

**UNIVERSIDAD SAN FRANCISCO DE QUITO**

**Diseño y desarrollo del geoportal piloto en el marco del Sistema  
Territorial de Información Integrada de los Patrimonios (STIIP) del  
Ministerio Coordinador de Patrimonio del Ecuador**

**Erick Patricio Bastidas Ruiz**

Tesis de grado presentada como requisito para la obtención del título de  
Ingeniería de Sistemas

Quito, mayo del 2012

**Universidad San Francisco de Quito  
Colegio Politécnico**

**HOJA DE APROBACION DE TESIS**

**Diseño y desarrollo del geoportal piloto en el marco del Sistema  
Territorial de Información Integrada de los Patrimonios (STIIP) del  
Ministerio Coordinador de Patrimonio del Ecuador**

**Erick Patricio Bastidas Ruiz**

Fausto Pasmay, M.S.  
Director de Tesis y  
Miembro del Comité de Tesis

---

Lorena Balseca, Ing.  
Miembro del Comité de Tesis

---

Carlos Mena, Ph.D.  
Miembro del Comité de Tesis

---

Santiago Gangotena, Ph.D.  
Decano del Colegio Politécnico

---

Quito, mayo del 2012

© Derechos de autor  
Erick Patricio Bastidas Ruiz  
2012

## AGRADECIMIENTO

Mi eterno reconocimiento y agradecimiento a las autoridades y profesores de la Universidad San Francisco de Quito, por la educación integral, humanística, científica y profesional dentro de la filosofía de las Artes Liberales que me impartieron; a mis padres y hermanos, por su respaldo incondicional y sabios consejos; y, a todas aquellas personas que han estado presentes a lo largo de mi vida universitaria, que de una u otra forma contribuyeron a mi crecimiento como persona.

*Erick Bastidas Ruiz*

## RESUMEN

El proyecto *Sistema Territorial de Información Integrada de los Patrimonios (STIIP)* se basa en la realización del diseño y desarrollo del prototipo de geoportal en el marco del diseño conceptual, para el Ministerio Coordinador de Patrimonio del Ecuador (MCP). Este proyecto tiene el propósito de integrar y georreferenciar la información de todas las instituciones coordinadas por el MCP para la creación, análisis, búsqueda y presentación de datos e indicadores patrimoniales del Ecuador.

En el **diseño conceptual**, se evalúa las instituciones del Consejo Sectorial del Patrimonio y se analiza los requerimientos para la creación de un Sistema de Gestión Documental interinstitucional. Además, se diseña una estructura de funcionamiento en diversos niveles y procesos para el sistema de información territorial. Luego, se desarrolla el **prototipo de un motor de búsqueda**, cuyo núcleo es el metadato en el índice del servidor Apache Solr, y un **geoportal**, con las herramientas de software MapServer y una base de datos de los territorios georreferenciada en PostgreSQL y PostGIS.

El rendimiento del motor de búsqueda registró un tiempo de búsqueda de 0.5 segundos para 1.2 millones de términos únicos. El proyecto STIIP logró integrar la información georreferenciada del patrimonio de todo el país y mejorar la eficiencia en el manejo de datos, toma de decisiones, transparencia de la información y participación ciudadana.

*Palabras claves:* STIIP, MCP, geoportal, motor de búsqueda, Solr, MapServer, PostGIS, diseño conceptual

## ABSTRACT

The Project *Integrated Territorial Information System of Heritage (ITISH, or STIIP in spanish)* is based in the elaboration of the design and development of the prototype of a geoportal under the conceptual design, for the Ministry Coordinator of Heritage of Ecuador (MCH). This project pretends to integrate and georeference the information of all of the MCH's coordinated institutions for the creation, analysis, search, and display of the heritage data and indicators from all over Ecuador.

The **conceptual design**, evaluate the institutions of the Heritage Sectorial Council and analyze the requirements for the creation of an interinstitutional Document Management System. It is also designed the operating structure in different levels and processes for the territorial information system. Later, it is developed the **prototype of a search engine**, which core is the metadata of the index from Apache Solr server; and a **geoportal**, using the software tools MapServer, and a georeferenced territories data base in PostgreSQL and PostGIS.

The overall performance of the search engine registered a query speed time of 0.5 seconds for 1.2 million unique terms. The Project ITISH achieved to integrate the georeferenced information of the country's heritage and improve the efficiency of data management the decision making, transparency of information, and citizen participation.

*Keywords:* STIIP, ITISH, MCH, MCP, geoportal, search engine, Solr, MapServer, PostGIS, conceptual design.

## ÍNDICE

	Pág.
<b>1 INTRODUCCIÓN</b> .....	2
1.1 Antecedentes del STIIP.....	2
1.2 Objetivo general.....	4
1.3 Objetivos específicos .....	6
1.4 Proyectos relacionados del STIIP .....	6
1.5 Justificación e importancia del STIIP .....	7
1.6 Centro de Transferencias y Tecnología de la USFQ .....	8
 <b>2 MARCO TEÓRICO</b> .....	 10
2.1 Definiciones generales .....	10
2.2 Qué es un Sistema de Información Geográfica.....	13
 <b>3 METODOLOGÍA</b> .....	 15
 <b>4 ANÁLISIS</b> .....	 17
4.1 Situación actual en el Consejo Sectorial de Patrimonio .....	17
4.2 Instituciones coordinadas por el MCP .....	19
4.2.1 Ministerio de Cultura del Ecuador.....	21
4.2.2 Ministerio del Ambiente.....	22
4.2.3 Ministerio del Deporte.....	22
4.2.4 Ferrocarriles del Ecuador.....	23
4.2.5 Instituto Nacional de Patrimonio Cultural – INPC.....	24
4.2.6 Dirección Nacional de Salud de los Pueblos Indígenas .....	25
4.2.7 Otras instituciones .....	25
4.2.7.1 Ecorae.....	25
4.2.7.2 Dirección de Educación Intercultural Bilingüe – DINEIB .....	26
4.3 Requerimientos del Consejo Sectorial de Patrimonio .....	26
4.4 Sistema de gestión documental.....	29
4.4.1 ¿Qué es un sistema de gestión documental?.....	29
4.4.2 Sistema de Gestión Documental aplicado al STIIP.....	31
4.4.3 Escenarios posibles.....	32

<b>5 DISEÑO CONCEPTUAL DEL STIIP</b> .....	34
5.1 Diagramas de Flujo de Datos del STIIP .....	34
5.2 Diagrama de contexto .....	35
5.3 Diagrama de nivel superior .....	37
5.4 Organización de los almacenes de datos.....	39
5.5 Procesos de creación del metadato y generación del índice .....	41
5.5.1 El metadato .....	41
5.5.2 El índice.....	45
5.5.3 El SDG-STIIP componente automático .....	46
5.5.4 Algoritmo de la araña .....	48
5.5.5 Protocolos de copia de datos .....	52
5.6 Procesos de protocolo de manejo de datos y de metadatos .....	55
5.6.1 El SDG-STIIP componente humano .....	57
5.6.2 Definición de los elementos del manejo de datos y de metadatos.....	57
5.6.3 Algoritmo del SGD-STIIP componente humano .....	60
5.7 Proceso de búsqueda .....	62
5.8 Proceso de geoportal .....	65
5.9 Proceso de espacialización.....	67
5.10 Proceso de modelado y simulaciones espaciales.....	69
5.11 Proceso de gestión y análisis de indicadores.....	70
5.12 Proceso de Patrimonio participativo .....	72
5.13 Resumen de las características del SGD-STIIP .....	75
5.14 Resumen de los procesos del STIIP .....	77
<b>6 DISEÑO TÉCNICO DEL MOTOR DE BÚSQUEDA Y DEL GEOPORTAL STIIP-PILOTO</b> .....	79
6.1 Campos del metadato .....	80
6.1.1 El sistema cartográfico base .....	80
6.1.2 Campos del Dublin Core .....	82
6.1.3 Campos de dominio de identidad y territoriales.....	85
6.1.4 Campos de otras secciones del metadato del STIIP .....	88
6.2 Arquitectura del motor de búsqueda.....	89
6.2.1 Lucene: Motor de búsqueda textual.....	90
6.2.2 Solr: Servidor empresarial de motor de búsqueda.....	91

	Pág.
6.2.3 Tika: Análisis de contenido textual .....	92
6.2.4 Nutch: Araña y motor de búsqueda .....	93
6.3 Arquitectura del geoportal .....	94
6.3.1 Estructura integrada de software del geoportal .....	96
6.3.2 Algunas definiciones y requerimientos de interoperabilidad .....	100
6.3.2.1 Servicios de mapas (Web Map Service, WMS).....	100
6.3.2.2 Servicio de publicación de entidades (Web Feature Service, WFS)....	101
6.3.2.3 Servicio de catálogo (Web Catalog Service, WCS).....	102
6.3.3 PostgreSQL: Sistema de gestión de base de datos .....	103
6.3.4 PostGIS: Extensión espacial de PostgreSQL.....	103
6.3.5 MapServer: Servidor de servicios geográficos .....	105
<b>7 DESARROLLO DEL STIIP-PILOTO .....</b>	<b>106</b>
7.1 Especificaciones técnicas para el uso o desarrollo del STIIP-Piloto .....	107
7.1.1 Para el usuario .....	107
7.1.2 Para el desarrollador .....	108
7.2 Motor de búsquedas y geoportal .....	110
7.2.1 Archivos de configuración.....	111
7.2.1.1 Definición del schema.xml.....	111
7.2.1.2 Definición del manejador de búsqueda “DisMax” .....	112
7.2.2 Integración del motor de búsqueda con el geoportal.....	113
<b>8 RESULTADOS .....</b>	<b>115</b>
8.1 Rendimiento del motor de búsqueda.....	115
8.2 Implementación posterior del STIIP .....	115
<b>9 CONCLUSIONES .....</b>	<b>117</b>
<b>10 RECOMENDACIONES .....</b>	<b>119</b>
<b>11 BIBLOGRAFÍA.....</b>	<b>122</b>
<b>12 GLOSARIO DE TÉRMINOS MÁS USADOS .....</b>	<b>125</b>

<b>13 ANEXOS</b> .....	126
Anexo 1. Código fuente.....	126
1. Código del motor de búsqueda (en java).....	126
2. Código del geoportal (en php).....	133
Anexo 2. Capturas de pantalla del proyecto SIPE .....	136
Anexo 3. Capturas de pantalla de Portal de promoción de la cultura de España .....	137
Anexo 4. Captura de pantalla de los resultados de la búsqueda de la palabra "riobamba" en el Google.....	138
Anexo 5. Captura de pantalla del geoportal STIP en la página del MCP .....	139
y acta de entrega del proyecto .....	139
Anexo 6. Lista de categorías para la búsqueda avanzada.....	143
Anexo 7. Manual de operación: Inicio rápido .....	146
Anexo 8. Manual de usuario.....	149
1. Mapa del sitio .....	149
2. Tipos de búsquedas.....	150
2.1 Búsquedas generales.....	151
2.2 Búsquedas específicas .....	154
2.3 Búsqueda avanzada.....	155
2.4 Búsquedas facetada o desde el árbol de categorías .....	156
3. Geoportal y el visualizador de resultado en el visor de mapas.....	159
4. Preferencias o sitio de administración.....	160
4.1 Directorios .....	161
4.2 Bases de Datos.....	163
4.3 Hojas de cálculo.....	166

## ÍNDICE DE FIGURAS

	Pág.
Figura 2-1 Modelo Vista Controlador .....	11
Figura 3-1 Desarrollo Rápido de Aplicaciones .....	16
Tabla 4-1 Estado del manejo de datos en instituciones coordinadas del MCP .....	20
Tabla 4-2 Áreas de un Sistema de Gestión Documental .....	30
Figura 5-1 Símbología de los Diagramas de Flujo de Datos (DFD) del STIIP.....	35
Figura 5-2 Diagrama de contexto del STIIP .....	36
Figura 5-3 Diagrama de flujo de datos de primer nivel del STIIP .....	38
Figura 5-4 Esquema del metadato base para el STIIP .....	44
Figura 5-5 Diagrama de red del consejo sectorial y el usuario .....	46
Figura 5-6 DFD de los procesos de creación automática de metadatos y generación del índice.....	47
Figura 5-7 La araña y su telaraña .....	48
Figura 5-8 Algoritmo de creación automática de metadatos .....	49
Figura 5-9 División del volumen de almacenamiento de datos.....	53
Figura 5-10 Estructura de un almacén periférico .....	54
Figura 5-11 Ejemplo de esquema de formulario de ingreso de dato - metadato .....	59
Figura 5-12 Algoritmo de ingreso datos y metadatos componente humano .....	61
Figura 5-13 DFD del proceso de búsqueda del STIIP .....	63
Figura 5-14 Proceso geoportal del STIIP .....	66
Figura 5-15 Algoritmo de espacialización o visualización de la información recuperada en un mapa .....	67
Figura 5-16 Proceso de conversión de resultados de búsqueda en capas de mapas dinámicos .....	69
Figura 5-17 DFD del proceso de modelado y simulaciones espaciales .....	70
Figura 5-18 DFD del proceso de gestión y análisis de indicadores.....	72
Figura 5-19 DFD del proceso de Patrimonio participativo .....	74
Figura 6-1 Elementos del sistema cartográfico base del STIIP .....	80
Figura 6-2 Correspondencia entre campos de metadatos y componentes de la cartografía base del STIIP .....	85
Tabla 6-1 Dominios de indentidad y campos del metadato del STIIP .....	87

Figura 6-3 Aplicación del softwares en los procesos del STIIP .....	90
Figura 6-4 Arquitectura integrada de software .....	98
Figura 13-1 Pantalla de SIPE .....	136
Figura 13-2 Pantalla del Portal de promoción de la cultura de España .....	137
Figura 13-3 Pantalla de la búsqueda de "riobamba" en Google .....	138
Figura 13-4 Pantalla del Geoportal STIIP actual en la página web del MCP .....	139
Figura 13-5 Estadísticas del índice de muestra usado en el STIIP-Piloto .....	148
Figura 13-6 Mapa del sitio .....	149
Figura 13-7 Página Inicio .....	149
Figura 13-8 Pantalla de la búsqueda por palabra clave "riobamba" .....	150
Figura 13-9 Ejemplo de formato de presentación de archivos generales en los resultados de una búsqueda .....	151
Figura 13-10 Ejemplo de formato de presentación de bases de datos como archivos del índice en el resultado de una búsqueda .....	152
Figura 13-11 Ejemplo de formato de presentación de hojas de cálculo como archivos del índice en el resultado de una búsqueda .....	153
Figura 13-12 Pantalla de la búsqueda específica de "Museos municipales" .....	154
Figura 13-13 Pantalla de búsqueda avanzada o por categoría .....	155
Figura 13-14 Pantalla de búsqueda de <i>bosque alfaró</i> y el árbol de categorías .....	157
Figura 13-15 Pantalla de la búsqueda de <i>azuay quechua</i> y el árbol de categorías (Vista de bases de datos y Archivos CSV) .....	158
Figura 13-16 Pantalla del visualizador de resultados de la palabra clave <i>orellana</i> .....	160
Figura 13-17 Pantalla de la herramienta para indexar directorios completos .....	161
Figura 13-18 Pantalla de la herramienta para añadir los datos de conexión a una base de datos .....	164
Figura 13-19 Pantalla de creación de vistas o máscaras en una base de datos .....	165
Figura 13-20 Pantalla de la herramienta para añadir un archivo de formato xls o csv .....	167
Figura 13-21 Pantalla de configuración de indexación de una hoja de cálculo .....	168

# 1 INTRODUCCIÓN

## 1.1 Antecedentes del STIIP

El Decreto de Emergencia del Patrimonio Cultural, fue emitido por el Presidente de la República, Rafael Correa Delgado, el 21 de diciembre del 2007 (No. 816), a partir del robo de la Custodia de Riobamba. Una parte de este proyecto se designó para el registro de bienes culturales, procediéndose a realizar un levantamiento del inventario a nivel nacional en cinco áreas específicas lo que permitió registrar más de 80 mil bienes, cifra importante, pero no suficiente, pues se estima que el Ecuador posee más de dos millones de bienes culturales.

Con este antecedente, uno de los aspectos de la gestión institucional en el Ministerio Coordinador de Patrimonio (MCP) es la forma o manera de identificar, seleccionar, almacenar, priorizar, monitorear y evaluar políticas y resultados del sector e identificar las unidades encargadas de conseguirlos; es por esta razón, que se requiere generar un sistema que permita bajar de los niveles de abstracción que es la definición de una política y operar en espacios territoriales más concretos y delimitados, sin perder de vista en el proceso la articulación de resultados a dichas políticas.

La dificultad que existe dentro de la administración del patrimonio de todo el país gestionada por el MCP, es la integración de datos o la comunicación entre todas las entidades a las cuales coordina. Los ministerios o instituciones bajo la coordinación del MCP son:

- Ministerio de Cultura
- Ministerio del Deporte
- Ministerio del Ambiente
- Consejo de Gobierno de Galápagos
- Ferrocarriles del Ecuador
- Instituto Nacional de Patrimonio Cultural
- Ciudad Alfarero
- Diálogo Intercultural
- Dirección Nacional de Salud de los Pueblos Indígenas

Además, el MCP en su calidad de Institución Coordinadora, trabaja de manera conjunta con las siguientes instituciones:

- Secretaría Nacional de Planificación y Desarrollo (SENPLADES)
- Instituto Geográfico Militar (IGM)
- Instituto Nacional de Estadísticas y Censos (INEC), cuyas demandas serán incorporadas a través de mecanismos de coordinación definidos por el MCP.

