7	40	0	30	1	28
Pragmática	1	0	0	3	1	0	2
Dinámica	3	2	30	1	33	1	24
Conceptual	0	0	1	0	1	1	1
Organizacional	3	2	28	2	24	1	43

Tabla 5. 18: Relaciones entre los niveles del modelo de interoperabilidad una vez validado

La figura 5.9 muestra las intensidades de cada uno de los niveles del modelo, atendiendo a los valores mostrados en la tabla 5.16.

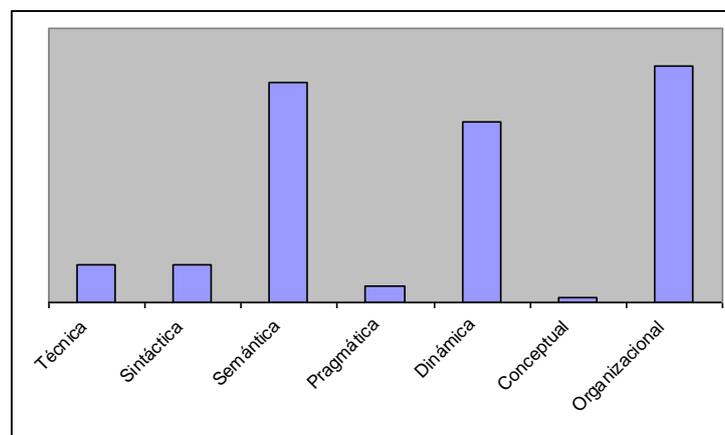


Figura 5.9: Histograma de los niveles de interoperabilidad en los ítems del núcleo de la norma de metadatos una vez validado

La figura 5.10 muestra gráficamente las intensidades de las relaciones entre niveles del modelo, una vez validado.

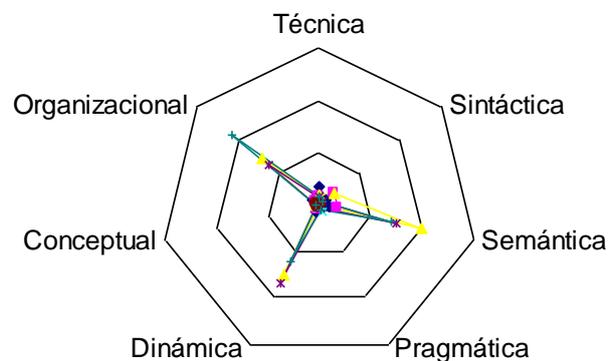


Figura 5.10: Intensidades de las relaciones entre los niveles del modelo una vez validado

Se ha calculado la correlación matemática existente entre las cantidades de ítems de metadatos que proporcionan interoperabilidad en los distintos niveles del modelo de interoperabilidad, comparando los valores iniciales y los finales del estudio, tras haber discutido y reajustado el modelo durante la validación. Para realizar este cálculo se han utilizado los valores de las tablas 5.17 y 5.18, respectivamente. El resultado del mismo es 0,9940, lo que indica el altísimo grado de acuerdo entre ambas tablas, que representan el modelo a validar y el validado. También se ha calculado dicha correlación entre el conteo de ítems de la norma completa de metadatos del capítulo 4 (tabla 4.7) y el modelo validado (tabla 5.18), obteniéndose como resultado 0,9235; este valor indica que existen un alto grado de acuerdo entre ambos estudios.

Teniendo en cuenta las interpretaciones, afirmaciones y consideraciones que se han ido realizando a lo largo del proceso de validación, los resultados de este proceso pueden resumirse de la siguiente forma:

- Dos de los cinco encuestados coinciden al identificar los niveles de interoperabilidad que proporcionan los metadatos; los otros tres colaboradores discrepan y enriquecen el modelo con sus criterios.
- Se ha realizado la agregación de los resultados de las encuestas siguiendo distintos criterios: (1) fusionando los resultados de los cinco encuestados más el propio modelo, (2) considerando los resultados de los cinco o (3) solo considerando los resultados de los tres más críticos.
- Se ha comprobado que los resultados son similares y se ha dado más relevancia a la suma de los resultados de los tres colaboradores más críticos.
- Se han analizado las discrepancias cuando coincidían todos o dos de ellos, y como consecuencia, se han estimado algunos cambios en los niveles de interoperabilidad que proporcionan algunos ítems del núcleo, y en otros casos se ha desestimado de forma justificada.
- Se ha vuelto a analizar el modelo en su conjunto y las intensidades de las relaciones entre niveles de interoperabilidad propiciadas por los ítems del núcleo de la norma, y se ha constatado que son básicamente las mismas que el modelo que se está validando.

- Se puede citar, más que destacar, que al incluir los ítems que proporcionan más niveles de interoperabilidad, el nivel que se ha visto ligeramente beneficiado es el sintáctico.
- Sobre todo lo anterior, destacar que los ítems de metadatos pueden usarse para dar soporte a los niveles de interoperabilidad.

## **5.6 Conclusiones**

Los estudios previos realizados en torno a los metadatos almacenados en la propia información, ya sea de forma implícita o explícita, y las formas de identificar el sistema de referencia espacial usado por las coordenadas manejadas en los mismos, permiten extraer las siguientes conclusiones:

- En primer lugar, destacar la gran variedad de formatos utilizados por la industria y por la comunidad científica para almacenar conjuntos de datos geográficos, o relacionados con el espacio. Se han identificado, analizado y descrito 71 formatos distintos agrupados en 7 categorías por contexto, tanto de uso, como de tipos de datos. Por término medio, para cada uno de estos formatos, se puede obtener 7-8 ítems de metadatos, muchos simples, aunque algunos complejos como las coordenadas máximas y mínimas, los estadísticos de las bandas, el conteo de filas y columnas o el conteo de geometrías de cada tipo. Pueden destacarse algunos tipos de formatos por su capacidad, desconocida en muchos contextos, disponible para almacenar metadatos, como los formatos jpeg y gtiff. En otros casos, el formato permite insertar en la cabecera del archivo un bloque de información, que bien puede ser un metadato completo, como ocurre con el formato JPEG2000. Las bases de datos con extensiones espaciales también sorprenden en algunos casos por la capacidad para almacenar los catálogos de objetos o incluir metadatos en las propias bases de datos geográficas, como ocurre en el caso de Esri.
- La segunda conclusión está relacionada con el estudio, en detalle, de la dificultad de interpretación del sistema de referencia espacial utilizado por las coordenadas en los distintos formatos de almacenamiento de los datos. Se han identificado y presentado ejemplos de la heterogeneidad de representaciones utilizadas: texto no estructurado, texto semi-estructurado, texto estructurado,

numéricas compartidas o consensuadas, numéricas particulares y, finalmente, nemotécnicos. Otro aspecto diferenciador detectado entre la heterogeneidad y los formatos analizados, es que en unos casos esta información se almacena como un solo atributo o metadato, y en otros se almacenan un conjunto de ítems, que identifican individualizadamente el elipsoide y el origen que define el datum, la proyección y el conjunto de parámetros que la definen. Este conjunto de hechos hace destacar que el tipo de usuarios al que está orientado este conocimiento es el experto en Geodesia y Cartografía. Afortunadamente, la estandarización y normalización en este campo (por ejemplo codificaciones EPSG y WKT) están permitiendo identificar, unívocamente y del modo más automático posible, los sistemas de referencia espaciales. La interpretación automática de los distintos tipos de codificaciones usadas por los formatos de almacenamiento, es una tarea extensa en trabajo y difícil de alcanzar, y una solución posible consistiría en identificar, para cada tipo de codificación, su equivalente en una codificación estandarizada como la propuesta por el EPSG.