En este sentido, la recopilación, búsqueda y presentación estructurada de todos los bienes patrimoniales, tanto del MCP como de las entidades a las cuales coordina, es de vital importancia para la toma de decisiones efectivas, consensuadas y apegadas a la ley sobre aspectos del patrimonio ecuatoriano. Para ilustrar lo mencionado hasta ahora se presenta un ejemplo práctico: en un cantón X del Ecuador, se cuenta con patrimonios ancestrales incas, además tiene asentamientos de poblaciones indígenas y mestizas, el cantón es también una

concentración de pesistas de Ecuador y por último es una zona natural protegida. En este momento, el MCP debe comunicarse con cada una de las entidades responsables para conocer la existencia de todo el patrimonio de una zona del país; en el ejemplo propuesto, debería recopilar datos del Ministerio de Cultura, Ministerio del Deporte, Ministerio del Ambiente, etc., unificarlos en un reporte final y en base a esto tomar una simple decisión, como podría ser la colocación de una estatua conmemorativa sin afectar el patrimonio de ningún tipo. Esta tarea se realizará suponiendo que el MCP tiene un método eficaz de recopilar todos los datos relevantes para la mejor toma de decisión que incluya a todos los sectores patrimoniales afectados. La inexistencia de este método eficaz en el MCP es el problema que se pretende resolver con el presente proyecto de tesis, al cual se lo ha denominado Sistema Territorial de Información Integrada de los Patrimonios (STIIP), el cual a su vez nace de una consultoría de contratación directa del MCP en el Portal de Compras Públicas del Ecuador.

## **1.2 Objetivo general**

El objetivo del proyecto STIIP, es en primera instancia diseñar una metodología que servirá como base para la construcción de instrumentos apropiados, que permitan al MCP integrar la información que las instituciones coordinadas generan, a partir de un proceso de agrupación de los indicadores de acuerdo a las prioridades resultantes de las políticas del Plan Nacional y según sus características comunes. La implementación de dichos instrumentos en el esquema institucional del sector público, en las entidades e instituciones involucradas, implica trabajar revisar y replantear fuentes de información, trabajar en el proceso de recolección y análisis de datos. Por otro lado, durante todo el proceso, se

realizará un trabajo continuo de transferencia de conocimientos a los responsables del Ministerio.

El alcance y solución al problema antes mencionado del proyecto STIIP se divide en dos aspectos. Primero es el **diseño conceptual** para la integración de datos de todo tipo de bienes patrimoniales que existan en el país, provenientes de fuentes diversas, como son todas las entidades coordinadas por el MCP. Dicho diseño conceptual, servirá de base para plasmar el siguiente producto deseado por el MCP: la creación de un geoportal o un Sistema de Información Georreferenciada (GIS, por sus siglas en inglés) de los bienes patrimoniales del país. El segundo aspecto de la consultoría es la creación de un **prototipo de software**, llamado STIIP-Piloto, que sea reflejo del diseño conceptual a elaborarse. Por su propio nombre, el prototipo de software contará con funcionalidades básicas (no desarrolladas en su totalidad) que muestren el funcionamiento de un geoportal acorde a las necesidades del MCP. Este además, servirá para la difusión y promoción del proyecto para una subsiguiente consultoría dentro del MCP, especialmente para tener un producto más palpable que muestre las capacidades y beneficios de un geoportal para personas no técnicas que auspiciarían la implementación del STIIP en su totalidad (subsecretarios, ministros, etc.). Esta hipotética subsiguiente consultoría futura, que está fuera del alcance del presente proyecto de tesis y de la primera consultoría ya contratada, dependerá de los recursos económicos e incentivos políticos del MCP y, supondría si se llegara a dar, la ejecución efectiva de los cambios de procesos en todas las entidades involucradas de acuerdo al diseño conceptual del proyecto STIIP y el desarrollo completo del geoportal para su puesta en producción. Dicho todo esto y delimitado el alcance, el objetivo general del presente proyecto de tesis es:

**Diseñar y desarrollar un prototipo de geoportal en el marco del diseño conceptual del proyecto *Sistema Territorial de Información Integrada de los Patrimonios (STIIP)*, del Ministerio Coordinador de Patrimonio del Ecuador.**

### **1.3 Objetivos específicos**

1. Analizar el contexto para la realización del proyecto STIIP, es decir, la situación de las instituciones coordinadas y sus requerimientos.
2. Generar el diseño conceptual del proyecto STIIP, que funcionará como marco teórico para el funcionamiento integrado de las instituciones coordinadas por el MCP y, además especificará el diseño del geoportal usando herramientas tecnológicas de código abierto o software libre, como establecen las políticas públicas del país con respecto al desarrollo del software.
3. Desarrollar el prototipo del geoportal, llamado STIIP-Piloto, que refleje la funcionalidad básica del geoportal y que servirá para la difusión y promoción del proyecto dentro del MCP.
4. Evaluación de rendimiento del prototipo desarrollado, elaboración del manual de usuario y recomendaciones para la continuación del desarrollo del software.

### **1.4 Proyectos relacionados del STIIP**

Previo a la realización del proyecto STIIP y su piloto, el Ministerio Coordinador de Patrimonio había desarrollado el Sistema Integrado de Información de los Patrimonios del Ecuador (SIPE), el cual mostraba datos en tablas, gráficos de barras y en mapas (a nivel

provincial únicamente) de la distribución geográfica de los bienes del patrimonio ecuatoriano (ver Anexo 2). En este sentido, el MCP propone el proyecto de construir un sistema con un grado de interactividad mayor, tanto para el uso del ciudadano común como para las autoridades para la toma de decisiones, planificación territorial, información, entre otros usos. Proyectos relacionados que actualmente estén en funcionamiento y sirvan como modelo de lo que se quiere lograr con el nuevo geoportal a desarrollar, se lo puede ver en la página del geoportal del patrimonio español (<http://geo.spainisculture.com/index.php/es>), en los que se cuenta con un buscador de bienes patrimoniales de España y se despliegan gráficamente los resultados en el mapa con toda la interactividad disponible con las nuevas tecnologías SIG (ver Anexo 3).

La tecnología de Sistemas de Información Geográfica (SIG), son servidos por bases de datos geográficas como lo son Google Maps, Bing Maps, OpenStreetMap, entre otras, que dan una idea de la interactividad que se puede alcanzar con un geoportal.

## **1.5 Justificación e importancia del STIIP**

Como justificación del proyecto, hay que mencionar en primera instancia que el proyecto de tesis que se plantea en este documento, surge como una contratación de consultoría directa del Ministerio Coordinador de Patrimonio del Ecuador (MCP) con el Centro de Transferencias y desarrollo de Tecnologías de la Universidad San Francisco de Quito (CTT-USFQ), bajo el sistema Portal de Compras Públicas del Ecuador, con código de proceso CDC-MCP-003-2010. Dicho proyecto fue realizado entre mayo y agosto del 2010, cuyo objetivo es sentar las bases (diseño conceptual y prototipo de software) para que un

futuro proyecto del mismo Ministerio contemple la implementación en su totalidad de la consultoría antes mencionada. Este nuevo proyecto se estimaba, a la fecha de finalización de la consultoría de la CTT-USFQ, que duraría aproximadamente dos años, repartidos entre la implementación completa del geoportal (en base al piloto realizado) y la recolección de datos georreferenciados de los bienes patrimoniales de todo el país.

Otro punto a tomar en cuenta es que bajo las políticas estatales que promueven el uso del software libre para el desarrollo tecnológico y el incentivo del ahorro del presupuesto estatal, la creación del geoportal se lo desarrollaría con este tipo de herramientas.

La importancia del proyecto se enmarca dentro del Decreto de Emergencia del Patrimonio Cultural, emitido por el Presidente de la República, Rafael Correa. Con lo cual se pretende el levantamiento del inventario a nivel nacional, de un estimado de dos millones de bienes culturales del Ecuador. "Pensar en Ecuador como Patria, nos remite a entender que la Patria es la suma de todos los patrimonios". El geoportal busca convertirse en el medio de difusión del patrimonio al ciudadano común. Esta información además permitiría la toma de decisiones, de autoridades nacionales como locales más eficaz, basada en información actualizada acorde a una planificación y a mejorar el ordenamiento territorial.

## **1.6 Centro de Transferencias y Tecnología de la USFQ**

El MCP dentro del Portal de Compras Públicas, realiza la contratación directa de consultoría (con código CDC-MCP-003-2010) en abril del 2010, con el objeto de contratación: "Consultoría – Diseño para desarrollar un Sistema Territorial de Información

Integrada de los Patrimonios (STIIP) como parte del sistema nacional de información que permita planificar y gestionar los territorios del país”. El Centro de Transferencias y Tecnologías de la Universidad San Francisco de Quito (CTT-USFQ), fue la seleccionada para la consultoría de desarrollo del STIIP, el cual es objeto del proyecto de tesis planteado en este documento. Cabe hacer mención especial al equipo de profesionales del CTT asignado a dicha consultoría: Carlos Mena, PhD. (director de proyecto); Francesco Pizzitutti, PhD. (consultor); Citlali Galindez (asistente) y finalmente mi persona, en el proyecto de consultoría me desempeñé como diseñador y desarrollador web, autor de este documento de tesis.

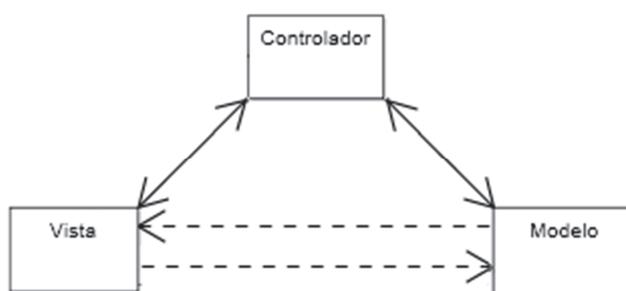
## 2 MARCO TEÓRICO

### 2.1 Definiciones generales

Lo primero que debe conceptualizarse es definir un Sistema de Información (SI). A un sistema se lo puede definir como un conjunto de varios elementos relacionados entre sí para cumplir un objetivo, en este caso el elemento es la información. Estos elementos deben estar enmarcados dentro de una estructura o el llamado diseño conceptual, en el cual se especificará la manera en que la información es extraída, transformada, almacenada, organizada y utilizada. La información o datos entonces, se trasladan por diferentes componentes que son necesarios definir:

- **Entidades externas:** personas ajenas al sistema informativo que entregan o reciben información. Las entidades externas pueden ser de distintos tipos y muchas veces se comunican entre ellas, pero esta forma de comunicación no necesariamente forma parte del sistema informativo mismo. Por ejemplo, entidades externas serían los usuarios que utilizan el sistema informativo, la ciudadanía en general, los que se encargan de su administración, la institución propietaria del sistema informativo, en este caso el MCP, que a su vez define exigencias, fines y provee informaciones acerca de datos y procesos preexistentes.
- **Almacenes de datos:** son los lugares físicos donde se almacenan datos. En ellos no se crean, destruyen ni se transforman datos; por lo tanto, la información guardada en los almacenes de datos es información semiprocesada, lo cual adelante se explica más en detalle. Cabe aclarar que los almacenes no se comunican directamente con entidades

externas, sino a través de un controlador, semejante al Modelo Vista Controlador (MVC). En la Figura 2-1 [1], se muestra en líneas punteadas las relaciones indirectas y en línea continua las directas. La ventaja de usar este modelo es que el funcionamiento de la aplicación es transparente al usuario, es decir, el usuario no conoce (ni le es útil conocer) las herramientas de software que se utilizan internamente para que el programa corra, sino que simplemente lo utiliza para obtener resultados.



**FIGURA 2-1 MODELO VISTA CONTROLADOR**

- **Procesos:** conjunto de actividades que explican cómo, cuándo y dónde se producen transformaciones en los datos que fluyen a través del sistema informativo.
  - Procesos externos: llamados interfaces o la parte Vista del MVC, en el que se define cómo el sistema presenta la información a los usuarios y recoge la interacción de los mismos.
  - Procesos internos: la parte Modelo del MVC, en donde se realiza la lógica del negocio y que los usuarios no pueden interactuar ni modificar.
- **Flujos de datos:** pasajes de datos de entidades a procesos, de procesos a almacenes de datos y viceversa.

En definitiva, un sistema de información transforma los datos crudos en información útil.

Para definir a estos dos elementos:

- **Dato:** “El dato es una representación simbólica (numérica, alfabética, algorítmica etc.), un atributo o una característica de una entidad. El dato no tiene valor semántico (sentido) en sí mismo, pero si recibe un tratamiento (procesamiento) apropiado, se puede utilizar en la realización de cálculos o toma de decisiones. Es de empleo muy común en el ámbito informático y, en general, prácticamente en cualquier disciplina científica.” [2]
- **Información:** “En sentido general, la información es un conjunto organizado de datos procesados, que constituyen un mensaje sobre un determinado ente o fenómeno. Los datos se perciben, se integran y generan la información necesaria para producir el conocimiento que es el que finalmente permite tomar decisiones para realizar las acciones cotidianas que aseguran la existencia.” [3]

Finalmente, hay que mencionar que el núcleo de un sistema de información, es el metadato. Los metadatos en su sentido más general se dice que son los “datos sobre los datos” que por ejemplo especifica el autor del contenido, fecha de publicación, tipo de contenido, entre otros. Existen varios estándares de metadatos que en capítulos posteriores se detallará a profundidad el usado para el STIIP.

## 2.2 Qué es un Sistema de Información Geográfica

Un Sistema de Información Geográfica (SIG), es un tipo de SI que almacena, maneja, analiza y presenta todo tipo de información georreferenciada [4]. Un SIG nos ayuda a interpretar rápidamente información visualizada en mapas. Los beneficios que nos brinda un SIG son múltiples y, en el caso específico del MCP se puede enumerar los siguientes:

- Ahorro de costos y aumento de eficiencia: esto se deriva propiamente del sistema de información, ya que se necesita una manera de integrar los datos de todas las instituciones coordinadas por el MCP, la georreferencia de los datos es un aspecto necesario para el correcto funcionamiento en cuestión de manejo del patrimonio.
- Mejora en la toma de decisiones: el SIG servirá para la toma de decisiones evitando la pérdida de información relevante. Cuando se tiene una gran cantidad de información, en este caso todos los datos de patrimonio del país, un SIG cuenta con las herramientas que facilitan la búsqueda y presentación de todos los datos relevantes para la toma de decisión más acertada o precisa.
- Mejora en la comunicación: aquellos usuarios de un SIG cuentan con las facilidades de intercambio de información en una interfaz más entendible. En el caso del MCP, todos los datos de patrimonio deben tener georreferencia y, lo más conveniente entonces sería presentar la información en mapas.
- Mejora el almacenamiento de registros: el diseño conceptual que se presta a realizar en el proyecto STIIP, definirá una estructura en la cual los datos de las diferentes entidades se ajusten a un marco común que facilita la recolección y el preprocesamiento para el almacenamiento.

- Administración geográfica: este es el objetivo final que se quiere lograr con el GIS. En el caso de administración de un país es imprescindible contar con herramientas para tratar información dependiendo de un espacio geográfico del país.

### 3 METODOLOGÍA

Debido a que el proyecto de consultoría STIIP generará un sistema informativo que integra a varias instituciones coordinadas del Ministerio Coordinador de Patrimonio, la mejor manera de desarrollar dicho producto es bajo el modelo de Desarrollo Rápido de Aplicaciones (RAD, por sus siglas en inglés).

El RAD facilita la construcción rápida de prototipos, mediante un proceso iterativo entre las fases de análisis, diseño e implementación del software, hasta lograr un prototipo final que será llamado la implementación o producto final, como se muestra en la Figura 3-1 [34]. La planificación de software y sistemas usando el RAD se intercala con la implementación del mismo software. La falta de una amplia planificación previa, por lo general, permite que el software se escriba de forma mucho más rápida y, hace que sea más fácil cambiar los requisitos iniciales en cualquier momento del proceso de desarrollo.

Debido al hecho de que se deben integrar varias instituciones, con procesos internos muchas veces muy diferentes, la comunicación sincronizada y provisión de información suficiente para el análisis del proyecto STIIP no se podrá realizar de manera eficiente. Por estos motivos se decide emplear la metodología RAD, ya que no se contará con la suficiente información previa para realizar una planificación sustanciosa. Las ventajas entonces de la creación de varios prototipos serán la utilidad y la rapidez de ejecución. En el presente documento, sin embargo, se detalla el análisis, diseño e implementación del producto final o último prototipo del proyecto STIIP.

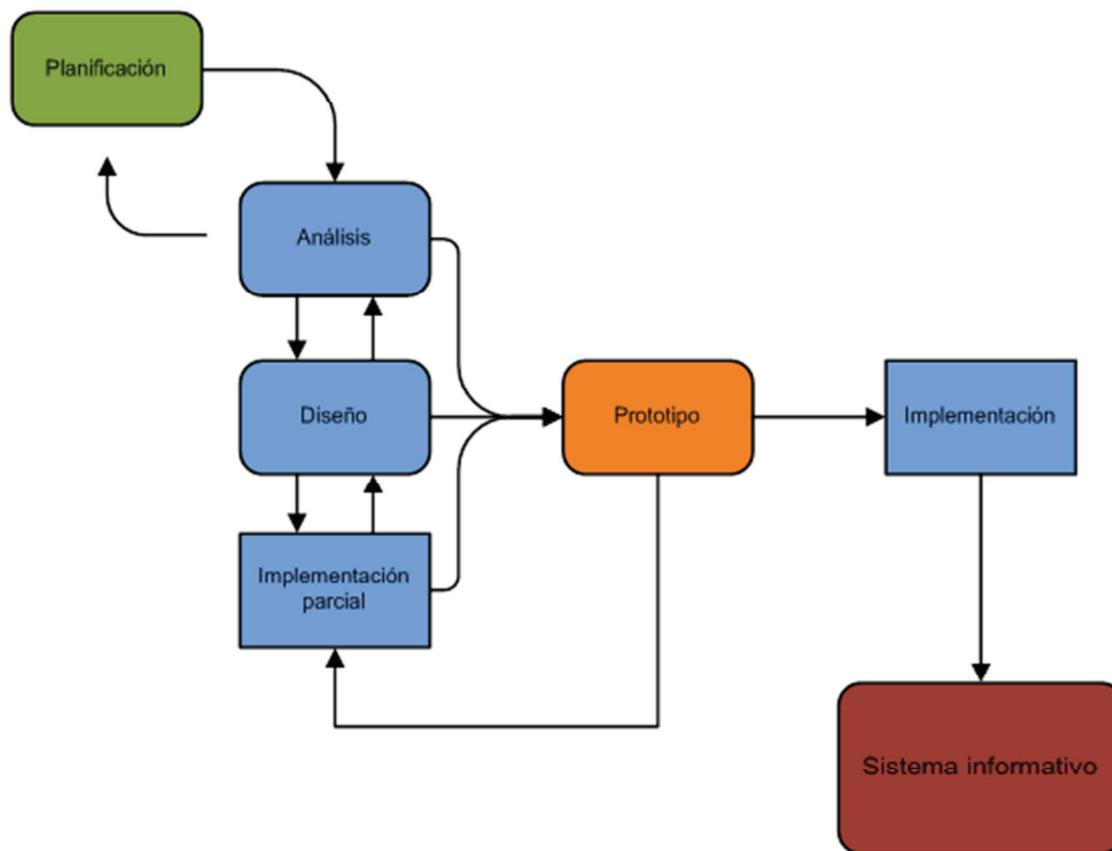


FIGURA 3-1 DESARROLLO RÁPIDO DE APLICACIONES

## 4 ANÁLISIS

### 4.1 Situación actual en el Consejo Sectorial de Patrimonio

El Consejo Sectorial de Patrimonio (CSP) lo conforman el mismo MCP más todas aquellas instituciones coordinadas por el MCP, más específicamente las siguientes:

- Ministerio de Cultura del Ecuador
- Ministerio del Deporte
- Ministerio del Ambiente
- Consejo de Gobierno de Galápagos
- Ferrocarriles del Ecuador
- Instituto Nacional de Patrimonio Cultural
- Ciudad Alfaro
- Diálogo Intercultural
- Dirección Nacional de Salud de los Pueblos Indígenas
- Ministerio Coordinador de Patrimonio

Todas estas instituciones presentan realidades humanas, organizativas y tecnológicas muy diferentes entre ellas. Además, los contenidos mismos y los tipos de datos manejados son de naturaleza diferente y tienen finalidades diferentes.

En aspectos generales, el manejo de datos en el CSP se encuentra en el siguiente estado:

1. Se cuenta con dos tipos de contenido acerca de patrimonio: digitales y físicos.
2. Con relación a los contenidos digitales se tiene archivos que además son muy diferentes en su formato: texto, bases de datos, multimedia (audio, video), formato de texto enriquecido (pdf, doc, ppt, docx, pptx), hojas de cálculo (xls,.xlsx), shape files, etc.; así como en su contenido: catálogos de especies, textos de interés cultural, acta de conferencias, reportes, indicadores, mapas.
3. En relación a los contenidos físicos, se puede decir que son aquellos que no son digitales o están en proceso de digitalización, es decir, aquellos que están solo en papel.
4. La ubicación física de los datos está muy fragmentada, lo cual es solucionable con un sistema de información. Tanto los contenidos físicos como los servidores que almacenan a los contenidos digitales se encuentran en las sedes centrales de cada institución como en sus entidades ejecutoras periféricas.
5. Las políticas de producción, análisis y distribución de los datos son muy diferentes, dependiendo de las prioridades de cada una de las instituciones. Al momento del inicio del proyecto STIIP, no se cuenta con ninguna estandarización de datos que contemple a todo el CSP, a pesar que se han hecho algunos intentos.
6. De acuerdo a las entrevistas con varios de los funcionarios que manejan la información en el MCP, la poca facilidad tecnológica de intercambio de información por lo mencionado hasta ahora, no promueve la rápida compartición de información entre instituciones del CSP, ya que se cuenta solo con procesos demasiado burocráticos (al interior de cada una) y poco eficientes para hacerlo. Todo esto, a pesar de que en niveles superiores del servicio público y legislativo, se estén creando estructuras organizativas que generen estándares para la comunicación entre diferentes

instituciones públicas. Si bien cada institución se maneja con su propio sistema interno de almacenamiento de información, no se cuenta con estándares para la integración de información interinstitucional.

7. Respecto a la transparencia y difusión de la información hacia la ciudadanía, en general, se puede indicar que está bastante desatendida, es poca la información relevante del patrimonio que se puede obtener desde los sitios web y es prácticamente nula la información integrada de patrimonio interministerial que cualquier ciudadano puede conseguir. El proyecto STIIP pretende resolver también este asunto.
8. La toma de decisiones en niveles medios o altos administrativos o políticos es bastante ineficiente e incluso sesgada, ya que resulta difícil el hecho de descubrir o acceder a todos los datos relevantes o necesarios para llegar a la mejor decisión posible en temas relacionados al patrimonio del país.

## **4.2 Instituciones coordinadas por el MCP**

Se realizará un análisis del estado de cada institución que se pudo acceder a información relevante para los propósitos de STIIP. La siguiente tabla muestra el resumen del estado del manejo de datos de las instituciones consultadas.

**TABLA 4-1 ESTADO DEL MANEJO DE DATOS EN INSTITUCIONES COORDINADAS DEL MCP**

<b>Institución</b>	<b>Probabilidad de fácil integración al STIIP (alta, media, baja)</b>	<b>Observaciones</b>
Ministerio de Cultura del Ecuador	Media	El <i>Sistema de Gestión y Seguimiento de Proyectos</i> , no se encuentra todavía en una fase operativa. Necesita de un SI previo a la integración al STIIP.
Ministerio del Ambiente	Alta	Se está planeando un sistema de gestión documental, se estima que estará operativo después de finalizar el STIIP.
Ministerio del Deporte	Media	El <i>Sistema Nacional del Deporte</i> está en desarrollo, el cual organizará todas las informaciones de este ministerio.
Ferrocarriles del Ecuador	Alta	Información accesible en el portal del <i>Sistema de Información para la Gobernabilidad Democrática (Sigob)</i> . Además se está desarrollando el SIG-TREN.
Instituto Nacional de Patrimonio Cultural	Alta	Publica mucha información en su sitio web y tiene un geoportal con servicios WMS. Su información está bien organizada.
Dirección Nacional de Salud de los Pueblos Indígenas	Alta	Gestión organizada de la información. Posible dificultad con las sedes provinciales que generan información en papel y no llega a la central rápidamente.
Ecorae	Alta	Implementación de un sistema similar al STIIP para la región amazónica.
DINEIB	Baja	Falta de copias de respaldo de informaciones. Parte de sus datos están administrados por el Ministerio de Educación.

#### 4.2.1 Ministerio de Cultura del Ecuador

El objetivo institucional del Ministerio de Cultura del Ecuador es “contribuir a la construcción de la identidad nacional desde las identidades diversas.” [5]

La situación a nivel de gestión de la información institucional relevante en el Ministerio de Cultura al presente no es muy definida. En el ministerio se han producido y se sigue produciendo muchas informaciones en forma de informes, libros, revistas, productos de programas y proyectos. Dentro del contenido en el sitio web se cuenta con una página de proyectos culturales en un mapa con las provincias del Ecuador que muy rudimentariamente podría ser considerado como manejo de patrimonio georreferenciado. Sin embargo, la información es escasa, sin estructura y de difícil uso por parte de los usuarios externos.

Internamente y contrariamente a los que se está desarrollando en otras instituciones del CSP, en el Ministerio de Cultura no se utiliza un sistema de gestión documental o de distribución de la información adecuado. De acuerdo a los funcionarios del ministerio, la única excepción es el *Sistema de Gestión y Seguimiento de Proyectos*, que no se encuentra todavía en una fase operativa.

Por la falta de estándares definidos, esta institución necesita de un SI estándar para los procesos de recolección de datos y posteriormente para la integración al STIIP.

#### **4.2.2 Ministerio del Ambiente**

El Ministerio del Ambiente “es el organismo del Estado ecuatoriano encargado de diseñar las políticas ambientales y coordinar las estrategias, los proyectos y programas para el cuidado de los ecosistemas y el aprovechamiento sostenible de los recursos naturales.” [6]

El Ministerio del Ambiente, a nivel de situación actual de manejo de información y gestión documental se encuentra en el mismo estado del Ministerio de Cultura. La diferencia es que en este caso, según los informantes entrevistados, se está planeando un proyecto de consultorías y procedimientos internos ya asignados y en fase de definición, el diseño y la implementación de un sistema de gestión documental, de recopilación de la información también en formato papel y de organización de la información espacial relevante. Por otro lado, el sitio web del ministerio no cuenta con ningún contenido de patrimonio para un usuario externo.

Si el proceso de recopilación de la información por el STIIP, se iniciará después de la realización de todos los sistemas de manejo de datos del Ministerio del Ambiente, la tarea debería resultar bastante sencilla ya que toda la información estaría disponible en una forma preprocesada.

#### **4.2.3 Ministerio del Deporte**

El Ministerio del Deporte se encarga de “desarrollar la Actividad Física y el Deporte Ecuatoriano mediante una gestión eficiente, integradora y transparente que priorice al ser humano [...] que define las políticas, los objetivos y las estrategias del sector.” [7]

Como el Ministerio del Ambiente, el Ministerio del Deporte está en el medio de un proceso de implementación de un sistema de gestión documental. También el Ministerio del Deporte se encuentra por finalizar los Términos de Referencia (TdR) para la realización de una encuesta nacional sobre el mundo deportista, las infraestructuras de deportes, las actividades gubernamentales deportivas, que servirá para la realización de una base de datos en que se irán a almacenar los datos recogidos. Estos datos irán a componer un sistema informático denominado *Sistema Nacional del Deporte*.

Si la fase de recopilación de información del STIIP, se realizará después de la actuación de dicha encuesta y de la implementación del sistema de información, esta tarea resultará más directa de integrar. Sin embargo, en el Ministerio del Deporte resulta ser un poco más incierto el cumplimiento de estos proyectos que en el caso del Ministerio del Ambiente.

#### **4.2.4 Ferrocarriles del Ecuador**

En la Empresa de Ferrocarriles de Ecuador, la información que se tiene actualmente no representa un volumen considerable comparado con instituciones como el Ministerio de Cultura o del Ambiente. El tiempo de tramitación, permisos y papeleos para la entrega de información patrimonial, solicitada a la empresa fue bastante largo, fue la última institución en entregar información al proyecto STIIP. La empresa reporta una cantidad considerable de material en físico y con respecto a la información sobre la gestión y los proyectos, la misma que se encuentra accesible en el portal del *Sistema de Información para la Gobernabilidad Democrática (Sigob)*.

La Gerencia de Planificación de la Empresa de Ferrocarriles del Ecuador tiene entre sus ejes estratégicos la programación, capacitación, implementación y desarrollo del proyecto – SIG TREN – (Sistema de Información Georreferenciado) y, como objetivo establecer los procedimientos para implementar un sistema que pueda sistematizar, organizar y difundir diversas capas temáticas actualizadas, complementadas y/o generadas por Ferrocarriles del Ecuador – Empresa Pública (FEEP). Adicionalmente, se tiene previsto ejecutar el Diseño, Desarrollo e Implementación de un Geoportal. Todo esto se encuentra en fases iniciales de definición al momento de empezar el proyecto STIIP.

#### **4.2.5 Instituto Nacional de Patrimonio Cultural – INPC**

El INPC tiene un Centro de Cómputo con el MCP utilizado para fines externos, publica mucha información en su sitio web y tiene un geoportal con servicios WMS. También el INPC tiene un servicio de denuncia ciudadana, el cual es un proceso similar a lo que se planea en el Proceso de Patrimonio participativo del STIIP. En general el estado de la información del INPC parece bien ordenado pero con algunas fallas en la fase de difusión y acceso a la información. El geoportal utilizado por ellos tiene ciertas fallas de interactividad o visualización, pero es el único geoportal encontrado en el CSP, el cual puede servir como guía.

#### **4.2.6 Dirección Nacional de Salud de los Pueblos Indígenas**

En el caso de esta institución, se observa como en los otros casos, la dispersión de la información, pero a diferencia de otras instituciones de mayor tamaño. La Dirección de Salud Intercultural, que es relativamente pequeña, se encuentra bien organizada por quienes gestionan la información, razón por la cual integrar la información no sería demasiado complicado. Lo que sí puede generar algunas dificultades es el hecho de que parte de la información es generada por las sedes provinciales y lo que llega a Planta Central es información procesada; es decir, resultados que se solicitan y muchas veces en formato papel. Es probable que la información generada en las sedes provinciales sea de interés para la difusión que se está proponiendo el STIIP.

#### **4.2.7 Otras instituciones**

##### *4.2.7.1 Ecorae*

El Instituto para el Ecodesarrollo Regional Amazónico (Ecorae), está trabajando para la implementación de un sistema muy similar al STIIP pero con información particularmente de la región amazónica. Para dicho proyecto ya tendrían aprobados los términos de referencia. Están coordinando este esfuerzo con la SENPLADES y Patrimonio, que tiene el objetivo de optimizar los servicios web. El proyecto está todavía en fase de primera definición.

#### 4.2.7.2 Dirección de Educación Intercultural Bilingüe – DINEIB

La situación del manejo de la información en la DINEIB no es adecuada, sobre todo por la falta total de copias de respaldo de informaciones que se consideran como muy valiosas. Sin embargo, una parte de los datos manejados por esta dirección, se presenta sistematizada en una base de datos denominada Archivo Maestro, que es administrada a nivel central por el Ministerio de Educación. Por el resto, la información se presenta almacenada en computadoras de los encargados de la Dirección y en sistemas archivos en sistemas periféricos. Por lo tanto, debido al pequeño tamaño de la institución no se prevé un volumen de trabajo muy grande en vista de la recopilación de los datos de la DINEIB.

### 4.3 Requerimientos del Consejo Sectorial de Patrimonio

En base a lo propuesto y discutido en las reuniones con el MCP, se han definido los límites y características generales que el futuro STIIP requerirá:

1. Ser un SI distribuido en Internet e Intranet.
2. Ser un sistema que pueda distribuir a varios niveles toda la información relevante producida en todas las instituciones coordinadas.
3. Ser un SI territorial: almacenar y organizar la información a través de criterios geográficos y espaciales.
4. Guardar compatibilidad con las directivas públicas en manejo, almacenamiento y distribución de datos e informaciones, es decir, cumplir un estándar definido.

5. Ser compatible con otros SI operantes en el país, en particular con el Sistema de Información Nacional (SNI) de la SENPLADES.
6. Ser un sistema lo bastante simple para poder funcionar de manera semiautomática y con pocas necesidades de manutención.
7. Tener tres posibles niveles de usuarios: tomadores de decisiones, técnicos y ciudadanía.
8. Ser capaz de poder acoger las diversas exigencias informativas de los diferentes usuarios.
9. Ser una estructura base para la organización, producción, almacenamiento y distribución para todas las instituciones coordinadas.
10. Tener una utilización simple, intuitiva y seguir estándares internacionales del manejo de datos (definidos en la fase de diseño).
11. Dar acceso directo, en los casos oportunos, a los datos crudos almacenados por las instituciones coordinadas.
12. Utilizar las nuevas tecnologías en el campo informático y de software libre.

Después de un análisis de los requerimientos, se ha definido que las posibles maneras de desarrollar un sistema de información que corresponda a las características detalladas anteriormente son dos:

- Un sistema de información sustentado en una base de datos (como base de dato relacional que es el modelo de base de datos más utilizado en este campo).

- Un sistema informativo basado en índice.

Los dos tipos de sistemas presentan ventajas y desventajas, según la situación que se tenga que enfrentar. A continuación, se analizará algunos detalles del entorno en que se implementará el STIIP. Se estima que, sin considerar las bases de datos, el número de documentos actualmente almacenados en el Consejo Sectorial por las instituciones coordinadas es del orden de magnitud de las decenas de millones. Esto implica que, considerando la exigencia de distribuir y publicar la totalidad de los documentos presentes, un acercamiento al problema sustentado en base de datos relacionales, simplemente no es el más adecuado. Es notorio que los motores de base de datos disponibles no presentan un buen comportamiento en cuanto a escalabilidad con la cantidad de información almacenada, lo cual significa que presentándose la necesidad de buscar un registro en todas las tablas de una base de datos de millones de registros, la tarea podría demandar minutos de latencia, afectando gravemente al rendimiento del sistema. Es principalmente por este motivo que el STIIP ha sido pensado en términos de un motor de búsqueda indexado.

Un motor de búsqueda es una arquitectura integrada de software que permite construir un índice de todos los archivos de un sistema: documentos de texto, bases de datos, multimedia, etc. El índice propuesto almacenaría ordenadamente cierta información acerca de los documentos que permita un eficiente proceso de búsqueda de los mismos. Los principales motores de búsqueda escalan muy bien con el número de documentos indicados y permiten encontrar una respuesta a una búsqueda entre millones de documentos en tiempos del orden de décimas de segundo [8], incluido la presentación de los resultados en

la pantalla. Por este motivo, el escenario de desarrollo e implementación adoptado por el STIIP será el modelo de sistema informativo basado en índice.

## **4.4 Sistema de gestión documental**

A continuación se presenta un análisis de cómo el sistema informativo a desarrollarse manejará las informaciones en un contexto de un Sistema de Gestión Documental (SGD).

### **4.4.1 ¿Qué es un sistema de gestión documental?**

“Se entiende por gestión documental el conjunto de normas técnicas y prácticas usadas para administrar el flujo de documentos de todo tipo en una organización, permitir la recuperación de información desde ellos, determinar el tiempo que los documentos deben guardarse, eliminar los que ya no sirven y asegurar la conservación indefinida de los documentos más valiosos, aplicando principios de racionalización y economía.” [9]

En general, los problemas que se quieren enfrentar desarrollando un sistema de gestión documental se resumen en la siguiente tabla:

TABLA 4-2 ÁREAS DE UN SISTEMA DE GESTIÓN DOCUMENTAL

<b>Almacenamiento</b>	¿Dónde guardaremos nuestros documentos?
<b>Recuperación</b>	¿Cómo se puede encontrar los documentos necesarios? ¿Cuánto tiempo se puede pasar buscándolo?
<b>Clasificación</b>	¿Cómo organizamos nuestros documentos? ¿Cómo aseguramos que los documentos estén archivados siguiendo el sistema más apropiado?
<b>Seguridad</b>	¿Cómo evitamos la pérdida de documentos, evitar la violación de la información o la destrucción no deseada de documentos?
<b>Custodia</b>	¿Cómo decidimos qué documentos conservar? ¿Por cuánto tiempo deben ser guardados? ¿Cómo procedemos a su eliminación (expurgo de documentos)?
<b>Distribución</b>	¿Cómo distribuimos documentos a la gente que la necesita?
<b>Flujo de trabajo</b>	¿Si los documentos necesitan pasar a partir de una persona a otra, cuáles son las reglas para el flujo de estos documentos?
<b>Creación</b>	¿Si más de una persona está implicada en creación o modificación de un documento, cómo se podrá colaborar en esas tareas?
<b>Autenticación</b>	¿Cómo proporcionamos los requisitos necesarios para la validación legal al gobierno y a la industria privada acerca de la originalidad de los documentos y cumplimos sus estándares para la autenticación?