Relacionado con los ítems de metadatos que pueden crearse automáticamente, hemos de destacar que la cantidad depende del tipo de información geográfica que se analice y del formato de almacenamiento utilizado, como se ha mostrado en el apartado de estudios previos. A continuación, se presentan algunos valores de carácter orientativo:

- Se puede crear un metadato asociado a datos ráster o imágenes de forma automática compuesto por 83 ítems, que podrán incrementarse si el número de bandas del formato es superior a una imagen perteneciente al espectro visible (RGB) con 3 bandas.
- Se puede crear un metadato asociado a datos vectoriales de forma automática compuesto por 69 ítems, que podrán incrementarse si el formato es capaz de almacenar más de un tipo de geometrías.
- Se puede crear un metadato asociado a modelos digitales de elevación de forma automática, compuesto por 68 ítems.

También se ha realizado el estudio de las funciones y los niveles del modelo de interoperabilidad aplicado a las IDE, favorecidos por los ítems de metadatos creados automáticamente, cabe destacar las siguientes conclusiones:

- Se han identificado explícitamente los ítems de metadatos que pueden crearse, bien por extracción, cálculo o inferencia (C), los que pueden tener cardinalidad mayor de uno (N) y los que pueden establecerse por parametrización del método u obtenerse del contexto (F), para las distintas categorías de datos geográficos: ráster vectoriales y modelos digitales del terreno. También se ha identificado las funciones que desempeña cada uno de los anteriores ítems, y los niveles de interoperabilidad que proporcionan.
- Se ha realizado el conteo de ítems que proporcionan los distintos niveles del modelo de interoperabilidad para cada tipo de dato y para los ítems del núcleo de la norma, mostrando los resultados en la tabla 5.16. Los porcentajes de dichos valores, frente a las cantidades máximas que pueden obtenerse cumplimentando todos los ítems de la norma, se han representado en la figura 5.1, y se puede concluir que facilitan de un modo homogéneo todos los niveles del modelo. También se puede observar que el número de ítems del núcleo fundamental que se pueden crear automáticamente es alto encontrándose en torno al 60%. Los valores que más destacan, relacionados con los niveles conceptual y sintáctico, tienen su justificación en el escaso número de ítems en la norma que lo facilitan.
- El análisis de las funciones que desempeñan los ítems de metadatos que se pueden crear automáticamente, indica que las funciones más favorecidas son la *localización* y *el uso* con 34 y 36% de los ítems, seguidos por la *evaluación* y el *acceso* con el 8 y el 6% de los mismos, respectivamente. Estos valores permiten afirmar que la creación automática de metadatos favorece las dos funciones extremas y básicas de los mismos, primero localizar, para finalmente usar, mientras que las funciones intermedias de evaluar y acceder a los datos no están siendo favorecidas en la misma medida.

Finalmente en este capítulo se presenta el trabajo de validación realizado en torno a la identificación de los distintos niveles de interoperabilidad, descrito en el capítulo 4, que proporcionan los ítems del núcleo fundamental de esta norma. Además de describirse la metodología seguida para realizar la validación, se han interpretando los resultados obtenidos con cada análisis realizado. A continuación, se sintetizan los resultados y las conclusiones alcanzadas con la validación:

- Cuatro de los cinco colaboradores en la validación han realizado críticas constructivas, identificando o cambiando los niveles de interoperabilidad proporcionados por los ítems de metadatos. Dos de los cinco han sido más críticos, y por esta razón, se ha realizado un análisis diferenciado con sus respuestas.
- Se ha comprobado que la interpretación de la interoperabilidad que pueden aportar los ítems de metadatos no es subjetiva, como cabía pensar. Las cantidades de ítems que proporcionan los distintos niveles del modelo son similares entre sí, y con el modelo objeto de la validación. Se ha comprobado que el análisis agregado de las respuestas de varios encuestados, atendiendo a las coincidencias y discrepancias en la interpretación de la interoperabilidad, arroja valores similares a los individuales, otra razón más que avala la validez del trabajo presentado en el capítulo 4.
- Se han propuesto algunos cambios en la interpretación de la interoperabilidad como consecuencia de este trabajo de validación, que pueden observarse en la figura 5.8. El resultado de analizar la correlación matemática sobre el conteo de los ítems de metadatos pertenecientes al núcleo fundamental de la norma ISO19115 que facilitan los niveles de interoperabilidad, ya sean nivel a nivel o parejas de niveles (tabla 5.18), con los valores antes de la validación, indica que existe un grado de asociación o parecido del 0.9940, hecho que valida el modelo previo.
- Se ha calculado la correlación existente entre los resultados del análisis de las interoperabilidades facilitadas por todos los ítems de la norma, frente al núcleo fundamental ahora validado, y se ha obtenido un resultado de asociación o parecido del 0,9235, validándose todo el trabajo realizado.
- Como consecuencia de la validación del análisis de la interoperabilidad que facilitan los ítems de metadatos, se pone de manifiesto que el uso de los mismos como mecanismo para permitir la interoperabilidad en las IDE, tiene un conjunto de fortalezas y otro de debilidades. Las fortalezas están del lado de los niveles de interoperabilidad organizacional, semántico y dinámico, mientras que las debilidades lo están de los de la interoperabilidad sintáctica, pragmática y conceptual. Esta conclusión está en la línea de las otras dos descritas en el capítulo 4, en las que se indica que la interoperabilidad sintáctica está garantizada por la propia norma y que las interoperabilidades

técnica y pragmática se han de garantizar por medio de otras normas, que estandaricen las tecnologías y las interfaces de los servicios.

## 6 CONCLUSIONES

En este último capítulo de la tesis se exponen las contribuciones científicas que este trabajo aporta; se revisan los principales objetivos de investigación y las cuestiones que se plantearon en el primer capítulo, y se muestra cómo se ha dado respuesta a las mismas. Al final, se plantean nuevas cuestiones, que quedan pendientes como futuros trabajos de investigación.

En primer lugar se resaltan las contribuciones científicas que este trabajo aporta:

- **El diseño de un nuevo modelo de interoperabilidad entre sistemas aplicable a las IDE:** Se ha diseñado un nuevo modelo de interoperabilidad fundamentado en la investigación de los modelos existentes en el contexto de los sistemas de sistemas, sobre los que se han incorporado los aspectos de organización propios de éste y se ha descrito la semántica de los niveles.
- **El diseño de una nueva metodología para la creación automática de metadatos relativos a la IG:** Se ha diseñado un nuevo método para automatizar la creación de metadatos aplicable a la IG, fundamentado en la investigación de las metodologías existentes, y de una nueva que surge del análisis de los metadatos implícitos y explícitos contenidos en los formatos de almacenamiento de la información geográfica. La nueva metodología extrae, calcula e infiere metadatos para, finalmente, estructurarlos y almacenarlos de un modo que facilite su intercambio, explotación e interoperabilidad.
- **El estudio de los metadatos de la IG desde el punto de vista de la interoperabilidad:** Se han analizado los ítems definidos por la norma internacional de metadatos ISO19115 desde un punto de vista novedoso: los niveles de interoperabilidad que proporcionan los metadatos. El análisis se ha aplicado en dos niveles de detalle: los ítems del núcleo de metadatos, para los datos, definidos por la norma y la totalidad del conjunto de ítems que la conforman, alcanzándose en ambos casos similares resultados. Como resultado, se aporta una nueva metodología de análisis de la interoperabilidad propiciada por los metadatos que es aplicable a nuevos perfiles o normas.

- **La aplicación y validación de los análisis de la interoperabilidad aportados por los ítems de la norma de metadatos:** Bajo la supervisión del autor, se ha realizado una encuesta orientada a validar la interpretación de los niveles de interoperabilidad aportados por los ítems del núcleo de la norma de metadatos. Los resultados de la encuesta validan la tipificación de la interoperabilidad facilitada por los ítems de la norma.
- **La aplicación de la metodología de creación automática de metadatos para distintos tipos de IG y formatos de almacenamiento:** Se han estudiado los resultados teóricos de la aplicación de la metodología de creación automática de metadatos, identificando los ítems que se crearían. La variedad y heterogeneidad de formatos y tipologías de IG hace imposible ofrecer resultados precisos sobre los ítems que pueden crearse, aportándose en su ausencia los resultados a nivel agregado por tipología de IG e identificando los ítems con cardinalidad mayor de la unidad.
- **Evaluación de la metodología de creación automática de metadatos desde el punto de vista de la interoperabilidad:** El estudio combinado de los ítems de metadatos que pueden crearse automáticamente, con la metodología y con el análisis de la interoperabilidad que aportan dichos ítems, permite evaluar, desde este último punto de vista, el grado de interoperabilidad que proporcionará la metodología de creación automática de metadatos.