#### 4.4.2 Sistema de Gestión Documental aplicado al STIIP

En el STIIP no todos los requisitos presentados en la tabla anterior serán implementados. Por ejemplo el requerimiento de tener un control en el flujo de trabajo de un documento no es parte de las funciones del STIIP, porque el control de las modificaciones a documentos es una tarea que debe realizarse en cada una de las instituciones coordinadas y no en el sistema central STIIP. Lo que interesa entonces, desde el punto de vista del análisis de un componente de gestión documental, es definir la manera en que los documentos, independientemente de sus contenidos son generados, asociados, almacenados, eventualmente actualizados y borrados.

De todas las características funcionales de un sistema de gestión documental, listadas en la tabla anterior, serán analizadas las necesarias para cumplir con las necesidades del MCP. Las características funcionales de un sistema de gestión documental que si serán implementadas en el STIIP son: almacenamiento, recuperación, clasificación, seguridad y distribución. En particular, se debe contestar las siguientes preguntas:

1. ¿Cómo se organiza los datos?
2. ¿Cómo se crea, guarda y actualiza los documentos?
3. ¿Cómo se evita la pérdida de documentos?
4. ¿Cómo se evita la violación de la información reservada?
5. ¿Cómo se decide qué documentos conservar?

#### 4.4.3 Escenarios posibles

Una posible respuesta a las preguntas propuestas en la sección precedente, podría ser implementar dentro del STIIP un software de gestión documental de código abierto presente en el mercado. Muchos son los programas de este tipo disponible en fase de desarrollo avanzado. Uno muy utilizado y difundido dentro de algunas de las implementaciones de varias de las instituciones coordinadas es Alfresco. Las ventajas de elegir una solución similar podrían ser:

- Tener un software de alta fiabilidad.
- Tener muchas funciones predefinidas.

Las desventajas serían:

- Necesidad de instalación y mantenimiento del software en los servidores y/o computadoras en red de todas las instituciones coordinadas.
- Capacitación de personal para manejar el software en cada institución coordinada.

Un escenario hipotético sería la selección de Alfresco para desarrollar el SGD para el STIIP. Alfresco cuenta con muchas de las funcionalidades requeridas para el manejo de documentos. Entre las funciones de Alfresco que serían útiles al STIIP se encuentran: interfaz de administración ya desarrollada, control automático de versiones (lo cual serviría para documentos variables del CSP especificados más adelante), niveles de seguridad y asignación de diferentes roles o usuarios.

La principal razón de no utilizar un SGD del mercado, por ejemplo Alfresco, es que el proyecto STIIP necesita de varias funcionalidades específicas para el CSP, en especial aquellas relacionadas a georreferencia; entonces, se decidió desarrollar un sistema de gestión documental interno en el STIIP, denominado SGD-STIIP. En este caso algunas de las ventajas son:

- Se diseña y se implementa según las necesidades específicas del STIIP.
- Se integra de manera nativa con el resto de la arquitectura del STIIP.
- Se puede extender fácilmente.

Las desventajas, sin embargo, de desarrollar un sistema nuevo serán:

- Trabajo de diseño, desarrollo y pruebas.
- No presenta todas las funciones predefinidas de un software ya desarrollado.
- Mantenimiento del software.
- Propiedad intelectual.

## 5 DISEÑO CONCEPTUAL DEL STIIP

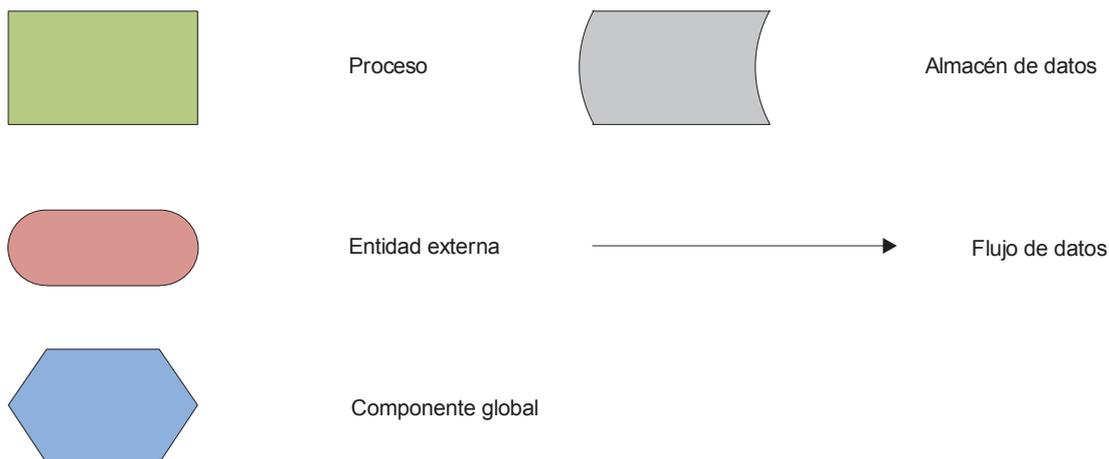
### 5.1 Diagramas de Flujo de Datos del STIIP

“Un Diagrama de Flujo de Datos (DFD) es una representación gráfica para la faceta del *flujo* de datos a través de un sistema de información. Los diagramas de flujo de datos pueden ser usados para proporcionar al usuario final una idea física de cómo resultarán los datos a última instancia, y cómo tienen un efecto sobre la estructura de todo el sistema.”  
[10]

En el diagrama de flujo de datos se describe el movimiento de los datos a través del sistema, especificando los siguientes componentes:

- Los lugares de origen y destino de los datos (los límites del sistema).
- Las transformaciones a las que son sometidos los datos (los procesos internos).
- Los lugares en los que se almacenan los datos dentro del sistema (depósitos).
- Los canales por donde circulan los datos (flujos).

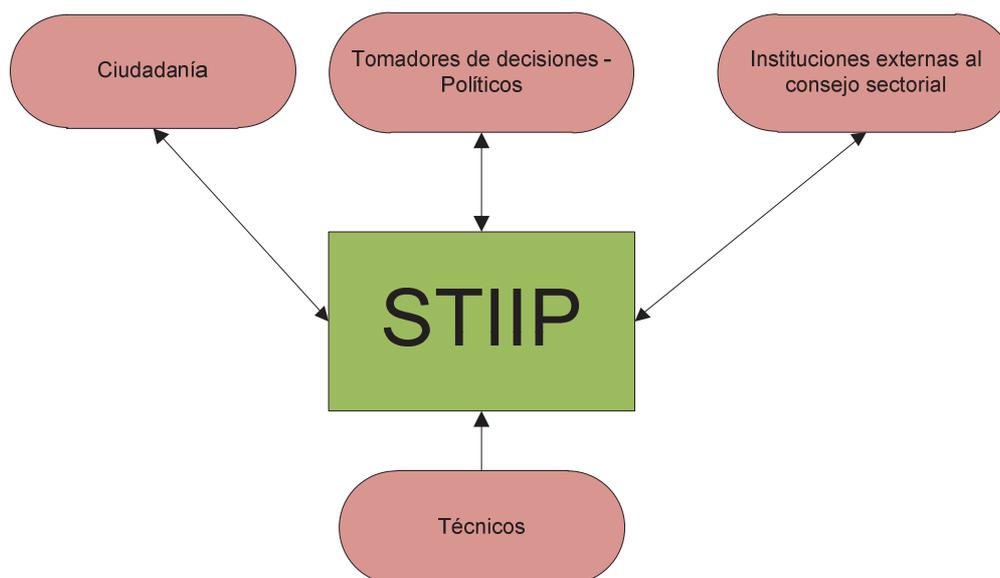
En el DFD cada componente es representado a través de un símbolo. En este documento utilizaremos la siguiente convención de simbología para los DFD [34]:



**FIGURA 5-1 SÍMBOLOGÍA DE LOS DIAGRAMAS DE FLUJO DE DATOS (DFD) DEL STIP**

## 5.2 Diagrama de contexto

Es preciso definir y dibujar primero un diagrama de contexto que muestra la interacción más general o flujos de datos, entre el sistema de información y las entidades externas. Luego, este primer diagrama de contexto se utiliza para mostrar más detalles del sistema que se está modelando en los siguientes DFD de niveles interiores. El diagrama de contexto del STIP es representado en la Figura 5-2 [basado en 34].



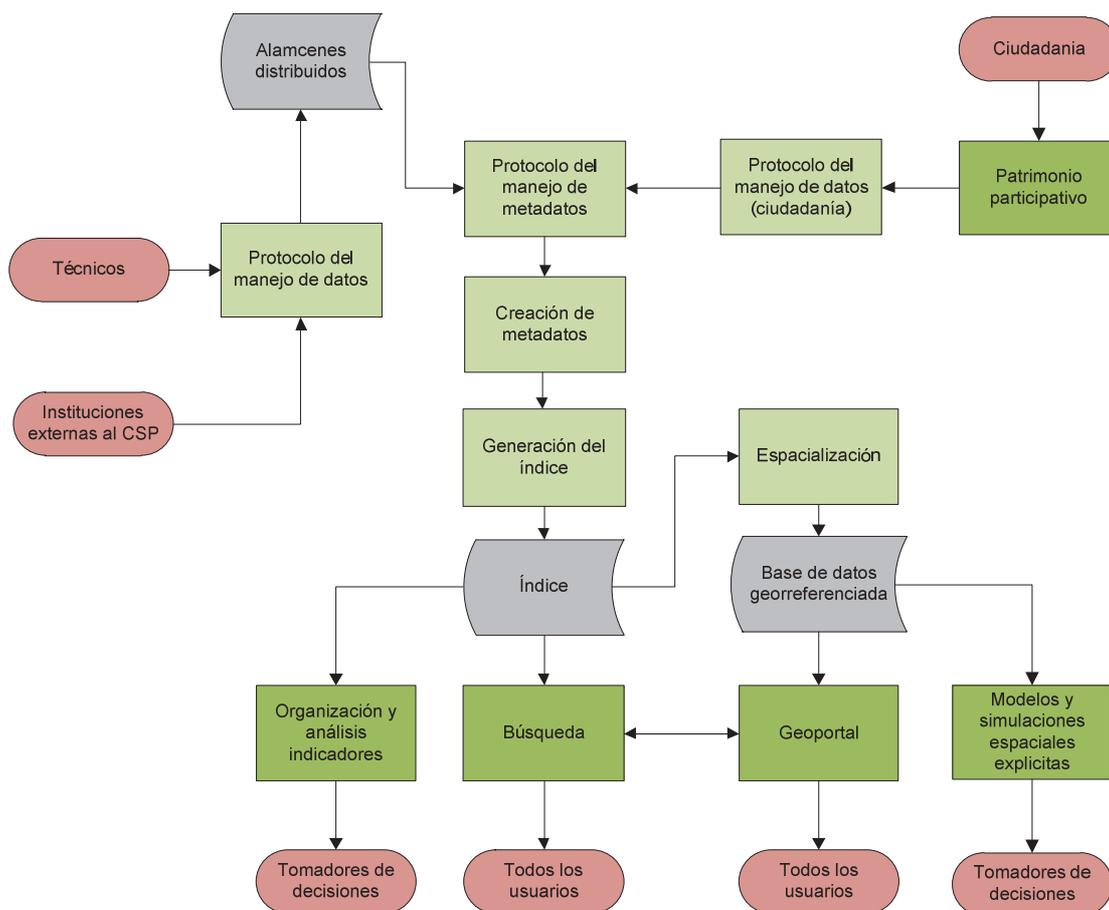
**FIGURA 5-2 DIAGRAMA DE CONTEXTO DEL STIIP**

Como se observa en la figura 5-2, los flujos de datos e informaciones principales son definidos como intercambios entre el sistema y tres tipologías distintas de entidades externas: la ciudadanía que interactúa con el sistema de información para conseguir datos y entonces genera exclusivamente flujos de datos de salida al sistema informativo aunque en algunos procesos y en forma limitada se permitirá el ingreso de datos también a la ciudadanía (ver Proceso de Patrimonio participativo). El segundo tipo de usuarios son los especialistas y personal técnico de las instituciones coordinadas y del MCP, que interactúan con el sistema de forma “profesional” con fines de planificación, de manejo y de producción de información. Este tipo de usuarios ingresarán y elaborarán los datos e información del STIIP, tendrán privilegios para visualizar toda la información con fines de monitoreo y mantenimiento. La tercera clase de entidad externa al STIIP está constituida por los tomadores de decisiones internos y externos al consejo sectorial o por los usuarios definidos como “políticos”. Estos últimos en general solo extraen la información procesada para la toma de decisiones por lo que su flujo de datos es de salida. Sin embargo, también

deben tener privilegios para visualizar y modificar todos los datos, pero sus fines serán determinar las prioridades estratégicas informativas. Una última clase de entidad externa al STIIP está constituida por instituciones externas al consejo sectorial (como por ejemplo la SENPLADES) que generan flujos de entrada y salida al STIIP. Por lo tanto, las características del STIIP deberán ser calibradas para satisfacer las exigencias de las cuatros entidades externas indicadas.

### **5.3 Diagrama de nivel superior**

El paso siguiente en la edificación del sistema de información, es diseñar los diagramas de flujo de datos del nivel superior, donde se especifica los distintos subsistemas funcionales en el que se va a descomponer el sistema de información del STIIP.



**FIGURA 5-3 DIAGRAMA DE FLUJO DE DATOS DE PRIMER NIVEL DEL STIP**

Desde el punto de vista tecnológico funcional, el STIP se ha conceptualizado como el conjunto de los siguientes componentes:

1. Almacenes distribuidos: son los almacenes de datos localizados en varias instituciones del consejo sectorial.
2. Índice: es el almacén (central) de los metadatos. Este almacén corresponde al repositorio del MCP, donde se almacenan datos de interés general y también una copia de parte de los datos contenidos en los almacenes distribuidos.

3. Base de datos (central) georreferenciada: contiene la información de coordenadas geográficas de los territorios del país.
4. Procesos externos (verde oscuro en el diagrama): búsqueda, geoportal, modelos y simulaciones espaciales explícitas, organización y análisis de indicadores, y patrimonio participativo. Estos procesos son llamados externos porque corresponden a funciones del STIIP visibles a las entidades externas.
5. Procesos internos (verde claro en el diagrama): generación del índice, creación de metadatos, protocolo del manejo de datos, protocolo del manejo de metadatos y espacialización.

En las siguientes secciones se detallará una descripción de los procesos mostrados en la Figura 5-3 [basado en 34], de tal manera de mostrar el diseño planeado para el funcionamiento del STIIP.

#### **5.4 Organización de los almacenes de datos**

Como hemos visto en los numerales precedentes, el STIIP será el sistema de información de una red de instituciones, cada una de las cuales presenta su propio sistema de almacenes de datos y, en algunos casos también un sistema de gestión documental propio. Además cada una de las instituciones coordinadas maneja volúmenes de datos muy grandes; por este motivo, no es recomendable en la fase de diseño del STIIP prever una estrategia de organización de los almacenes de datos, basada en un proceso de copia de cualquier tipo de datos situados en los almacenes periféricos de las instituciones y de acumulación de estas

copias en un almacén central organizado según criterios estándares. Siguiendo esta estrategia, el tamaño total de los datos que se deberían copiar y el trabajo que se realizaría para actualizar el sistema del almacén central, implica un esfuerzo demasiado grande no solo en términos de organización y sistematización, sino también de la infraestructura que se necesitaría (gran tamaño de los almacenes centrales y de ancho de banda de conexión entre servidores).

Una orientación más sencilla y también menos onerosa en términos de tamaño del almacén central, es la de no mover la mayoría de los datos y dejarlos en los almacenes de cada una de las instituciones coordinadas, pidiendo al mismo tiempo a las instituciones definir de manera definitiva una colocación de cada documento, que servirá para el correcto y eficiente funcionamiento del STIIP. Las instituciones coordinadas deberán definir un sistema de almacenes lo más constante posible en el tiempo y comunicar a la administración del STIIP cada cambio de la organización de sistemas de almacenamiento. Por supuesto, las instituciones no deberán indicar la dirección física de cada una de las parcelas de datos que tiene almacenados, pero dar una dirección del sistema físico (servidores, computadoras) dentro del cual estén almacenados los datos, por ejemplo el sistema de archivos o ficheros usado en un determinado servidor periférico al alcance del STIIP. Será después incumbencia de los algoritmos del STIIP el escanear cada almacén indicado por las instituciones y construir automáticamente la lista de todos los archivos en las direcciones definidas. La manera de construir lo dicho hasta aquí se lo especifica en el diseño de los Procesos de creación del metadato y generación del índice como se detalla a continuación.

## 5.5 Procesos de creación del metadato y generación del índice

El núcleo del STIIP es la creación de los metadatos y utilización de estos en la generación del índice. Antes de seguir con la descripción de estos dos procesos es necesario explicar que son los metadatos y el índice.

### 5.5.1 El metadato

“Metadatos: (del griego *μετα*, *meta*, «después de» y latín *datum*, «lo que se da», «dato»), literalmente «sobre datos», son datos que describen otros datos. En general, un grupo de metadatos se refiere a un grupo de datos, llamado *recurso*. El concepto de metadatos es análogo al uso de índices para localizar objetos en vez de datos. Por ejemplo, en una biblioteca se usan fichas que especifican autores, títulos, casas editoriales y lugares para buscar libros. Así, los metadatos ayudan a ubicar datos.” [11]

Esta definición corresponde también al uso que se va a dar al metadato en el STIIP. Formalmente el metadato es una lista de parejas campo–valor. Cada campo puede tomar más de un valor y los valores asociados a cada campo son informaciones a propósito del dato asociado al metadato. Es claro que no es posible definir un estándar único de metadato que dé una descripción completa para todos los tipos de datos, por ejemplo, un metadato relativo a una sinfonía presentará sin duda el campo “fecha de composición”, mientras un metadato relativo a una especie animal presentará el campo “familia” pero no el campo “fecha de composición”. Sin embargo, hay algunos campos comunes a todo tipo de datos: según los estándares internacionales adoptados también en Ecuador, cualquier dato debe ser relacionado a un metadato que contenga por lo menos el *Dublin Core* [12]. En el

Dublin Core hay campos relativos a nombre, fecha de edición, autor, formatos, idioma, derechos, etc. que todo tipo de datos debería presentar. A partir de este núcleo base se puede asociar al metadato un número cualquiera de campos, hasta a definir un *estándar* adecuado para el tipo de datos que se va a manejar.

Por lo que concierne el STIIP, después de definir un estándar base de campos necesarios para el CSP, irá siendo modificado de acuerdo a las necesidades. El STIIP se imaginará siempre como un sistema modular y flexible capaz de englobar constantemente nuevos elementos funcionales y nuevos tipos de contenidos. El esquema del metadato base es representado en la Figura 5-4 [34].

Es claro, que la naturaleza territorial del STIIP implica que una parte importante y obligatoria de campos del estándar del metadato utilizado será acerca de atributos geográficos. Habrá campos de coordenadas geográficas y campos espaciales de atributos geográficos como provincia, cantón, parroquia, vías, parque nacional, etc.

Otros campos presentes en el estándar STIIP serán los sectoriales que especifican el sector al cual pertenece el dato, así como los campos de actores que especificaran quien se encargó de recopilar y modificar el dato y otros campos. Este tipo de metadato se debe pensar como un modelo dinámico, porque siempre se podrá expandir y porque presenta algunos campos relativos a vistas de las bases de datos administradas en el consejo sectorial. En la fase de diseño del motor de búsqueda se especificará cada uno de los campos del metadato base del STIIP. Por el momento basta decir que todo dato que entra a los almacenes del STIIP, cualquiera sea el formato de almacenamiento, deberá ser asociado a un correspondiente metadato.

Cabe indicar que un campo importante del estándar del STIIP de metadatos, es el asociado al resultado de un análisis textual de cualquier tipo de texto asociado al dato. En la fase de creación del índice, la utilización de este campo nos permitirá ubicar los datos también en base al contenido de texto de los mismos y facilitará las búsquedas en base a recurrencias de términos; posteriormente, en la fase de diseño del motor de búsqueda, se hablará de esto.

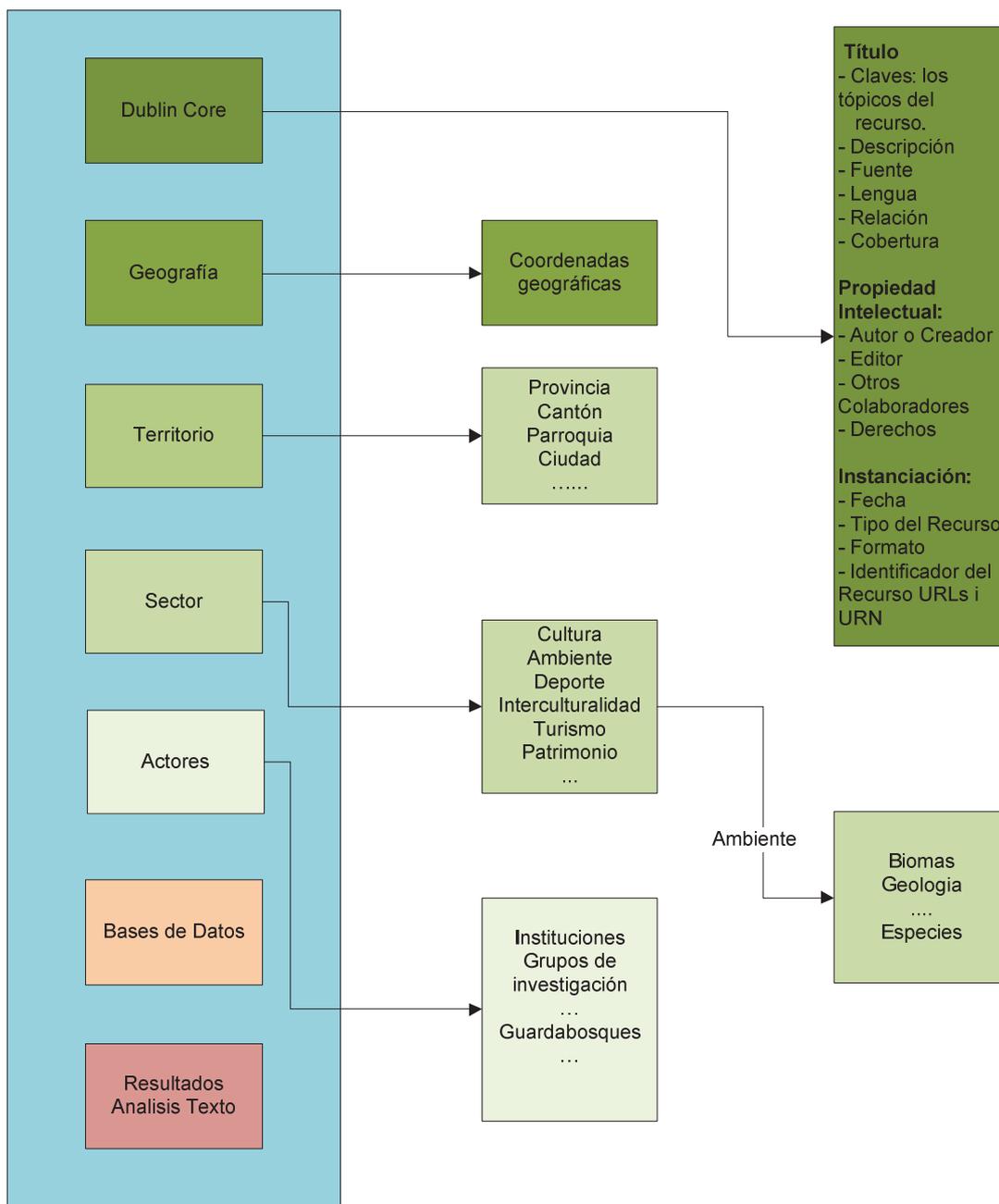


FIGURA 5-4 ESQUEMA DEL METADATO BASE PARA EL STIP

### 5.5.2 El índice

El índice será el corazón del STIIP. Una vez que cada dato genere su metadato asociado, habrá un proceso automático de generación del índice a partir de todos los metadatos. El índice puede ser pensado como un almacén en el que todos los metadatos, correspondientes a cada dato, son organizados en tal forma que permita al proceso de búsqueda realizar las búsquedas de documentos de manera extremadamente eficiente. Por el hecho de haber puesto entre los campos del metadato un campo relativo a los resultados del análisis textual, automáticamente el índice contiene una copia de todos los textos de los documentos procesados en el STIIP. Esto nos permite decir que el índice es también una copia de seguridad del texto de los documentos, mas no de todo el contenido del documento (especialmente aquel contenido que tenga gran peso de almacenamiento). Explicado de otra forma, muchos documentos como gráficos, imágenes, sonidos, videos, entre otros documentos del orden de magnitud de varios megabytes o gigabytes, ocupan mucho espacio en los servidores en donde residen y además la transmisión completa de ellos hacia el servidor central del MCP ocuparía mucho tiempo, lo cual haría ineficiente al proceso de indexación del STIIP. Para este tipo de archivos grandes, sus metadatos generados que serán enviados al almacén de metadatos central en el servidor del MCP, serán una copia de los campos básicos definidos (Dublin Core, territorio, sector, etc.), mas no de su contenido esencial. En otras palabras, el índice estará poblado de metadatos de todos los documentos del CSP en formato de texto, que tendrá campos de vinculación, si fuese el caso, hacia archivos de tamaño grande guardados en los almacenes distribuidos de la institución propietaria y responsable del archivo. En la fase de diseño del motor de búsqueda se explicará el detalle técnico del metadato y el índice, de acuerdo a las herramientas tecnológicas usadas.

### 5.5.3 El SDG-STIIP componente automático

Es importante notar que a diferencia del proceso de protocolo de manejo de datos, el proceso de creación metadatos es un proceso completamente automático.

En la figura 5-5 [34] se puede ver representado el STIIP en términos de red de conexión entre las diferentes instituciones coordinadas. Esta red posee una estructura de tipo centralizado en estrella, donde el servidor del MCP es el nodo central.

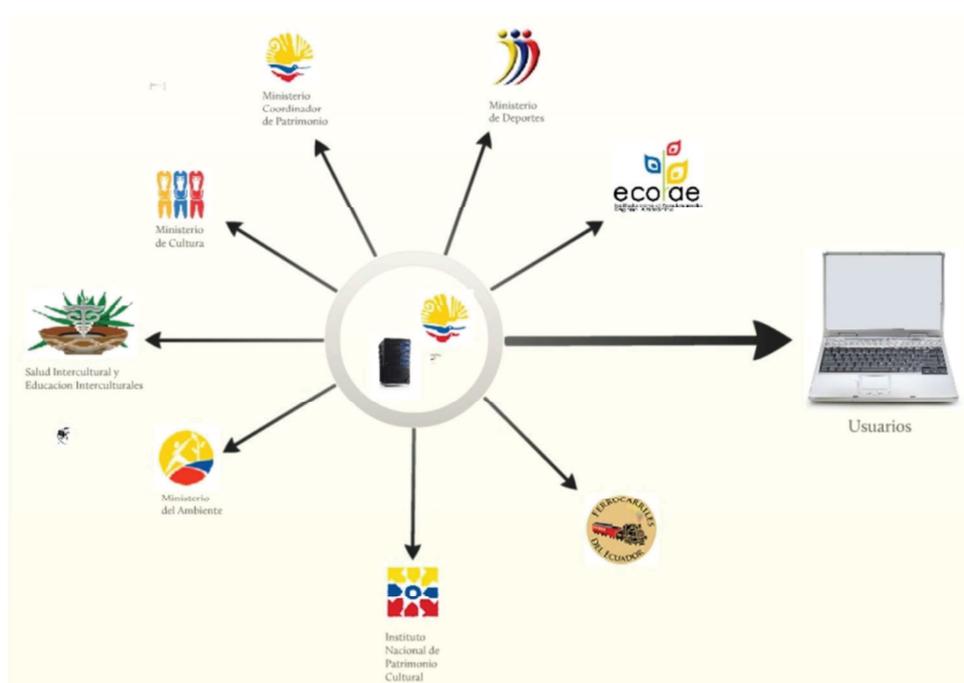
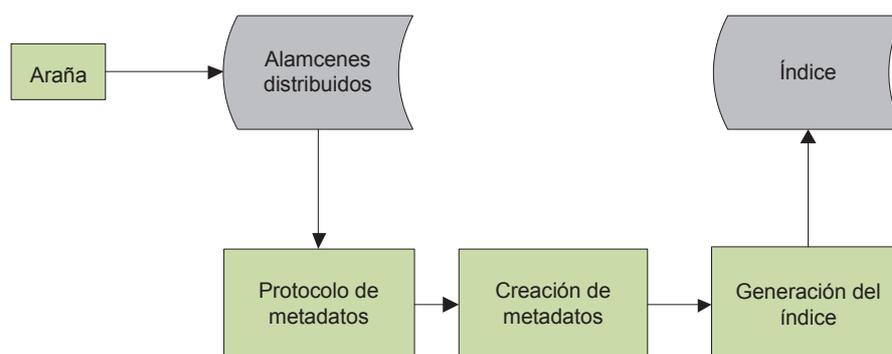


FIGURA 5-5 DIAGRAMA DE RED DEL CONSEJO SECTORIAL Y EL USUARIO

En el nodo central serán físicamente instalados los almacenes de metadatos del índice y de otros datos de carácter general. Los restantes nodos son nodos periféricos y corresponden a los almacenes distribuidos, indicados en el DFD de primer nivel del STIIP. El STIIP conceptualmente posee un algoritmo que se puede pensar como un *web crawler*, que llamaremos en este documento con el nombre de *araña*. La araña, moviéndose por la

telaraña constituida por la red de conexión de servidores del consejo sectorial, irá escaneando los documentos, las páginas web y las bases de datos que encontrará con el fin de generar automáticamente metadatos. Una vez generados todos los metadatos relativos a un nodo, estos serán comprimidos y enviados al servidor central que descomprimiéndolos procederá, a través del proceso de generación del índice, a actualizar el índice con los nuevos metadatos recién llegados. De esta manera con un proceso automático, el sistema es capaz de generar un índice que nos permite hacer funcionar el motor de búsqueda del STIIP, como se puede ver en la figura 5-6 [basado en 34].



**FIGURA 5-6 DFD DE LOS PROCESOS DE CREACIÓN AUTOMÁTICA DE METADATOS Y GENERACIÓN DEL ÍNDICE**

Además, el funcionamiento del STIIP se lo definió como una aplicación web distribuida; por lo tanto, se debe especificar una dirección de acceso remoto a los almacenes de cada institución. Esto implica otorgar los privilegios de acceso adecuados para que sea posible acceder a los datos de forma remota desde otra ubicación en una red.

### 5.5.4 Algoritmo de la araña

Una vez definidos los almacenes fijos de cada institución y, especificada una dirección de acceso remoto a los mismos, el STIIP podrá poner en marcha el proceso de indexación automática. Este proceso será realizado por el conjunto de algoritmos que anteriormente hemos llamado *araña*. A continuación se definirán las tareas de la araña.

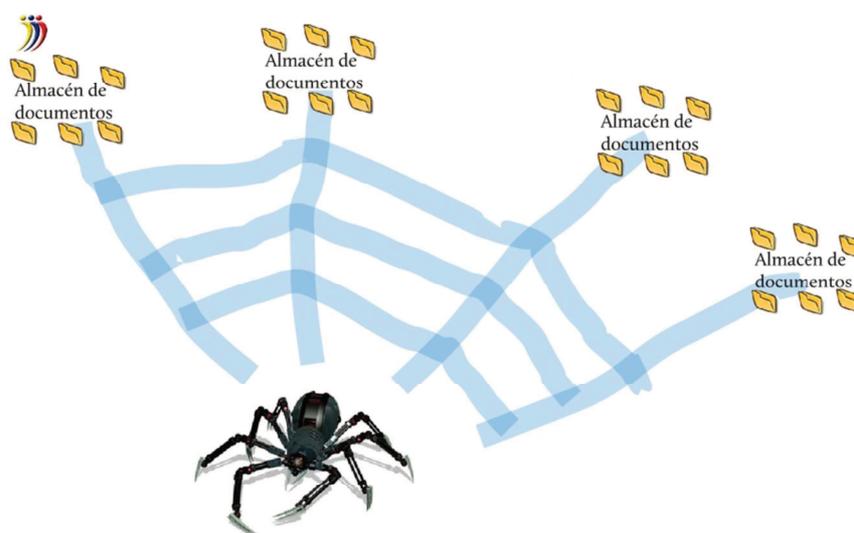


FIGURA 5-7 LA ARAÑA Y SU TELARAÑA

Como se puede ver en la Figura 5-7 [34] y como se indicó en los numerales anteriores, la araña del STIIP vive en la red (o telaraña) de comunicación telemática definida por las conexiones de cada uno de los almacenes de datos que las instituciones coordinadas especifiquen, en base a lo solicitado en la sección precedente.

Con intervalos fijos (semanal, mensual u otro dependiendo del tipo de datos contenidos en los diferentes almacenes), la araña saldrá de su ubicación en el servidor central del MCP e

irá visitando los almacenes de datos periféricos de las institución coordinadas; en términos técnicos, se realizará una conexión remota para la transferencia de archivos. Al llegar a un almacén, la araña iniciará el trabajo localmente, empezando por los puntos de entrada que se definirá a priori por cada almacén de datos de cada institución del CSP. Por ejemplo, dado un servidor, el punto de entrada podrá ser el nivel superior de un sistema de ficheros (o el directorio raíz), la dirección de un servidor de motor de base de datos, etc. A partir de cada punto de entrada la araña ejecutará el algoritmo de creación de metadatos graficado en la figura 5-8 [basado en 34].

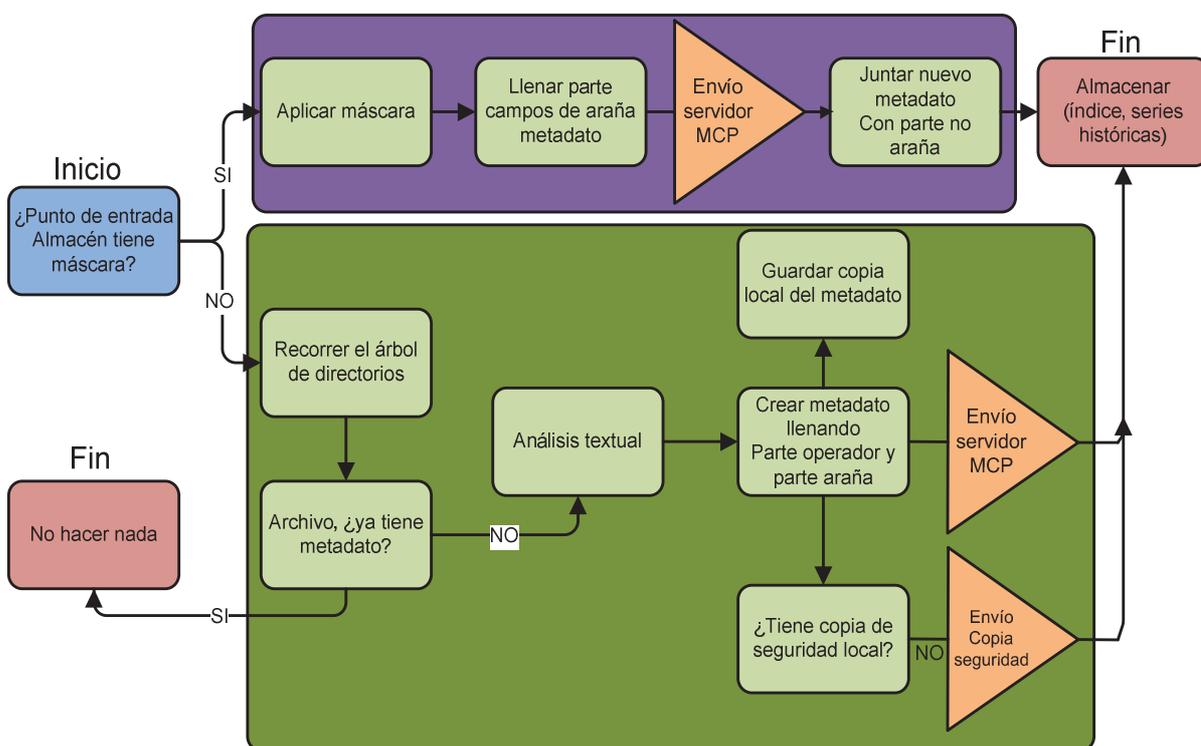


FIGURA 5-8 ALGORITMO DE CREACIÓN AUTOMÁTICA DE METADATOS

Analizando la parte superior (color morado) de la figura 5-8 de este algoritmo se tiene:

- Desde el punto de entrada, la araña hará distinción entre almacenes de datos con máscara y sin máscara. Una máscara es el equivalente a una vista de una base de datos. Se puede decir que algunos tipos de almacenes como las bases de datos tienen máscaras para permitir a la araña extraer conjuntos de informaciones específicas desde las bases de datos, mas no toda la base de datos y con estas llenar una parte de un metadato. De esto se hablará más en el proceso de gestión y análisis de indicadores.
- En el caso en que el almacén tenga máscara, la araña avanzará aplicando la máscara y con las informaciones extraídas llenará una parte de un metadato que enviará al almacén central del MCP, para que esta información sea unida con la parte de metadato ya presente en el servidor central acerca de la base de datos procesada.
- El algoritmo seguirá almacenando los nuevos metadatos en el almacén central de metadatos e ingresando los mismos en el índice.