## 6.1 Revisión de las cuestiones de investigación

Para facilitar la revisión de las cuestiones de investigación, se vuelven a presentar las cuestiones planteadas en el capítulo 1 para contestarlas, posteriormente, en el mismo orden:

1. ¿Se puede formalizar un modelo de interoperabilidad de sistemas para las infraestructuras de datos espaciales?
2. ¿Cuál es la aportación, en términos de interoperabilidad, de la información contenida en los metadatos?
3. ¿Se pueden crear metadatos útiles de la información geográfica de un modo automático y eficaz?

4. ¿Qué propuesta de validación es más adecuada para validar un modelo de interoperabilidad de sistemas, en el contexto de las infraestructuras de datos espaciales?
5. ¿Cuáles son las fortalezas y las debilidades de los metadatos generados manual y automáticamente, desde el punto de vista de la interoperabilidad de los sistemas que los explotarán (IDE)?

### **6.1.1 Modelo de interoperabilidad para las IDE**

Primera cuestión de investigación planteada: *¿Se puede formalizar un modelo de interoperabilidad de sistemas para las infraestructuras de datos espaciales?*

En relación con la primera cuestión de investigación, propongo la extensión del modelo de interoperabilidad LCIM definido para los Sistemas de Sistemas (SoS) con un nivel adicional, para dar soporte a los aspectos legales y organizativos de las IDE. La formalización del modelo se ha fundamentado en: (a) la hipótesis de que una IDE, analizada como un sistema de información, es un caso particular de los SoS; (b) en la revisión y el análisis de los modelos existentes, tanto en este contexto como en el de los SIG; (c) en el análisis de los objetivos propuestos en la literatura revisada para cada nivel de interoperabilidad.

Se han revisado 9 modelos (LISI, EIMM, OIMM, OIAM, LCIM, Goodchild *et al.* 1997, Bish, Intermodel5, InterOP) y las clasificaciones de la interoperabilidad propuestas por un total de 27 autores/fuentes, identificándose 15 niveles de interoperabilidad diferentes. Se han revisado más de 100 definiciones u objetivos de la interoperabilidad y su análisis ha permitido desestimar unos y reclasificar otros niveles relacionados para, finalmente, proponer un modelo aplicable a las IDE. Además de analizar la interoperabilidad en el contexto IDE, se ha revisado la literatura relacionada con su medida, verificación concluyendo que los principales objetivos son: detectar obstáculos y garantizar un determinado nivel. Para conseguir estos objetivos se requiere definir los indicadores sobre los que realizar las medidas de interoperabilidad y un patrón o escala de medida sobre el que comparar los resultados. Este patrón se ha obtenido al analizar la interoperabilidad proporcionada por todos los ítems de la norma de metadatos y puede utilizarse como referencia para realizar la medida de interoperabilidad proporcionada por un metadato.

Una vez que se ha respondido afirmativamente a la primera cuestión de investigación proponiéndose un modelo de interoperabilidad para las IDE compuesto por siete niveles que proceden del modelo LCIM, y de la revisión y análisis detallado antes. Los niveles del modelo propuesto son: técnico, sintáctico, semántico, pragmático, dinámico, conceptual y organizacional, el siguiente paso ha consistido en analizar el papel que desempeñan los metadatos en el modelo.

### **6.1.2 Análisis de la interoperabilidad proporcionada por los metadatos**

La segunda cuestión de investigación se redactó como: *¿Cuál es la aportación, en términos de interoperabilidad, de la información contenida en los metadatos?* La hipótesis inicial es que los metadatos facilitan la interoperabilidad en uno o varios niveles del modelo simultáneamente. Esta premisa ha motivado el análisis de la interoperabilidad proporcionada por los ítems de metadatos a distintos niveles de granularidad. Para responder a la segunda cuestión de investigación se han analizado tanto los ítems de metadatos pertenecientes al núcleo metadatos para los datos, como a la totalidad de los ítems que la conforman, identificando los niveles de interoperabilidad proporcionados por cada uno de ellos.

Se ha realizado el recuento de ítems que favorecen cada nivel o grupos de niveles de interoperabilidad (parejas, tríos, etc.) por medio de una encuesta realizada a 5 expertos diferentes. La comparación de los resultados obtenidos para los ítems del núcleo y para el conjunto de la norma son semejantes (correlación 0.943), detectándose que los niveles más favorecidos son: organizacional, semántico y dinámico respectivamente, y que el 46% de los ítems favorecen los tres de manera simultánea. También se puede destacar que la norma enfatiza en aspectos de interoperabilidad organizacional, ya que el 92% de los ítems que sólo favorecen un nivel lo hacen con el citado. Igualmente es notorio el bajo impacto de los metadatos en el nivel sintáctico, siendo nuestro argumento ante esta situación, que la propia norma de metadatos dota de interoperabilidad sintáctica al contenido, definiendo las reglas de codificación.

En relación con los niveles técnico y pragmático, pensamos que estos aspectos de la interoperabilidad se han de garantizar por el camino de las normas y estándares que definan los protocolos, las interfaces, etc. Finalmente, en relación la interoperabilidad conceptual, cabe destacar la importante carencia de ítems que proporcionan dicho nivel.

Adicionalmente hay que indicar que los resultados del análisis de la interoperabilidad proporcionada por los ítems de la norma de metadatos en su conjunto pueden utilizarse como patrón o escala de referencia para realizar medidas de la interoperabilidad proporcionada por los metadatos.

### **6.1.3 Metodología para la creación automática de metadatos**

En relación con la tercera cuestión de investigación que dice “¿*Se pueden crear metadatos útiles de la información geográfica de un modo automático y eficaz?*”, se propone una nueva metodología apta para automatizar la creación de metadatos una vez construido el sistema que la implementa. Nuestra hipótesis inicial es que la creación de metadatos puede automatizarse, evitando de este modo tareas rutinarias y monótonas que predisponen negativamente tanto a las organizaciones como a los operadores encargados de estas tareas. Esta hipótesis ha motivado la revisión de las metodologías existentes, la identificación de la información contenida en la IG y de aquella que está implícitamente relacionada con el formato de almacenamiento.

La metodología propuesta consta de un conjunto de etapas encargadas de realizar la extracción de la información almacenada explícitamente en la IG: interpretar o identificar datos (sistemas de referencia espaciales, formatos), realizar cálculos sobre los datos (conversión transformación de coordenadas, obtener topónimos de un nomenclátor geográfico), identificar el tipo de contenido almacenado (mediante los valores estadísticos en el caso de las imágenes, con los nombres de los atributos y los tipos de entidades en los vectoriales), derivar categorías de los temas y palabras clave que cataloguen los contenidos, sugerir un título para el metadato, identificar la estructura de la información en los formatos para argüir un modelo de datos y, finalmente,

empaquetar la información en un metadato nuevo o sobre una plantilla suministrada.

La metodología propuesta puede ser integrada en los distintos flujos de creación y actualización de metadatos detallados en la literatura: (a) creación puramente automática; (b) automática complementada por el experto en el tema de la IG; (c) automática complementada con el experto en catalogación, o finalmente (d) automática complementada por ambos (experto en el tema y en catalogación).

La eficacia del método puede ser analizada desde varios puntos de vista. En esta tesis se ha analizado dicha eficacia desde dos: las funciones que desempeñan los ítems de metadatos creados automáticamente y los niveles de interoperabilidad proporcionados por los mismos. Atendiendo a las funciones que posibilitan, los ítems creados con esta metodología favorecen las funciones de localización y el uso de los datos de un modo claro y, limitadamente, las funciones de evaluación y acceso. El segundo aspecto se trata en la quinta cuestión de investigación.