Al analizar la parte inferior del algoritmo (color verde oscuro) de la figura 5-8 obtenemos lo siguiente:

- Si el almacén de datos no tiene máscara, significa que se trata de documentos almacenados en un sistema de carpetas: la araña transversará todo el árbol de directorios procesando cada elemento encontrado.

- En el caso que el documento no posea un metadato (la araña llegará a cada almacén periférico con una lista de los documentos encontrados en el precedente pasaje y con la información si cada documento ya tenía o no un metadato).
- El procesamiento consistirá en extraer informaciones de los documentos a través de un análisis textual. Los resultados del análisis textual serán almacenados en un metadato con los campos que la araña pueda llenar automáticamente, especialmente aquellos del Dublin Core.
- Se guardará una copia del metadato en el mismo almacén del documento.
- Se enviará una copia del metadato generado al servidor central.
- En el caso que el documento no tenga copias de seguridad en la institución, se enviará una copia del dato al servidor central.
- El algoritmo seguirá almacenando los nuevos metadatos, en el almacén central de metadatos e ingresando los mismos en el índice.
- En el caso que el documento posea ya un metadato, el algoritmo pasa al siguiente documento sin hacer nada más.

En ningún caso las informaciones generadas de forma automática, con análisis textual por la araña, podrán sobrescribir campos del metadatos llenados por un humano, la araña entonces realiza un preprocesamiento de los datos, ayudando a la generación de metadatos. Se asocia una importancia más grande a las informaciones ingresadas por operadores humanos, que a la generada automáticamente por los algoritmo de análisis textual.

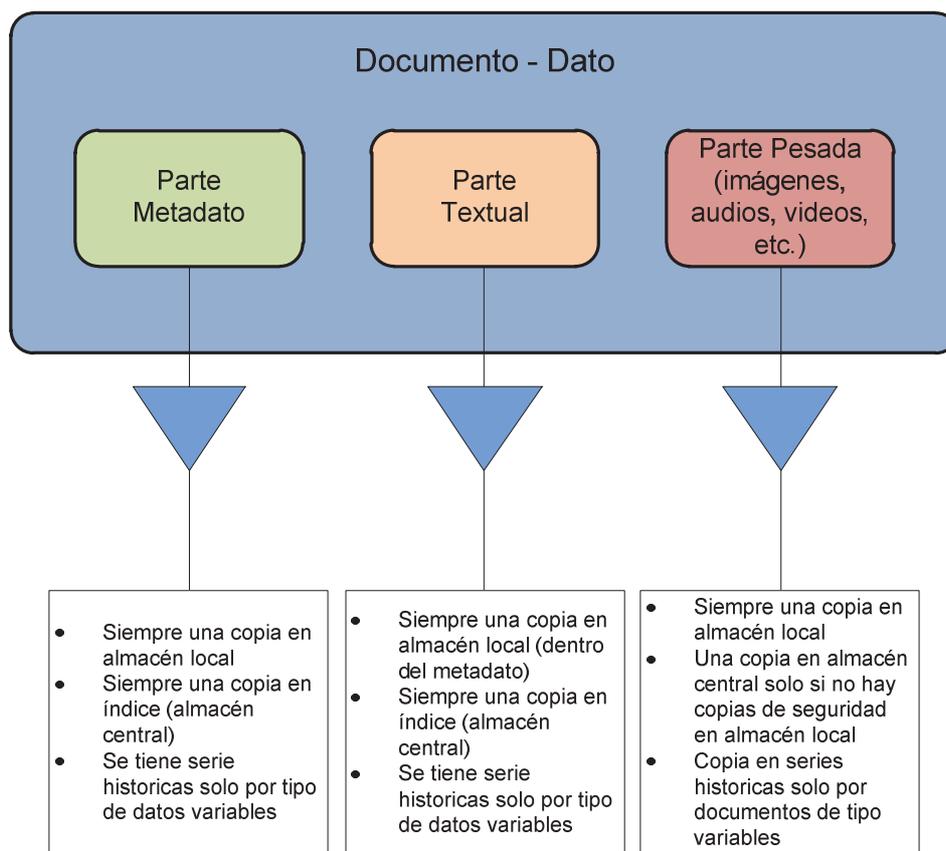
Cuando la araña termina su recorrido a lo largo de todos los sistemas de almacenamiento, se podrá contar automáticamente con un índice preprocesado, el cual tiene el vínculo de

unión hacia todos los datos y sus repositorios o almacenes periféricos de las instituciones coordinadas, que servirá de núcleo para correr el motor de búsqueda y posteriormente el geoportal.

#### **5.5.5 Protocolos de copia de datos**

Se ha visto que algunos de los datos que no tengan una copia de respaldo en los almacenes periféricos de las instituciones coordinadas, serán copiados en el servidor central del STIIP. Esto para garantizar que ninguna información de ningún tipo se pierda en los procesos de deterioro del hardware de cada institución, cumpliendo con el área de seguridad del SGD-STIIP.

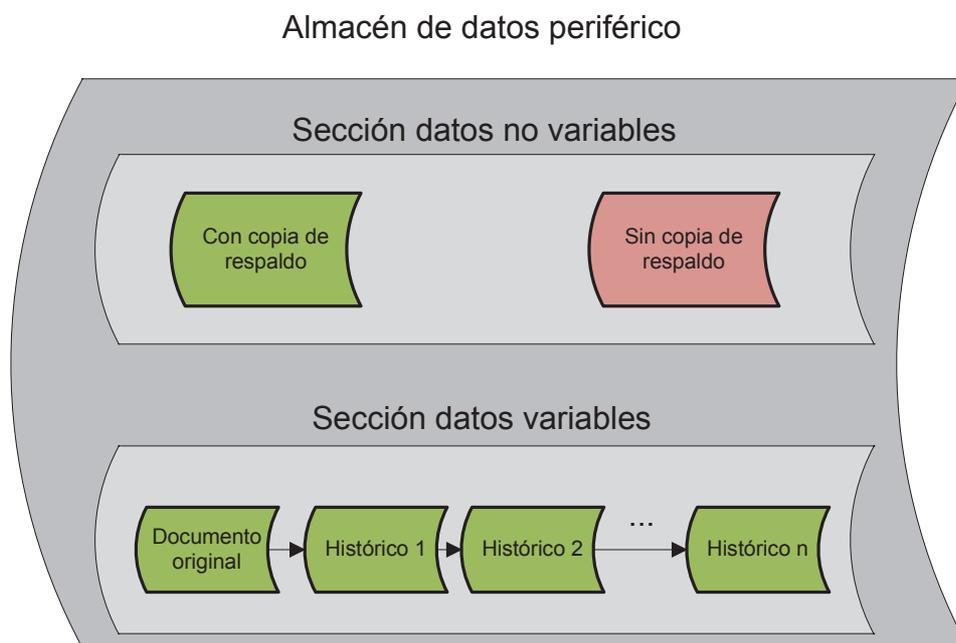
Además, se ha determinado que algunos de los datos y documentos serán definidos “variables” y otros “no variables”. La definición de esta característica dependerá de la persona que ingrese los datos en un inicio. Con esto se quiere decir que habrá algunos datos (variables) como los contenidos en las bases de datos, que se van continuamente actualizando o los documentos que se van produciendo a lo largo de la realización de proyectos o programas, que cambiarán en el tiempo. Por otra parte, habrá también documentos fijos (no variables) que no cambien en el tiempo, los reportes de investigaciones, los archivos de audio de entrevistas y otros.



**FIGURA 5-9 DIVISIÓN DEL VOLUMEN DE ALMACENAMIENTO DE DATOS**

Ahora se especificará en detalle cuales serán los protocolos de copia de los varios tipos de datos. Se puede ver el esquema de la Figura 5-9 [basado en 34] en donde se ha separado el volumen de almacenamiento relacionado a cada dato o documento en: metadato, parte textual y parte pesada que es la que se refiere a contenidos de imágenes, videos, audios y otros contenidos pesados. También, un documento como un pdf o de texto enriquecido tiene su parte pesada, que será la parte de imágenes u otro que contiene por su formato. Como se puede ver en la figura anterior, la parte textual y de metadato siempre se tendrá una copia en el almacén local de la institución y otra copia en el almacén central del MCP. Esto porque el volumen de texto y metadatos no representan un tamaño grande de almacenamiento para cualquier tipo de documento y principalmente porque se necesitan

estos valores para poblar el índice, para que el motor de búsqueda trabaje localmente en el servidor central del MCP, mejorando el rendimiento del proceso de búsqueda. La parte pesada del volumen de almacenamiento se copiará en el servidor central del MCP exclusivamente en el caso que el dato no posea una copia de seguridad en los almacenes periféricos y, dependiendo también de la disponibilidad de almacenamiento en el servidor central. Otros tipos de datos (bases de datos y documentos que se actualizan dependiendo de procesos dinámicos como seguimiento de proyectos), serán almacenados como series históricas. Por estos motivos los almacenes de datos de cada institución deberían organizarse por lo menos desde un punto de vista conceptual, como se muestra en la figura 5-10 [basado en 34].



**FIGURA 5-10 ESTRUCTURA DE UN ALMACÉN PERIFÉRICO**

Otra posible distinción dentro de la estructura de un almacén periférico podría ser el nivel de confidencialidad de los documentos. Esto para separar los procesos de búsqueda y

geoportales, que estarán disponibles para todos los usuarios, con los procesos de modelado, simulación, reportes e indicadores, disponibles solo a los tomadores de decisiones.

Otra función que la araña realizará para potenciar el algoritmo es el análisis de documentos duplicados. Periódicamente (trimestral o anualmente dependiendo), en su recorrido la araña efectuará un control de la unicidad de cada documento registrado en el índice a través de una comparación textual basada en el motor de búsqueda. Cuando resulta que dos documentos tengan una coincidencia textual de más del 95%, se pondrán los dos documentos en una lista especial en espera de una decisión (eliminar duplicado, guardar duplicado) por parte de un operador. Esto para evitar el duplicado del mismo documento dentro del índice, debido al almacenamiento del mismo en diferentes instituciones. Por ejemplo, es común que se de el caso de un acuerdo interministerial, que cada ministerio se quede con copias de las actas o documentos del tratado y los almacene en sus respectivos repositorios.

## **5.6 Procesos de protocolo de manejo de datos y de metadatos**

Como se ha visto en la fase de análisis, hasta la fecha no existe una manera estándar de manejar datos en las instituciones coordinadas; es decir, no existe un Sistema de Gestión Documental (SGD) común para todas las instituciones. Esto es un aspecto del STIIP muy importante, porque el propósito de encontrar datos e informaciones no puede ser cumplido si estos no tienen ubicaciones y estándares de almacenamientos establecidos y fijos en el tiempo. Aunque en el STIIP se prevé la implementación de un algoritmo capaz de recolocar en el índice datos que hayan sido movidos o ligeramente transformados, la

condición mínima para que un SI distribuya informaciones es la de tener un principio ordenador en la manera de almacenar documentos y datos. Por ejemplo, el esfuerzo de los motores de búsqueda actuales y de renombre, como Google, Yahoo! o Bing, para indexar páginas web y poder presentar resultados a búsquedas realizadas a sus bases de datos o índices, no serviría de nada si las páginas web cambian de dirección cada semana. Lo mismo se puede decir acerca de la dirección de un servidor que almacena documentos del CSP: el esfuerzo de construcción del STIIP debe estar sincronizado a un esfuerzo de organización en la gestión de datos, dentro de cada una de las instituciones coordinadas.

Es necesario entonces definir un protocolo estándar de manejo de datos, lo cual implica que dentro de cada institución coordinada se deberá decidir de forma definitiva los lugares y las maneras en que se irán almacenando los datos relevantes al propósito integrador del patrimonio del STIIP. De esta manera, cualquier técnico o persona que esté encargada de ingresar datos en su institución, lo hará de forma en que los datos sean almacenados en lugares precisos y que posibilite el funcionamiento del STIIP para distribuirlos a todo el CSP.

De la misma manera, es necesario definir un estándar de manejo de metadatos que acompañe la gestión de datos. En el STIIP cada dato está acompañado por un metadato, el ingreso de cada dato igualmente será acompañado por el ingreso de un metadato. Cada actualización de un dato corresponderá a una actualización de un metadato. Este proceso de gestión de metadatos, operados por las mismas entidades que se ocuparán de ingresar datos, es muy importante porque nos permite conseguir una información mucho más detallada y precisa acerca de nuestros datos, la que podríamos conseguir solamente a través del proceso de generación automática de metadatos.

El STIIP no pretende imponer un protocolo único y rígido de manejo de datos en todas las instituciones del CSP, a fin de dejar inválido cualquier sistema que se esté usando hasta el presente. Además, el MCP por su mismo nombre, es una entidad coordinadora, no regidora, por lo que el proyecto STIIP busca adaptarse a los sistemas de organización preexistentes en cada institución e integrarlos de tal manera que se genere información valiosa tras la unión de todos ellos. En este sentido, no se especificará un modelo de manejo de datos a seguir, pues es competencia de cada entidad del CSP definir su propio modelo de manejo de datos, muchas veces ya establecido, pero sí se exige una continuidad con su modelo para el funcionamiento óptimo del STIIP.

### **5.6.1 El SDG-STIIP componente humano**

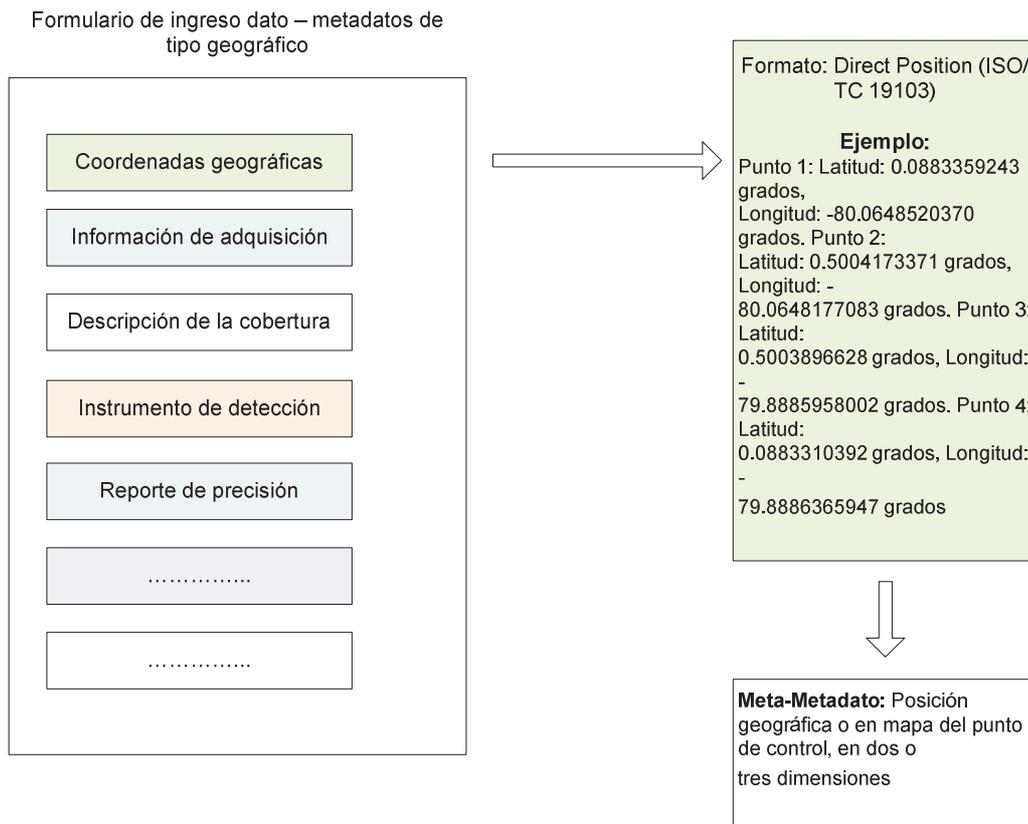
En este caso, contrariamente a lo que ocurre en el SGD-STIIP componente automático, el ingreso y generación de datos y de metadatos será de manera manual: los operadores humanos podrán realizar las dos tareas gracias a las interfaces del manejo de datos y de metadatos. A continuación, se incorpora algunas definiciones que nos servirán para entender cómo funciona este proceso.

### **5.6.2 Definición de los elementos del manejo de datos y de metadatos**

Desde el punto de vista conceptual, la interfaz del manejo de datos y de metadatos, será constituida por los elementos siguientes:

- **Tipos:** son un conjunto de formatos relacionados con la información que es ingresada. Corresponden casi siempre a campos del metadato. Cada tipo tiene un formato y un meta-metadato. Ejemplo: el tipo “*coordenadas geográficas*” tiene formato “(*coord\_latitud, coord\_longitud*)”, es decir, es un vector de dos dimensiones que tiene formato de números reales.
- **Formularios:** son agregados estándares de tipos. Cada formulario tiene: una lista fija de tipos y un meta-metadato. Ejemplo: el formulario correspondiente al Perfil Ecuatoriano de metadatos geográficos.
- **Meta-metadatos:** son metadatos de metadatos: proveen informaciones sobre los tipos y los formularios y también sobre los metadatos existentes. De esto se expondrá en detalle en la fase de diseño del motor de búsqueda, de acuerdo a las herramientas tecnológicas usadas.