#### **6.1.4 Validación del modelo de interoperabilidad**

El modelo de interoperabilidad para las IDE, fundamentado en los metadatos y formalizado como consecuencia de las dos primeras cuestiones de investigación, podría ser cuestionado aduciendo subjetividad en la interpretación de la interoperabilidad que proporcionan los ítems de la norma de metadatos. Para dar respuesta a la cuarta cuestión de investigación concerniente a la validación del modelo de interoperabilidad: “*¿Qué propuesta de validación es más adecuada para validar un modelo de interoperabilidad de sistemas, en el contexto de las infraestructuras de datos espaciales?*”, se ha diseñado una encuesta cuyo fin es validar las interpretaciones y, por tanto, el modelo propuesto.

Se ha elegido una muestra controlada y representativa de colaboradores y se les ha solicitado que identifiquen los niveles de interoperabilidad proporcionados por los ítems del núcleo fundamental, suministrándoles las descripciones de los niveles de interoperabilidad del modelo. Se ha utilizado el núcleo por haberse demostrado la alta correlación existente (0.943) entre los resultados obtenidos al

analizar el núcleo y el conjunto de ítems que conforman la norma de metadatos. Los resultados de analizar los datos contenidos en las encuestas son: (a) cuatro de los cinco encuestados han aportando comentarios y cambios (añadir o retirar niveles de interoperabilidad aportados por los ítems); (b) se han analizado los resultados tras realizar varias agregaciones con las encuestas (los tres más críticos y los cinco encuestados) y se ha reproducido el análisis de la interoperabilidad aportada por los ítems del núcleo de metadatos generado, una vez justificados e incorporados los cambios. (c) Se ha encontrado un altísimo grado de correlación (0.994) con el modelo a validar (núcleo fundamental) y también un alto grado de correlación (0.9235) con el modelo completo (todos los ítems de la norma).

Estos resultados ratifican las conclusiones acuñadas en la respuesta a la segunda cuestión de investigación y validan el modelo, al no encontrarse discrepancias importantes en la interpretación de la interoperabilidad.

### **6.1.5 Interoperabilidad proporcionada por los metadatos creados automáticamente**

Para analizar las fortalezas y las debilidades de los metadatos generados manual o automáticamente, y de este modo responder a la quinta cuestión de investigación: “¿*Cuáles son las fortalezas y las debilidades de los metadatos generados manual y automáticamente, desde el punto de vista de la interoperabilidad de los sistemas que los explotarán (IDE)?*”, se han revisado e identificado los ítems de metadatos que se encuentran almacenados junto a la IG, y se ha revisado las formas de representar los sistemas de referencia espacial usados por la industria y la academia.

Basándose en la metodología propuesta se han identificado los ítems de metadatos que pueden crearse automáticamente para cada tipología de IG. Los resultados de éste análisis indican que en término medio se pueden crear 83 ítems para los datos raster, 69 para los vectoriales y 68 para los modelos digitales de elevación, pudiéndose ver incrementados estos valores cuando los datos ráster contienen más de 3 bandas o si los datos vectoriales almacenan más de un tipo de geometrías.

Al analizar la interoperabilidad proporcionada por los ítems de metadatos creados automáticamente se ha comprobado que aproximadamente el 60% de los ítems pertenecientes al núcleo fundamental de la norma de metadatos ISO19115 pueden crearse automáticamente. También se ha comprobado que el porcentaje de ítems creados automáticamente, respecto al conjunto de ítems de la norma de metadatos, es uniforme para todos los niveles del modelo,.

Estos resultados demuestran las fortalezas y las debilidades de los metadatos creados automáticamente con la metodología propuesta. En relación con los metadatos creados manualmente, esta tesis proporciona un marco sobre el que analizar las fortalezas y debilidades. Este marco de trabajo está formado por el modelo de interoperabilidad para las IDE, la metodología que permite analizar los niveles facilitados por los ítems de metadatos y los patrones de interoperabilidad teórica proporcionada por los ítems (tanto del núcleo fundamental como del conjunto de la norma) que posibilitan su medida.

## **6.2 Futuras líneas de investigación**

Los resultados de investigación en torno a los modelos de interoperabilidad y la creación automática de metadatos son novedosos y útiles para que los sistemas de una IDE operen entre sí. Sin embargo a lo largo del desarrollo de la tesis han surgido otras cuestiones de investigación que se proponen como futuras líneas de trabajo.

Algunas de las preguntas que han surgido son:

¿Cómo se podrían inferir o determinar las categorías de la IG almacenada como datos vectoriales? Para determinar o inferir las categorías de la IG almacenada se debería de investigar en torno a las técnicas de razonamiento, basadas en reglas almacenadas, y en los algoritmos de minería de datos para determinar su aplicabilidad en este contexto. Estas técnicas se nutrirían de los catálogos de objetos, almacenados junto a los datos, y de los nombres de las capas y de los atributos de los datos. También se debería analizar su aplicabilidad para

seleccionar palabras clave descriptivas, pertenecientes a tesauros multilingües, que ayudan a catalogar el recurso.

¿Cuál debería de ser el núcleo de metadatos que describan conjuntos de datos para que ayuden a interoperar a las IDE atendiendo al modelo propuesto? El núcleo de metadatos definido por ISO en su norma 19115 está definido para facilitar las funciones de localización y uso de los datos. Se ha demostrado que la interoperabilidad facilitada por dichos ítems no es homogénea y está focalizada en los niveles organizacional, semántico y dinámico. Por tanto la definición de un núcleo de metadatos que maximice la interoperabilidad en sus distintos niveles es un objeto de investigación para el futuro.

¿Se podría definir un núcleo fundamental de metadatos útiles para los servicios de las IDE que maximice uniformemente la interoperabilidad proporcionada por los ítems? La investigación realizada en el marco de esta tesis ha estado enfocada a los metadatos que describen los conjuntos de datos. La interoperabilidad en las IDE además de ser promovida por estos metadatos, también ha de serlo por los servicios que los ofrecen o los explotan.

¿Cuáles o cómo deberían de ser los ítems de metadatos que proporcionen interoperabilidad conceptual al describir los servicios? En el contexto de los datos, sus modelos conceptuales descritos con lenguajes de ingeniería parecen ser válidos. ¿Es extensible este razonamiento a los modelos de los servicios?

## 7 BIBLIOGRAFÍA

ALCTS (2004). ALCTS/CCS/Committee on Cataloguing: Description and Access Task Force on Metadata, Summary Report. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html>

Antonovic, V. y Novak, I. (2006). GISOSS - One-Stop-Shop GIS. Shaping the Change, *XXIII FIG Congress, Munich, Germany*, October 8-13, 2006. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://www.fig.net/pub/fig2006/papers/ts37/ts37\\_02\\_antonovic\\_novak\\_0569.pdf](http://www.fig.net/pub/fig2006/papers/ts37/ts37_02_antonovic_novak_0569.pdf)

ANZLIC (1996). ANZLIC Guidelines: Core Metadata Elements Version 1 Report: ANZLIC Working Group on Metadata, July 1996,

ANZLIC (1997). Core Metadata Elements for Land and Geographic Directories in Australia and New Zealand. The Australian New Zealand Land Information Council. Documento Web, <http://www.auslig.gov.au/pipc/anzlic/metaelem.html>.

ANZLIC (2005). Draft Anzlic ISO metadata profile. Documento Web, enlace: <http://www.gsdi.org/SDI-AP/docs2005/drftmtdt.pdf>

Arms, W., Hillmann, D., Lagoze, C., Krafft, D., Marisa, R., Saylor, J. Y Terrizzi, C. (2002). A Spectrum of Interoperability: The Site for Science Prototype for the NSDL. *D-Lib Magazine*, 8(1), January 2002. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.dlib.org/dlib/january02/arms/01arms.html>

Assche, F. (2006). An Interoperability Framework. (Learning Interoperability Framework for Europe). Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.intermedia.uio.no/confluence/display/life/An+Interoperability+Framework>

Athena Integrated Project (507849) (2005). Framework for Establishment and Management Methodology, Deliverable DA1.4 Documento Web. Visitado el 1 Octubre 2008 enlace: <http://modelbased.net/aif/methodology/eimm.html>

Bailer, W. y Schallauer, P. (1998). Metadata in the Audiovisual Media Production Process. *Studies in Computational Intelligence (SCI)* 101, 65–84 (2008).

Baird, K. (2006). Automated Metadata. March 2006, *Jorum Team*.

Balfanz, D. (2002). Automated Geodata Analysis and Metadata Generation. Society of Photo-Optical Instrumentation Engineers -SPIE-, Bellingham/Wash.: *Visualization and Data Analysis*

2002. Proceedings : 21 - 22 January 2002, San Jose, USA Bellingham/Wash.: SPIE, 2002 (SPIE Proceedings Series 4665) ISBN: 0-8194-4405-7

Batcheller, J. (2008). Automating geospatial metadata generation—An integrated data management and documentation approach. *Computers & Geosciences* 34 (2008) 387–398

Beard, K. (1996). A Structure for Organizing Metadata Collection. *Third International Conference/Workshop on Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, USA, January 21-25, 1996

Béjar, R., Noguera-Iso, J., Muro-Medrano, P. y Zarazaga-Soria, F. (2008). Systems os Systems as a Conceptual Framework for Spatial Data Infrastructures. Article Under Review for the *Internacional Journal of Spatial Data Infrastructure Research*.

Bermudez, L. (2004). Tesis Doctoral: ONTOMED: Ontology Metadata Framework, Drexel University, December 2004. Documento Web, [http://dspace.library.drexel.edu/bitstream/1860/376/8/Bermudez\\_Luis.pdf](http://dspace.library.drexel.edu/bitstream/1860/376/8/Bermudez_Luis.pdf)

Bernard, L. et al. (2005). Towards an SDI Research Agenda. Proceedings of 11th EC-GIS, Alghero, Sardinia, Italy

Bernard, L., Kanellopoulos, I., Annoni, A. y Smits, P. (2005). The European Geoportal - One step towards the Establishment of a European Spatial Data Infrastructure. *Computers, Environment and Urban Systems*, 29: 15-31.

Bishr, Y. (1998). Overcoming the semantic and other barriers to GIS interoperability. *Int. Journal for Geographical Information Science*, 12(4):299–314, 1998

C2SS WG (1996). C2SS Working Group, 1996 C2 Support Study Phase 1 Report.

Campbell, T. (2008). Fostering a Culture of Metadata Production. GSDI10: *Tenth International Conference for Spatial Data Infrastructure*, St. Augustine, Trinidad February 25-29, 2008. Enlace: <http://www.gsdi.org/gsdi10/papers/TS8.2paper.pdf>

CAPLAN, P. (1995). You call it corn, we call it syntax-independent metadata for document-like objects. *The Public Access Computer Systems Review*, v. 4, n. 6, 1995.

Caplan, P. (2003). *Metadata Fundamentals for All Librarians*. Chicago: American Library Association.

Carnegie-Mellon University Software Engineering Institute (CMU-SEI). Online Interoperability Guide. Documento Web. Visitado el 1 Octubre 2008 enlace:  
<http://www.sei.cmu.edu/isis/guide/engineering/procurement.htm>.

Carney, D., Smith, J. y Place, P. (2005). Topics in Interoperability: Infrastructure Replacement in a System of Systems (CMU/SEI-2005-TN-031). Pittsburgh, Pa: Software Engineering Institute, Carnegie Mellon University, November 2005. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.sei.cmu.edu/pub/documents/05.reports/pdf/05tn031.pdf>

CGIAR-CSI (2004). Metadata Tips. Documento Web. Visitado el 1 de Octubre de 2008 enlace:  
[http://www.csi.cgiar.org/metadata/Metadata\\_tips.asp](http://www.csi.cgiar.org/metadata/Metadata_tips.asp)

Chen, D., Daclin, N. (2007). Barriers Driven Methodology For Enterprise Interoperability. Virtual Enterprises and Collaborative Networks, 2007, pp:453-460

Chen D, Dassisti M, Elvesaeter B, (2007). Interoperability Knowledge Corpus, intermediate report IST-508 011. Documento Web

Clark, T. y Jones, R. (1999). Organisational Interoperability Maturity Model for C2. Documento Web. Visitado el 1 Octubre 2008 enlace:  
[http://www.dodccrp.org/events/1999\\_CCRTS/pdf\\_files/track\\_5/049clark.pdf](http://www.dodccrp.org/events/1999_CCRTS/pdf_files/track_5/049clark.pdf)

Codd, E. F. (1990). "The Relational Model for Database Management", Addison - Wesley, 538, 2 2.

Colleman, A. (2002). Metadata: The Theory behind the Practice. *4th State GILS Conference*, April 2002, Scottsdale, AZ.

Crompvoets, J. and Bregt A. (2003). World Status of National Spatial Data Clearinghouses. URISA Journal. Urban and Regional Information Systems Association, Vol. 15, No. 2.

Currier, S., Barton, J. y otros (2004). Quality assurance for digital learning objects repositories: issues for the metadata creation process. *ALT-F, Research in Learning Technology*, Vol.12, No. 1, March 2004

Daclin, N., Chen, D. y Vallespir, B. (2006). Enterprise interoperability measurement - Basic concepts. Proceedings of the *Open Interop Workshop on Enterprise Modelling and Ontologies for Interoperability*. ISSN 1613-0073.

Daclin, N., Chen, D. y Vallespir, P. (2008). Methodology for Enterprise Interoperability. Proceedings of the 17th World Congress: The International Federation of Automatic Control

Seoul, Korea, July 6-11, 2008. Documento Web. Visitado el 1 Octubre 2008 enlace:  
<http://www.nt.ntnu.no/users/skoge/prost/proceedings/ifac2008/data/papers/2896.pdf>

Dangermond, J. (1991). The Commercial Setting of GIS in Maguire, D.J., Goodchild, M. F. & Rhind, D.W. *Geographical Information Systems: Principles and Applications*. Longman.

Danko, D. (2002). ISO/TC 211 Geographic information/Geomatics The Standards in Action Workshop in Gyeongju, Korea: Implementation. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.isotc211.org/WorkshopGyeongju/Presentations/Metadata.ppt>

Danko, D. (2005). Interoperability & Metadata. At the *19th ISO/TC 211 Plenary meeting* in Palanza, Italy, the TC decided to install the Focus Group on Data Producers (FGDP). Enlace: [http://www.isotc211fgdp.info/docs/figws/session2/Dave\\_Danko\\_%20Metadata.ppt](http://www.isotc211fgdp.info/docs/figws/session2/Dave_Danko_%20Metadata.ppt)

Dekkers, M. (2007). Metadata and modelling for Interoperability. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://library2.nalis.gov.tt/Portals/0/cdl\\_Makx\\_Dekkers\\_20070712.pdf](http://library2.nalis.gov.tt/Portals/0/cdl_Makx_Dekkers_20070712.pdf)

Díaz, L. Granell, C. Beltrán, A. Llaves, A. y Gould, M. (2008a). Extracción Semiautomática de Metadatos: Hacia los metadatos implícitos. *II Jornada de SIG Libre*. Universidad de Girona.

Díaz, L., Gould, M., Beltrán, A., Llaves, A. y Granell, C. (2008). Multipurpose Metadata Management in gvSIG. Proceeding of the academic track of the *2008 Free and Open Source Software for Geospatial (FOSS4G) Conference*, 29 September – 3 October 2008, Cape Town, South Africa. ISBN: 978-0-620-42117-1, pp 90-99

Ding, H. (2005). Challenges in Building Semantic Interoperable Digital Library System. Documento Web. Visitado el 1 Octubre 2008 enlace:  
<http://www.idi.ntnu.no/grupper/su/courses/dif8901>

Duval, E., Hodgins, W., Sutton, S., y Weibel, S. (2002). Metadata Principles and Practicalities. *D-Lib Magazine*. Documento Web. Visitado el 1 Octubre 2008 enlace:  
<http://www.dlib.org/dlib/april02/weibel/04weibel.html>.