Un ejemplo de esquema del formulario se muestra en la figura 5-11 [34].



**FIGURA 5-11 EJEMPLO DE ESQUEMA DE FORMULARIO DE INGRESO DE DATO - METADATO**

Cabe observar que, una parte muy importante de los tipos definidos en el metadato del STIIP, será correspondiente a las componentes de la cartografía de base del STIIP. Se introducirá la cartografía de base del STIIP en el capítulo de diseño del geoportal del STIIP.

Además, definimos tres tipos de operadores humanos, según diferentes privilegios, que podrán interactuar con el sistema desde el punto de vista del manejo de datos – metadatos:

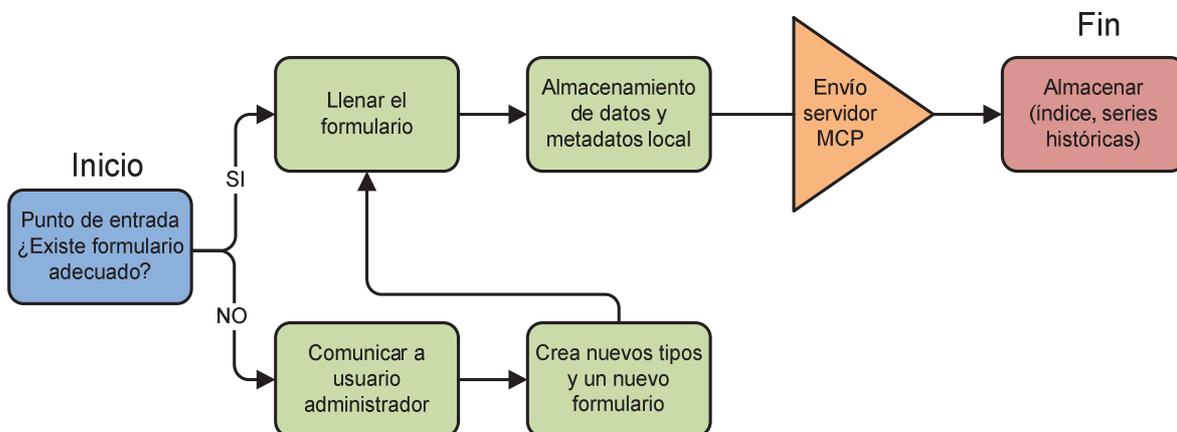
- **Usuario ordinario:** Solo busca, visualiza y descarga documentos.

- **Usuario técnico:** busca, visualiza y descarga documentos; ingresa, elimina y actualiza documentos.
- **Usuario administrador:** busca, visualiza y descarga documentos; ingresa, elimina y actualiza documentos; también crea nuevos formularios, tipos y meta-metadatos.

La interfaz de manejo datos-metadatos presentará la posibilidad a cada usuario, según sus privilegios, poder interactuar y utilizar los tres elementos definidos arriba. En este contexto por ejemplo, utilizando un formulario el usuario técnico puede: especificar la ubicación de la fuente física del dato que quiere ingresar (archivo, base de datos, etc.), llenar el metadato, enviar la actualización al sistema central, que se ocupará de forma automática de integrarlo al núcleo del STIP o índice.

### **5.6.3 Algoritmo del SGD-STIP componente humano**

El proceso de manejo de datos y metadatos permitirá a los operadores humanos ingresar nuevos datos con sus respectivos metadatos. Los datos serán ingresados directamente en el sistema de almacén de la institución desde la cual el operador se está conectando, para luego pasar a formar parte de la definición estructural de los metadatos, es decir, los meta-metadatos.



**FIGURA 5-12 ALGORITMO DE INGRESO DATOS Y METADATOS COMPONENTE HUMANO**

En la misma tarea de llenar los campos del formulario que se está utilizando, el operador indicará al sistema si es necesario enviar una copia de seguridad del dato al servidor central del STIIP, cuál debe ser el nivel de acceso al dato, si el dato es variable o no variable, si es necesario almacenar el dato dentro de una serie histórica, entre otras definiciones.

Para ingresar datos y metadatos, el operador seguirá el algoritmo dibujado en la Figura 5-12 [basado en 34], en los siguientes puntos se describe el algoritmo:

- Un usuario técnico tiene que ingresar un dato, abre la interfaz y busca un formulario apropiado para aquel dato.
- Si el formulario existe, el técnico lo llena, lo ingresa al sistema y automáticamente el sistema manda copia del metadato al servidor central y si fuese necesario una copia del dato también.

- Si el formulario adecuado no existe, el usuario técnico deberá contactar a un administrador del sistema y pedir que se cree un formulario adecuado a sus tipologías de datos.

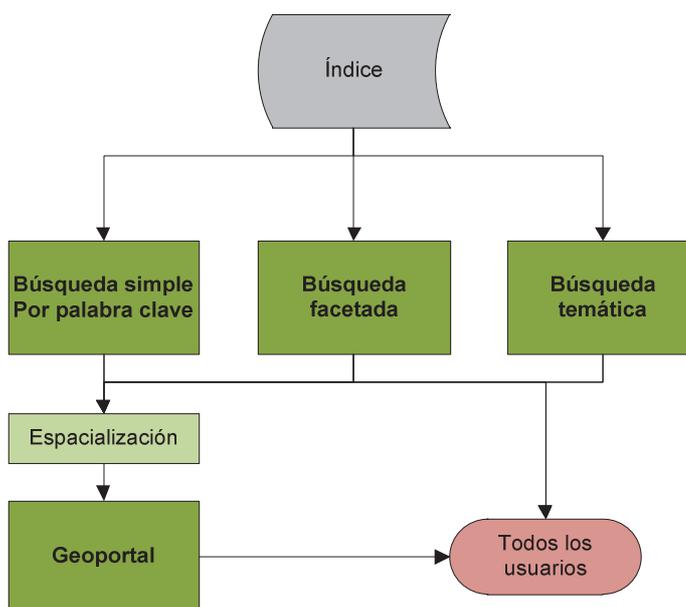
Es claro que, el sistema en su implementación inicial presentará un número suficiente de formularios ya predefinidos y que no será tan frecuente la exigencia de crear nuevos formularios. Esta posibilidad es prevista claramente porque el STIIP se imagina como un sistema modular y expandible, por lo que entonces será siempre posible añadir nuevos tipos de datos y metadatos.

También será posible llenar formularios de forma semiautomática, utilizando formularios ya llenados como plantillas. Una vez llenado el formulario, todos los subsecuentes procesos son los mismos del caso de ingreso automático, con relación a la generación del índice en el servidor central.

## **5.7 Proceso de búsqueda**

Todos los días millones de internautas utilizan servicios idénticos a lo que es el proceso de búsqueda del STIIP. Portales como Google, Yahoo! y otros son basados en motores de búsqueda muy eficientes que disponen de robots muy sofisticados (los equivalentes de nuestra araña), que escanean el Internet e indexan los sitios web, permitiendo búsquedas supereficientes de datos en la web. La manera de recolección de datos en el proyecto STIIP será similar.

En la Figura 5-13 [basado en 34], está representado el diagrama de flujo del proceso de búsqueda. Como se puede ver el proceso de búsqueda ha sido separado en tres subprocesos distintos: la búsqueda simple por palabra clave, el proceso de búsqueda facetada y el proceso de búsqueda temática.



**FIGURA 5-13: DFD DEL PROCESO DE BÚSQUEDA DEL STIP**

La búsqueda por palabra clave es la más usada en los motores de búsqueda, es decir, ingresar palabras características de un documento, por ejemplo realizar la búsqueda de “bosques amazónicos”, para encontrar una lista con resultados como “bosque tropical amazónico”, “bosque lluvioso”, “Amazonas”, etc.

El proceso de búsqueda facetada es algo que desde algunos años se va desarrollando en los principales motores de búsqueda de Internet: como resultado de una búsqueda por palabra

clave, las páginas de resultados presentan la lista de los documentos encontrados más una lista de categorías a las cuales el motor de búsqueda los asocia (ver Anexo 4) en el STIIP. Por ejemplo en una búsqueda, el motor de búsqueda del STIIP podrá descubrir que en el índice hay documentos con referencias acerca de la ciudad de Riobamba, del cantón Riobamba, de la calle Riobamba en Quito o la calle Riobamba en Guayaquil y además, de presentar los resultados de la búsqueda, presentará una lista de categorías, en este ejemplo, podría ser una lista con las categorías Ciudad: 1 resultado, Cantón: 1 resultado, Calle: 2 resultados. Esta manera de presentar los datos encontrados es muy útil para la búsqueda y encarna un verdadero primer nivel de análisis de datos ya que: 1) permite al usuario filtrar de manera muy eficaz los datos encontrados: si busco imágenes de la ciudad de Riobamba, primero selecciono en la columna de la izquierda imágenes y después los documentos acerca de la ciudad de Riobamba que excluyen a todas las categorías que no sean imágenes; y, 2) esta función de búsqueda facetada permite descubrir informaciones que el usuario no se le ocurriría encontrar: por ejemplo, el usuario que busca “osos de anteojos” podría aparecerle una lista facetada con categorías Pastaza, Monumentos, Especies y descubrir que existen monumentos de osos de anteojos en el país. Esto se logra con búsquedas facetadas dinámicas, es decir que las categorías presentadas no sean fijas sino dependientes de la primera búsqueda por claves realizada.

Un tipo de proceso de búsqueda adicional es la temática. Esta es una búsqueda en que el usuario puede encontrar los datos catalogados en un orden temático establecido a priori, como por ejemplo en la red se encuentra la información en las *Yahoo! Directories*. Este tipo de búsqueda, es como la búsqueda facetada dinámica, pero con una lista fija de las categorías más importantes para el CSP.

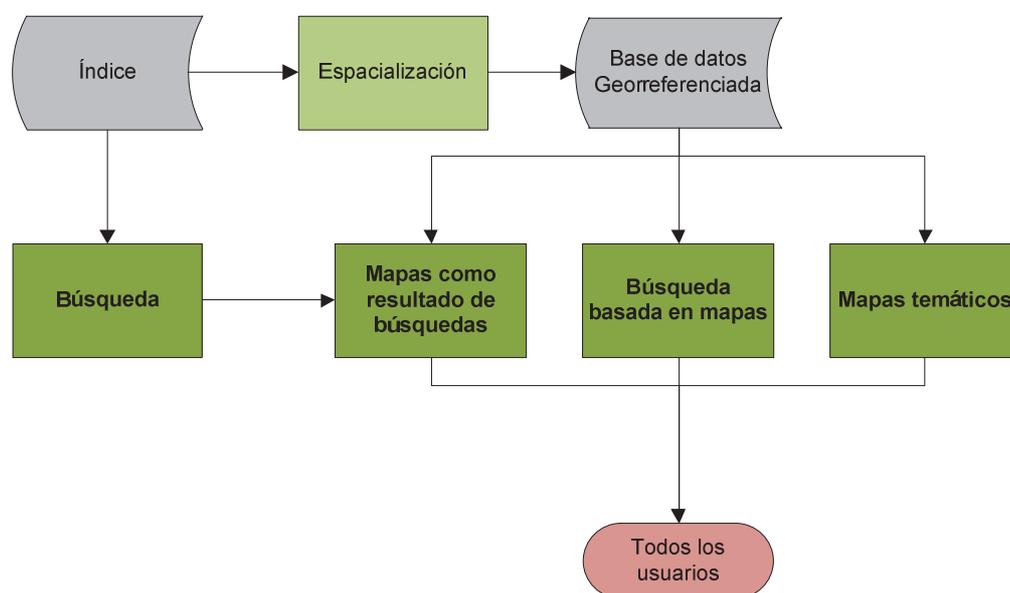
Una última observación, como se puede ver en la Figura 5-13, los procesos de búsqueda son conectados al proceso geoportal a través de varios flujos de datos en los dos sentidos, esto simplemente significa que se podrán encontrar documentos e informaciones desde los mapas dinámicos y, también crear mapas dinámicos a partir de documentos encontrado a través del proceso de búsqueda. El detalle técnico del proceso de búsqueda se hablará con mayor profundidad en la sección de diseño del motor de búsqueda.

## 5.8 Proceso de geoportal

Siendo el STIIP un sistema informativo territorial, el proceso de geoportal es uno de los más importantes en este sistema. El proceso geoportal es un proceso que se basa en la estructura geográfica y espacial del metadato. A través de las informaciones contenidas en el metadato es posible en principio construir una base de datos georreferenciada, que permitirá construir mapas dinámicos. Los mapas dinámicos construidos permitirían efectuar búsquedas basadas en mapas, como ejemplo de esto, actualmente se puede ejecutar en internet con el famoso *Google Maps*. Utilizando un servicio similar se pueden buscar documentos simplemente de forma visual, enfocando nuestra atención en las áreas geográficas de nuestro interés. Una búsqueda de este tipo es muy simple, intuitiva y más interactiva, que permitiría a cualquier usuario tener una idea muy clara de la situación patrimonial territorial de cada área del país. Desde otro punto de vista, el proceso del geoportal puede ser imaginado como una función del STIIP, donde el usuario puede construir mapas en base a las búsquedas efectuadas en el proceso de búsqueda. Por ejemplo, ingresando las palabras claves “oso de anteojos” en el motor de búsqueda, se podrán visualizar distribuidos en un mapa todo lo relacionado acerca de documentos con

alguna relación a los resultados de la búsqueda “oso de anteojos” y de esta manera hacerse una idea, por ejemplo de la distribución geográfica de los hábitats de tal animal. Otro ejemplo es la visualización en el mapa de los yacimientos arqueológicos de todo el país.

Otra función del geoportal, es la producción de mapas temáticos, que son mapas generados a partir de los datos contenidos en la base de datos georreferenciada y utilizando también otros tipos de fuente de datos. Estos mapas serán enfocados en temas particulares como podrían ser los parques nacionales o las áreas de riesgo del patrimonio cultural, dependiendo básicamente de la información que el CSP provea. Dichos tipos de mapas serán utilizados sobre todo por funcionarios que toman decisiones. En la figura 5-14 [basado en 34] se puede ver el diagrama de flujo del proceso de geoportal.



**FIGURA 5-14 PROCESO GEOPORTAL DEL STIIP**

## 5.9 Proceso de espacialización

El proceso desde un mapa, donde se puede pasar a una lista de documentos, es el proceso de espacialización o visualización de los resultados de la búsqueda en un mapa.

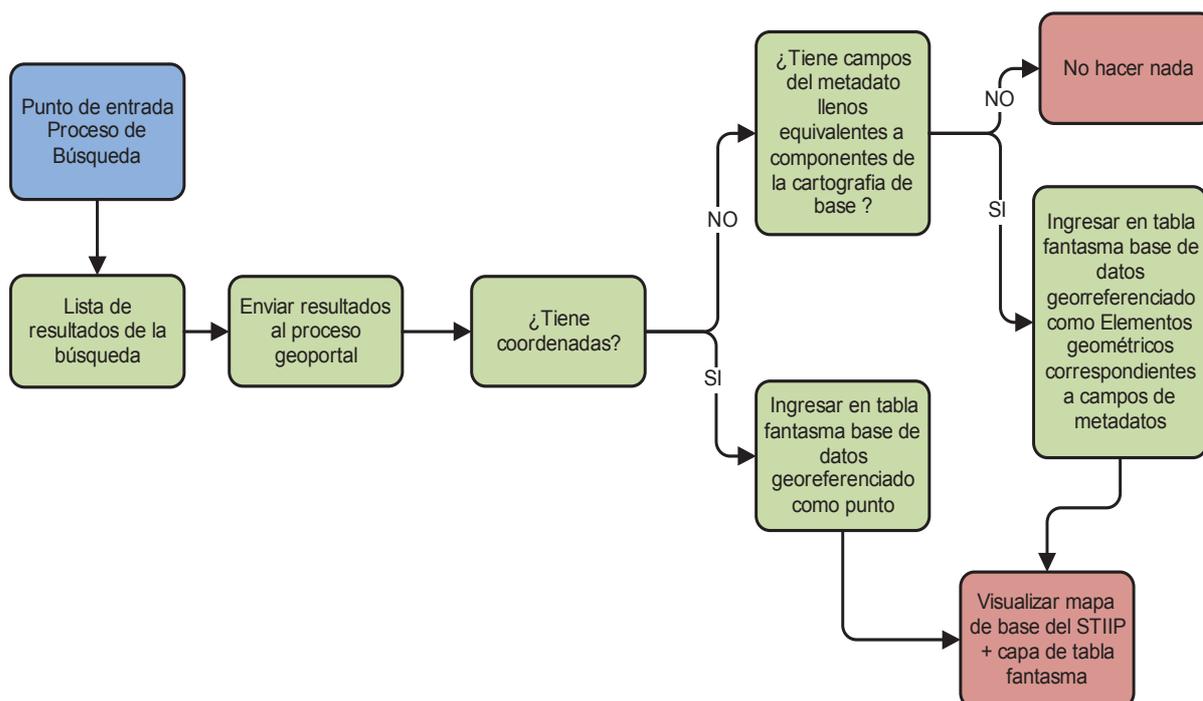
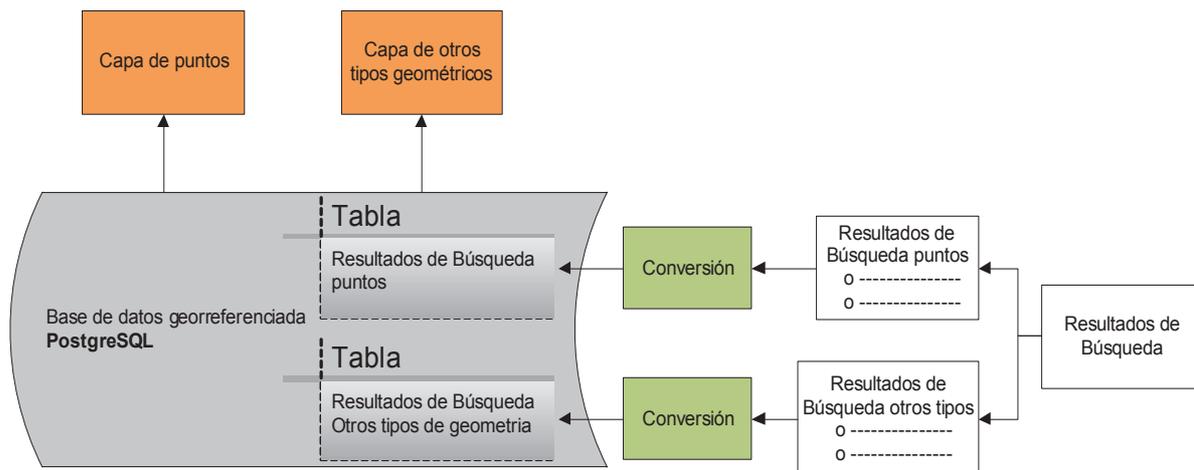