ECNBII, (2003). FGDC Biological Data Profile As it maps Dublin Core. GeoConnection. Documento Web. Visitado el 1 Octubre 2008 enlace:  
[http://www.geoconnections.org/developersCorner/devCorner\\_devNetwork/meetings/2003.06.10/Present/ECNBII.ppt](http://www.geoconnections.org/developersCorner/devCorner_devNetwork/meetings/2003.06.10/Present/ECNBII.ppt)

Ercegovac, Z. (1999). Introduction. *Journal of the American Society for Information Science*, v. 50, n. 13, p. 1165-1168, 1999

- Flater, D. (2002). Impact of Model-Driven Standards, *Proceedings of the 35th Hawaii International Conference on System Sciences – 2002*. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://csdl2.computer.org/comp/proceedings/hicss/2002/1435/09/14350285.pdf>
- FGDC (2000). Content Standard for Digital Geospatial Metadata Workbook, version 2.0. Federal Geographic Data Committee (USA), 2000.
- Gayatri y Ramachandran, S. (2007). Understanding Metadata. *The Icfai Journal of Information Technology*, March 2007.
- Georgiadou, Y., Puri S. y Sahay S. (2005). Towards a potential research agenda to guide the implementation of Spatial Data Infrastructures - A case study from India. *International Journal of Geographical Information Science* 19(10): 1113—1130
- Gilliland-Swetland, A. (2000). Setting the Stage, in *Introduction to Metadata: Pathways to Digital Information*. Los Angeles: Getty Research Institute. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://www.getty.edu/research/conducting\\_research/standards/intrometadata/2\\_articles/index.html](http://www.getty.edu/research/conducting_research/standards/intrometadata/2_articles/index.html)
- Goh, C. (1997). Representing and Reasoning about Semantic Conflicts in Heterogeneous Information Sources, Ph.D. Thesis, MIT Sloan School of Management, 1997
- Goodchild, M., Egenhofer, M., Fegeas, R. (eds.) (1997). Interoperating GISs: Report of a specialist meeting held under the auspices of the VARENIUS project. *Panel on computational implementation of Geographic concepts*. Santa Barbara (California), 5-6 December 1997. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://www.ncgia.ucsb.edu/Publications/Varenius\\_Reports/Interop.pdf](http://www.ncgia.ucsb.edu/Publications/Varenius_Reports/Interop.pdf)
- Goodchild, M. (2007). Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0, *International Journal of Spatial Data Infrastructures Research*, 2007, Vol. 2, 24-32.
- Gordon, D. (2003). What is Interoperability and why does it matter? *MapInfo Magazine Volume 7, n°4. 2003*. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://resource.mapinfo.com/static/files/document/1074288321775/mapinfomag\\_summer2003.pdf](http://resource.mapinfo.com/static/files/document/1074288321775/mapinfomag_summer2003.pdf)
- Greenberg, J. (2004). Metadata extraction and harvesting: a comparison of two automatic metadata generation applications. *Journal of Internet Cataloging* 6 (4), 59–82.
- Greenberg, J., Spurgin, K. y Crystal, A. (2006). Functionalities for automatic metadata generation applications: a survey of metadata experts' opinions. *Int. J. Metadata, Semantics and Ontologies, Vol. 1, No. 1, 2006*.

Groot, R. y J. McLaughlin, Eds. (2000). Geospatial data infrastructure - Concepts, cases, and good practice. Oxford, Oxford University Press.

Guptill, S.G. (1999). Metadata and data catalogues. In: Longley, P., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (Eds.), *Geographical Information Systems*. Wiley, Chichester, pp. 677–692

Guy, M., Powell, A. y Day, M. (2004). *Improving the Quality of Metadata in Eprint Archives*. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.ariadne.ac.uk/issue38/guy/>

Hamilton, J., Rosen, J. y Summers, P. (2004). Developing Interoperability Metrics, in joint command and control interoperability: cutting the gordian knot, Chapter 6 (2004).

Harvey, F., Kuhn, W., Pumdt, H. y Bishr, Y. (1999). Semantic interoperability: A central issue for sharing geographic information. *The Annals of Regional Science*, Vol. 33, No. 2. (10 May 1999), pp. 213-232

Hedorfer, M. y Bianchin, A. (1999). The Venice Lagoon Experimental GIS at the IUAV. *Interop99: The 2nd International Conference on Interoperating Geographic Information Systems*. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.hedorfer.it/docs/rsalv/rsalv1io-ENG.pdf>

Heery, R. Biblink: LB4034 D1.1 metadata formats. [S. l.] : Biblink, 1996.

IDABC (2006). Interoperability for eGovernment Dec 2006. European Interoperability Framework for pan-European eGovernment services. Office for Official Publications of the European Communities, Luxembourg. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://ec.europa.eu/idabc/en/document/3473/5585>.

INSPIRE Metadata IR (2008). Draft Guidelines – INSPIRE metadata implementing rules based on ISO 19115 and ISO 19119. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://www.ec-gis.org/inspire/reports/ImplementingRules/metadata/Draft\\_Guidelines%20INSPIRE\\_metadata\\_implementing\\_rules.pdf](http://www.ec-gis.org/inspire/reports/ImplementingRules/metadata/Draft_Guidelines%20INSPIRE_metadata_implementing_rules.pdf)

IEEE (1990). Standard Computer Dictionary—A Compilation of IEEE Standard Computer Glossaries. New York, NY: 1990.

ISO-14258 (1999). Industrial automation systems - Concepts and rules for enterprise models.

ISO 19101 (2002). Geographic Information – Reference Model. International Standard Organization

ISO 19115 (2003). Geographic Information – Metadata

ISO 19119 (2005). Geographic Information – Services. International Standard Organization

ISO 19128 (2005). Geographic Information – Web Map Service Interface

ISO 19142 (2008). Geographic Information – Web Feature Service

ISO 23950 (1998). Information and documentation -- Information retrieval (Z39.50) -- Application service definition and protocol specification. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://www.iso.org/iso/catalogue\\_detail?csnumber=27446](http://www.iso.org/iso/catalogue_detail?csnumber=27446)

Janowicz, K., Raudal, M., Schwering, A. y Kuhn, W. (2008). Semantic Similarity Measurement and Geospatial Applications. *Transaction in GIS* December 2008, Volume 12, Issue 6, pp 651-659.

Johnson, F. (1995). Automatic abstracting research, *Library Review*, Vol. 44, No. 8, pp.28–36.

JORUM (2004). The JISC Online Repository for [learning and teaching] Materials: *JORUM Scoping and Technical Appraisal Study*, Volume V: Metadata

Johnston, P. (2005). Good Practice Guide for Developers of Cultural Heritage Web Services. Research Officer, UKOLN. Enlace: <http://www.ukoln.ac.uk/interop-focus/gpg/Metadata/>

Jones, M. y Taylor, G. (2003). Metadata: Spatial Data Handling and Integration Issues. School of Computing Technical Report. Issued: February 2003.

Kalantari, Mohsen, Rajabifard, Wallace y Williamson (2006). An interoperability toolkit for e-Land administration, Williamson, Enemark and Wallace (eds), Sustainability and Land Administration Systems, Department of Geomatics, Melbourne , 213-222

Kasunic, M. y Anderson, W. (2004). Measuring systems interoperability: challenges and opportunities, Software engineering measurement and analysis initiative, Technical note CMU/SEI – 2004 – TN – 003, 2004.

Kildow, M. (1996). The value of Metadata (A NSDI report). US Fisheries and Wildlife Services. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.r1.fws.gov/metadata/meta.html>.

- Kingston, G., Fewell, S. y Richer, W. (2005). An Organisational Interoperability Agility Model. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA463924&Location=U2&doc=GetTRDoc.pdf>
- Korth, H. F. and A. Silberschatz (1991). "Database Systems Concepts". Singapore, McGraw - Hill, 694, 2 2.
- Kuhn, W. (2003). Semantic Reference Systems. *International Journal of Geographical Information Science* 17(5): 405-409
- Kuhn, W. y Raubal, M. (2003). Implementing semantic reference systems. *The 6th AGILE Conference in GIScience*, Lyon France, April 24-26 2003.
- Kuhn, W. (2005). Geospatial Semantics: Why, of What, and How? *Journal on Data Semantics III* (2005), pp. 1-24
- Lemmens, R. (2006). Semantic interoperability of distributed geo-services. Publication on Geodesy 63. NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission. Delft September 2006.
- Lewis, G. y Wrage, L. (2005). Approaches to Constructive Interoperability (CMU/SEI-2004-TR-020 ESC-TR-2004-020). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 2005. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.sei.cmu.edu/publications/documents/04.reports/04tr020.html>
- LISI (1997). Levels of Information Systems Interoperability. Software Engineering Institute. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.sei.cmu.edu/isis/guide/introduction/lisi.htm>
- Longley, P., Goodchild, M., Maguire, D. y Rhind, D. (2000). *Geographic Information Systems and Science*, 2nd Edition, Wiley Europe.
- Longhorn, R. (2008). Burning Issue. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://www.geoconnexion.com/burning\\_issue\\_int.php](http://www.geoconnexion.com/burning_issue_int.php)
- Losee, R. (2003). Adaptive organization of tabular data for display, *Journal of Digital Information*, Vol. 4, No. 1, Documento Web. Visitado el 1 Octubre 2008 enlace: <http://jodi.ecs.soton.ac.uk/Articles/v04/i01/Losee/>.
- Maier, M. (1996). Architecting Principles for Systems-of-Systems. In *6th Annual International Symposium of INCOSE*, Boston, MA, USA, p. 567-574.

Manso, M., Noguerras, J., Zarazaga, J. y Bernabé, M. (2004). Automatic Metadata Extraction from Geographic Information. *AGILE 2004. 7th Conference on Geographic Information Science, conference Proceedings*, 379-385.

Manso, M. y Bernabé, M. (2005). CSCAT: Translation Web Service & Coordinate Systems Catalogue. *GISPlanet 2005 Conference*, 30 May - 4 June 2005, Estoril, Portugal

Miller, P. (2000). Interoperability. What is it and Why should I want it? *Ariadne Issue 24*. 21-Jun-2000. Documento Web. Visitado el 1 Octubre 2008 enlace:  
<http://www.ariadne.ac.uk/issue24/interoperability/intro.html>

Milstead, J. y Feldman, S. (1999). Metadata: Cataloging by any other name. Online 25-31. Documento Web. Visitado el 1 Octubre 2008 enlace:  
<http://www.onlineinc.com/onlinemag/OL1999/milstead1.html>.

Moellering, H. y Brodeur, J. (2006). Towards a North American Profile of the ISO 19115 World Spatial Metadata Standard. *GSDI-9 Conference Proceedings*, 6-10 November 2006, Santiago, Chile. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://gsdidocs.org/gsdiconf/GSDI-9/papers/TS12.4paper.pdf>

Mohammadi, M., Binns, A., Rajabifard, A. y Williamson, I. (2006). Spatial Data Integration. *17th UNRCC-AP Conference and 12th Meeting of the PCGIAP*, Bangkok, 18-22 September 2006. Documento Web. Visitado el 1 Octubre 2008 enlace:  
[http://www.geom.unimelb.edu.au/research/SDI\\_research/publications/files/Spatial%20Data%20Integration.doc](http://www.geom.unimelb.edu.au/research/SDI_research/publications/files/Spatial%20Data%20Integration.doc)

Morris, S., Nagy, Z. y Tuttle, J. (2007). North Carolina Geospatial Data Archiving Project. *NCSU Libraries and North Carolina Center for Geographic Information & Analysis*. Documento Web. Visitado el 1 Octubre 2008 enlace:  
[http://www.digitalpreservation.gov/partners/ncgdap/high/NCGDAP\\_InterimReport\\_June2008\\_final.pdf](http://www.digitalpreservation.gov/partners/ncgdap/high/NCGDAP_InterimReport_June2008_final.pdf)

Mustacoglu, A. (2007). Event-Based Model for Reconciling Digital Entries. Ph.D. Proposal Report. Department of Computer Science, Indiana University, Bloomington. May 3 2007.

Nadkarni, P., Chen, R. y Brandt, C. (2001). UMLS concept indexing for production databases: a feasibility study, *Journal of the American Medical Information Association*, Vol. 8, No. 1, pp.80-91

Najar, C. (2006). A model-driven approach to management of integrated metadata – spatial data in the context of spatial data infrastructure. Doctoral Thesis. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://e-collection.ethbib.ethz.ch/view/eth:28733>

Najar, C. y Giger, C. (2006). Spatial Data and Metadata Integration for SDI interoperability. Under review for the *international Journal of Spatial Data Infrastructures Research*, submitted 2006-09-25.

NATO (2004). NATO C3 Board: NATO C3 System Architecture Framework (NAF) (Document AC/322-D(2004)0002 (INV)) NATO, 2004

Nebert, D. (2004). Developing Spatial Data Infrastructures: The SDI Cookbook. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf>

Nedovic-Budic, Z. y Pinto, J. (2001). Organizational (soft) GIS interoperability: lesson from the U.S. *International Journal of Applied Earth Observation and Geoinformation*, Volume 3, Number 3, 2001, pp. 290-298(9). Elsevier.

Nell, J. D. (1996) TC184 SC5 WG1 convener Report.

Nogueras-Iso, J., Zarazaga-Soria, F. y Muro-Medrano P. (2005). *Geographic Information Metadata for Spatial Data Infrastructures. Resources, Interoperability and Information Retrieval*. Springer editions. ISBN:3-540-24464-6.

Nowak, J. y Nogueras, J. (2005). Issues of multilingualism in creating a European SDI – The perspective for Spatial Data Interoperability. 29 June- 1 July 2005 Sardinia. 11th EC-GI & GIS Workshop. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.ec-gis.org/Workshops/11ec-gis/>

NISO (2004). National Information Standards Organization. (2004). *Understanding Metadata*. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

OpenGIS Guide (1998). Open GIS Consortium Technical Committee, *The OpenGIS Guide: Introduction to Interoperable Geoprocessing and the OpenGIS Specification, Third Edition, Draft*, June 3, 1998.

Oosterom, P. (2004). *Geo-information Standards in Action*. ISO TC211/Metadata (Danko, D.). Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.ncg.knaw.nl/Publicaties/Groen/pdf/42Standards.pdf>

Ostensen, O. y Danko, D. (2005). Global Spatial Metadata Activities in the ISO/TC211 Geographic Information Domain. World Spatial Metadata Standards: Scientific and Technical Descriptions, and Full Descriptions with Crosstable. H. Moellering, H.J.G.L. Aalders & A. Crane (Editors). Elsevier Ltd.

Ouksel, M. (1999). A Framework for a Scalable Agent Architecture of Cooperating Heterogeneous Knowledge Sources in Intelligent Information Agents: Agent-Based Information Discovery and Management of the Internet, Matthias Klusch Ed., Springer Verlag, 1999.

Ouksel, A. y Sheth, A. (1999). Semantic Interoperability in Global Information Systems: A Brief Introduction to the Research Area and the Special Section. SIGMOD Record 28(1): 5-12(1999)

Pasquinelli, A. (2003). Information technology directions in libraries: a sun microsystems white paper. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.sun.com/products-n-solutions/edu/libraries/libtechdirection.html>.

Pierkot, C. (2006). Using Metadata to Help the Integration of Several Multi-source Sets of Updates. *GSDI-9 Conference Proceedings*, 6-10 November 2006, Santiago, Chile. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.gsdi9.cl/english/papers/TS12.2paper.pdf>

Phillips, A. Willianson, I and Ezigbalike, C. (1998). The importance of Metadata Engines in Spatial Data Infrastructures. The 26th Annual Conference of AURISA Perth, Western Australia, 23-27 November 1998. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://www.sli.unimelb.edu.au/research/publications/IPW/phillips\\_NZ\\_981.htm](http://www.sli.unimelb.edu.au/research/publications/IPW/phillips_NZ_981.htm)

Pokraev, S., Reichert, M., Steen, M. y Wieringa, R. (2005). Semantic and Pragmatic Interoperability: A Model for Understanding. *Proceedings of the CAiSE*. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://ceur-ws.org/Vol-160/paper21.pdf>

Pridmore, J. y Rumens, D. (1989). Interoperability-how do we know when we have achieved it? Command, Control, Communications and Management Information Systems, May 1989, 192-205, ISBN:-85296-380-7

Probst, F. (2006). Ontological Analysis of Observations and Measurements. *International Conference on Geographic Information Science* No4, Münster, ALLEMAGNE (2006), vol. 4197, pp. 304-320. ISBN 3-540-44526-9; 978-3-540-44526-5

Rajabifard, A. y Binns, A. (2004). Facilitating Virtual Australia Utilizing an SDI Enabled Platform. December 2004. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://datasmart.oesr.qld.gov.au/Events/datasmart.nsf/0/F0E6094FCB6E7F124A256FA2007C2274/\\$FILE/QLD%20Seminar-CRC-Project%203.1%20-%20Dec%202004.pdf](http://datasmart.oesr.qld.gov.au/Events/datasmart.nsf/0/F0E6094FCB6E7F124A256FA2007C2274/$FILE/QLD%20Seminar-CRC-Project%203.1%20-%20Dec%202004.pdf)

- Rawat, S. (2003). Interoperable Geo-Spatial data model in the Context of the Indian NSDI. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://www.itc.nl/library/Papers\\_2003/msc/gfm/sujata.pdf](http://www.itc.nl/library/Papers_2003/msc/gfm/sujata.pdf)
- SAGA (2006). Standards and Architectures for eGovernment Applications SAGA. October 2006. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.epractice.eu/document/3210>
- Schekkerman, J. (2004). President of the Institute for Enterprise Architecture Development (IFEAD). Roadmap to implementation. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://web-services.gov/region4soa10104.ppt>
- Senso, J. y Rosa Piñero, A. (2003). El concepto de metadato. Algo más que descripción de recursos electrónicos. *Ci. Inf., Brasília*, v. 32, n. 2, p. 95-106, maio/ago. 2003.
- Shanzhen, Y., Qi, L. y Jicheng, C. (1999). An interoperability GIS model based on the spatial information infrastructure. *Geoinformatics and Socioinformatics. The Proceedings of Geoinformatics'99 Conference. Ann Arbor, 12-21 June*, pp, 1-5.
- Shekhar, S. (2004). Spatial Data Mining and Geo-spatial Interoperability. Report of the NCGIA Specialist Meeting on Spatial Webs, Santa Barbara, December 2-4 2004, National Center for Geographic Information and Analysis, University of California
- Sheldon, T. (2001). Linktionary. Entrada «Metadata». Visitado 29 mayo 2006
- Steinacker, A., Ghavam, A., y Steinmetz, R. (2001). Metadata Standards for Web-Based Resources. *IEEE MultiMedia*, enero-marzo 2001.
- Swick, R. (2002). Metadata Activity Statement. Visitado 29 mayo 2006
- Stroetmann, K. (2005). Towards an Interoperability framework for European e-Health Research area. Brussels, Feb. 14-15, 2005. Documento Web. Visitado el 1 Octubre 2008 enlace: [http://www.i2-health.org/i2-h-presentations/WHO-WS\\_Semantic-IOP\\_KAS\\_2005-02-14.pdf](http://www.i2-health.org/i2-h-presentations/WHO-WS_Semantic-IOP_KAS_2005-02-14.pdf)
- Struetmann, K. (2005). Towards an Interoperability framework for European e-Health Research area. Locating the Semantic Interoperability Domain. WHO/EC Workshop on semantic interoperability, Brussels, Feb. 14-15, 2005. Documento Web. Visitado el 1 Octubre 2008 enlace: <https://www.who.int/classifications/terminology/stroetmann.pdf>
- SUMM (1991). Semantic Unification Meta Model, ISO/IEC JTC1 SC2 WG3, N1360, 1991-Oct

- Taylor (2004). *The Organization of Information*. 2nd ed. Westport, CN: Libraries Unlimited
- Tolk, A. (2003). Beyond Technical Interoperability—Introducing a Reference Model for Measures of Merit for Coalition Interoperability, Proceedings of the 8th ICCRTS, Washington, D.C., June 17-19, 2003
- Tolk, A. y Muguira, K. (2003). The Levels of Conceptual Interoperability Model (LCIM). Proceedings of the 2003 Fall Simulation Interoperability Workshop, Orlando, FL, September 2003
- Turnitsa, C. y Tolk, A. (2006). Battle Management Language: A Triangle with Five Sides Proceedings of the Simulation Interoperability Standards Organization (SISO) Spring Simulation Interoperability Workshop (SIW), Huntsville, AL, April 2-7, 2006
- Tolk, A., Diallo, S. y Turnitsa, C. (2007). Applying the levels of Conceptual Interoperability Model in Support of Integrability, Interoperability, and Composability for System-of-Systems Engineering. *International Journal Systemics, Cybernetics and Informatics*. Volume 5 – Number 5
- Vckovski, A. (1998). *Interoperable and Distributed Processing in GIS (Research Monographs in GIS)*. Taylor and Francis.
- West, J. y Hess, T. (2002). Metadata as a knowledge management tool: supporting intelligent agent and end user access to spatial data. *Decision Support Systems* 32, 247–264
- Wieringa, R. (2003). *Design Methods for Reactive Systems: Yourdon, Statemate, and the UML*. Morgan Kaufmann, 2003. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://www.mkp.com/dmrs>
- William A et al. (2002). A Spectrum of Interoperability: The Site for Science Prototype for the NSDL. *D-Lib Magazine*, 8(1), January 2002. Documento Web. Visitado el 1 de Octubre de 2008 enlace: <http://www.dlib.org/dlib/january02/arms/01arms.html>
- Williamson, I., Rajabifard, A. y Feeney, M. (2003). *Developing Spatial Data Infrastructures: From Concept to Reality*, ISBN 0-415-30265-X, Taylor & Francis, U.K.
- Williamson, I. (2004). Building SDIs—the challenges ahead. In Proceedings of the 7<sup>th</sup> *International Conference: Global Spatial Data Infrastructure*, 2–6 February, Bangalore, India.
- Wilson, E. (1998). *Manuals Go Click. The Age*. Melbourne, Australia

Whitman, L., Santanu, D. y Panetto, H. (2006). An enterprise model of interoperability. *Information Control Problems in Manufacturing*, Volume 12, Part 1. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://hal.archives-ouvertes.fr/docs/00/08/20/37/PDF/366a.pdf>

Woodley, M. S., Clement, G. y Winn, P. (2003). DCMI Glossary. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://dublincore.org/documents/2003/08/26/usageguide/glossary.shtml>

Woodley M, Clement, G y Winn, P. (2005). Dublin Core Metadata Initiative. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://dublincore.org/documents/usageguide/glossary.shtml>

Woodley, M. (2004). Dublin Core Metadata Initiative. Documento Web. Visitado el 1 Octubre 2008 enlace: <http://dublincore.org/documents/usageguide/glossary.shtml>

Wright-Patterson, T.P. (1936). Factors Affecting the Cost of Airplanes, *Journal of Aeronautical Sciences*, 3(4) (1936): 122-128.

Wyoming University. Metadata Resources at the University of Wyoming. <http://www.sdvc.uwyo.edu/metadata/why.html>

Yilmaz, L. yTolk, A. (2006). Engineering Ab Initio Dynamic Interoperability and Composability Via Agent-Mediated Introspective Simulation. In *Proceedings of the 2006 Winter Simulation Conference*, page(s): 1075-1082. Monterey, CA, USA Dec 2006. ISBN: 1-4244-0500-9.

Zeigler, B. y Hammonds, P. (2007). Modelling and simulation-based data engineering. Introducing Pragmatics into Ontologies for Net-Centric Information Exchange. Elsevier Academic Press. ISBN:9780123725158