FIGURA 5-15 ALGORITMO DE ESPACIALIZACIÓN O VISUALIZACIÓN DE LA INFORMACIÓN RECUPERADA EN UN MAPA

A través del proceso de búsqueda, como hemos visto, el usuario podrá utilizar filtros sintácticos, geográficos (coordenadas, elementos geográficos) y otras herramientas para refinar sus búsquedas basadas en palabras claves. En cualquier momento de una búsqueda el usuario tendrá una lista de documentos y metadatos relacionados como resultado. El proceso de búsqueda STIIP será conectado con el geoportal, de manera que los resultados de la búsqueda, cuando estén georeferenciados o asociados a elementos de la cartografía

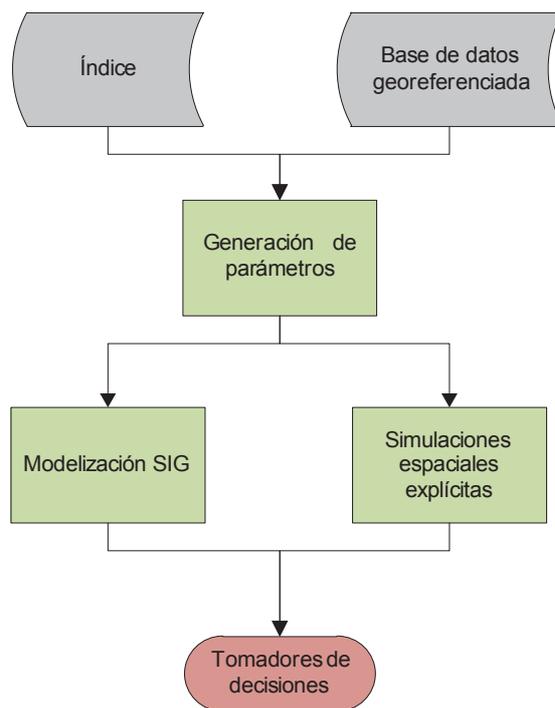
base del STIIP a través de sus metadatos, se puedan visualizar en un mapa a través de un solo clic. En la figura 5-15 [34] se ha dibujado el algoritmo de funcionamiento del procedimiento de visualización de la información recuperada, a través del proceso de búsqueda en el mapa base del geoportal del STIIP. Como ya se ha mencionado, el STIIP posee una cartografía base y un sistema cartográfico temático que serán almacenados en una base de datos georreferenciada, en resumen el mapa base, es el mapa del Ecuador con varios elementos geográficos o capas, por ejemplo provincias, cantones, ciudades, ríos, carreteras, etc. En esta base de datos se definirán dos tablas fantasmas o temporales (por usuario que realiza una búsqueda), que serán utilizadas para almacenar en forma de registros georreferenciados los resultados de la búsqueda. Los datos que se irán a almacenar en la tabla fantasma serán ingresados como elementos geométricos distintos según su origen. La primera tabla fantasma se llenará con los documentos encontrados de tipo georreferenciado (que tengan coordenadas geográficas explícitamente indicadas en el metadato con longitud y latitud) y serán ingresados en la tabla fantasma como puntos. Los documentos referenciados a áreas geográficas, por ejemplo ciudades, cantones o ríos, se ingresarán a la segunda tabla fantasma, como figuras geométricas o poligonales equivalentes a la geometría o perímetro del elemento del correspondiente componente del sistema cartográfico del STIIP. Una vez llenadas ambas tablas con todos los resultados de la búsqueda, con algún tipo de georreferencia, será dibujado como una capa más en el mapa base del STIIP. De esta forma, se logrará los requerimientos de información integrada territorial de los documentos y datos encontrados a través del proceso de búsqueda. En la Figura 5-16 [34] se puede observar el proceso de conversión de tipos geográficos en los mapas dinámicos. Más adelante y con mayor detalle se describe las arquitecturas informáticas del STIIP y el proceso tecnológico de conversión de los resultados de búsqueda en registros de la tabla fantasma.



**FIGURA 5-16 PROCESO DE CONVERSIÓN DE RESULTADOS DE BÚSQUEDA EN CAPAS DE MAPAS DINÁMICOS**

### 5.10 Proceso de modelado y simulaciones espaciales

Los datos y las informaciones procesadas en el proceso del geoportal y también las contenidas en el índice, podrán ser utilizados para elaborar modelos y simulaciones espaciales explícitas. En este proceso se irán generando proyecciones y predicciones que serán muy útiles para los usuarios que toman decisiones. El proceso indicado se desarrolla a nivel central, bajo las indicaciones del MCP.



**FIGURA 5-17 DFD DEL PROCESO DE MODELADO Y SIMULACIONES ESPACIALES**

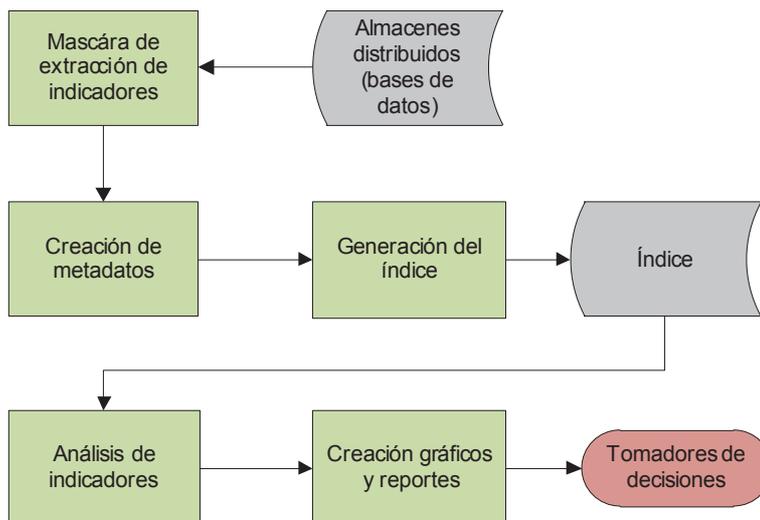
Los resultados de las simulaciones y de los modelados serán distribuidos a través del STIIP como indicadores y por lo tanto accesibles a través del proceso de organización y análisis de indicadores. A ese proceso serán asociados como herramientas externas al STIIP algunos programas de modelado SIG y programas de simulación espacial explícita, como se muestra en la figura 5-17 [basado en 34].

### **5.11 Proceso de gestión y análisis de indicadores**

Este es un proceso en que todos los indicadores de cualquier tipo encontrados en los sistemas de almacenes periféricos son agrupados y analizados de una forma tal que sea más fácil encontrarlos en el STIIP y sea posible visualizarlos en una forma adecuada. Este

proceso se ha diseñado para el STIIP, precisamente porque los indicadores (en particular los indicadores de la Agenda Sectorial del CSP) son datos muy importantes para el manejo del patrimonio cultural y también para la toma de decisiones a todos los niveles.

La manera de ser presentados y la forma de ser almacenados en el caso de todo tipo de dato dentro del STIIP, no van a ser cambiados por el proceso de indexación. Los indicadores de cada institución se podrán utilizar no sólo para producir gráficos y reportes sino también para cruzar esas informaciones con otros datos e indicadores. En el proceso de gestión y análisis de indicadores por cada indicador, donde sea necesario, se debe construir una máscara que nos permita extraer los datos más importantes del indicador y de utilizarlos para la producción de gráficos y reportes. En este proceso los indicadores no van a perder la posibilidad de ser buscados a través de los procesos de búsqueda ya que, van a ser almacenados en el mismo índice, sino que van a ser evidenciados de manera apropiada para la creación de reportes e indicadores del CSP. En la Figura 5-18 [basado en 34], se muestra el DFD de este proceso.



**FIGURA 5-18 DFD DEL PROCESO DE GESTIÓN Y ANÁLISIS DE INDICADORES**

## 5.12 Proceso de Patrimonio participativo

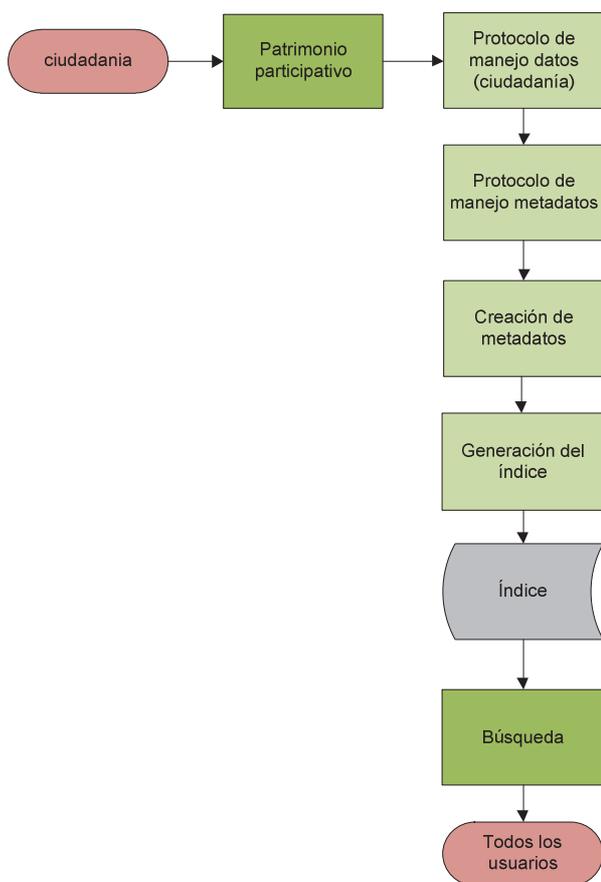
El propósito de este proceso es dar a toda la ciudadanía un instrumento de acercamiento al manejo del patrimonio ambiental y cultural del Ecuador. La idea es abrir una parte del sistema informativo y permitir a todos, empezando por las administraciones locales, el poder ingresar datos, informaciones acerca del patrimonio del país. Esas informaciones serán ingresadas a través del proceso de Patrimonio participativo y automáticamente asociadas a metadatos, así que estarán de forma automática disponibles a la búsqueda. Cualquier ciudadano podrá pedir ser acreditado como usuario de este servicio y, así conseguir los privilegios de enviar las informaciones a este proceso del STIIP. Este proceso se ha diseñado para que las entidades externas que contribuyan a ingresar datos, sean las municipalidades y las administraciones locales en general.

Este proceso podrá ser utilizado por la ciudadanía, por ejemplo, para señalar nuevas rutas de interés turístico y ambiental, así como la difusión de actividades de carácter cultural como fiestas tradicionales y celebraciones religiosas. A través de ese proceso, se podrán ingresar datos en formatos muy diferentes como videos, imágenes, audios, textos. Sin embargo, el nivel de confiabilidad de los datos producidos por la ciudadanía será menor a los datos ingresados por las instituciones del CSP, por lo que su indexación llevará un campo dentro de su metadato para especificar su grado bajo-medio de validez. Esto será controlado por el Protocolo del manejo de datos (ciudadanía). Lo que técnicamente implica es que, dentro del motor de búsqueda, los documentos ingresados por el CSP primarán sobre los documentos ingresados por la ciudadanía. A pesar de tener restricciones la información ciudadana se considera muy valiosa por el tamaño que puede adquirir, dependiendo de las campañas que se realicen para su utilización, además de ir acorde de las políticas públicas vigentes, respecto a la participación ciudadana.

Las entidades externas que más podrían aprovechar de este proceso, aparte de la ciudadanía en su totalidad, serían los tomadores de decisiones, porque tal repertorio de información actualizada acerca del patrimonio sería muy importante en la fase de planificación. Si los resultados participativos a esta iniciativa resultarán ser positivos y, la cantidad de datos ingresados es suficientemente grande, se podría pensar en utilizar este repositorio de datos para crear un directorio de mapas interactivos de turismo, cultura, ambiente, cuyos datos tendrían como origen la ciudadanía misma. Finalmente, este proceso se lo considera experimental y se requiere que los campos de ingreso de los datos básicos sean:

- Nombre del patrimonio.
- Tipo de patrimonio: cultural, natural.

- Ubicación del patrimonio: municipio, parroquia, cantón, provincia.
- Descripción del patrimonio.
- Documentación: Imágenes, fotografías, mapas.
- Datos de la persona que reporta el patrimonio.



**FIGURA 5-19 DFD DEL PROCESO DE PATRIMONIO PARTICIPATIVO**

En la figura 5-19 [basado en 34] se muestra el DFD del proceso de Patrimonio participativo.

### 5.13 Resumen de las características del SGD-STIIP

Después de la fase de diseño, se puede finalmente responder las preguntas realizadas en la sección de análisis para la creación del SGD-STIIP. A continuación se indica las respuestas a las preguntas realizadas.

#### 1. *¿Cómo se organiza los datos?*

Se ha señalado que la organización de los almacenes de datos, se realiza solicitando a cada institución lo siguiente:

- a) Que defina los almacenes físicos de datos fijos en el tiempo, de manera que las direcciones de acceso a los datos no cambien con el tiempo.
- b) Que estructure los almacenes de datos de tal manera que se formen secciones dedicadas a clases de datos diferentes: los datos variables y no variables, los datos con copias de seguridad y sin copia de seguridad, secciones dedicadas a datos con nivel de privilegios equivalentes.
- c) Que indique una dirección de acceso remoto a los almacenes.

#### 2. *¿Cómo se crea, guarda y actualiza los documentos?*

Habrán dos tipos de procesos para cumplir con esta tarea: el componente SGD-STIIP automático (araña) y el SGD-STIIP humano. Ambos componentes se encargarán de generar, guardar y actualizar los metadatos de cada documento, que serán almacenados en el índice central, para la realización de las demás funciones requeridas por el STIIP.

### 3. *¿Cómo se evita la pérdida de documentos?*

A través de la utilización de los metadatos y del cumplimiento de las tareas de organización de los almacenes. El texto de todos los documentos, de todo tipo, será incluido en uno de los campos de su respectivo metadato, almacenado en el índice o almacén central. La pérdida de documentos cuya información sea diferente de texto, se prevendrá con la creación de copias de seguridad en el almacén central, solo para aquellos documentos que requieren una copia de seguridad por su sensibilidad y de acuerdo a la disponibilidad de almacenamiento del servidor central.

### 4. *¿Cómo se evita la violación de la información reservada?*

Por medio de la definición de un campo dentro del metadato, con el nivel de privilegio para especificar qué tipo de usuario (ciudadanía, técnicos, tomadores de decisión) o proceso puede acceder al dato.

### 5. *¿Cómo se decide qué documentos conservar?*

Este aspecto depende de las especificaciones de cada institución del CSP y la manera de organización de sus documentos. La información extraída por los procesos del STIIP en los almacenes periféricos, será almacenada en el índice central y también será competencia del MCP el manejo de los metadatos; ya que en general, los metadatos contienen un vínculo a la ubicación del dato original; la eliminación o modificación del documento en los servidores periféricos y fijados para la utilización del STIIP, provocará su cambio mediante el SGD-STIIP componente automático, con una revalidación periódica de su metadato anidado.

## 5.14 Resumen de los procesos del STIIP

Para tener un punto de vista general se puede hacer una distinción entre procesos internos, que corresponden a los componentes internos acerca del funcionamiento del sistema y, procesos externos que corresponden a los componentes externos que presentan una interfaz al usuario. Los procesos internos son:

1. El proceso del protocolo de manejo de datos, a través del cual se definen los estándares de producción, ingreso y modificación de datos.
2. El proceso del protocolo de manejo de metadatos, mediante el cual se definen los estándares de producción, ingreso y modificación de metadatos.
3. El proceso de creación de metadatos, por el cual se procederá a llenar automáticamente algunos campos de los metadatos en todo el sistema de información.
4. El proceso de generación del índice, con el cual se creará el índice, basado en los metadatos y, será la base de todo el sistema de búsqueda de datos del sistema de información.
5. El proceso de espacialización, que a partir del índice se construirá una base de datos georreferenciados.

Asimismo, los procesos externos que se han definido son:

1. El proceso de búsqueda, a través del cual se podrá buscar de diferentes maneras la información en el sistema.

2. El proceso del geoportal, mediante el cual se podrá generar mapas dinámicos, efectuar búsquedas de datos en mapas, visualizar los resultados de búsqueda en mapas y también visualizar mapas temáticos dinámicos.
3. Procesos de modelado y simulaciones espaciales, a través de los cuales los datos en la base de datos georreferenciada podrán ser utilizados para elaborar modelos y simulaciones espaciales explícitas.
4. El proceso de gestión y análisis de indicadores, que podrán buscar visualizar y analizar datos de indicadores de todas las instituciones del CSP, para la utilización en tomas de decisiones del MCP.
5. El proceso de Patrimonio participativo, con el cual la ciudadanía podría ingresar datos al sistema de información relacionado al patrimonio cultural y ambiental del Ecuador.

## **6 DISEÑO TÉCNICO DEL MOTOR DE BÚSQUEDA Y DEL GEOPORTAL STIIP-PILOTO**

En este capítulo se expondrán los detalles técnicos del diseño, que fueron efectivamente desarrollados luego en el prototipo de software del proyecto STIIP, en base al diseño conceptual previamente realizado. El prototipo desarrollado se lo llamará STIIP-Piloto, el cual incluye varias funcionalidades para cumplir algunos de los objetivos del proyecto STIIP.

Si bien el alcance del proyecto no incluye un desarrollo completo del software, el STIIP-Piloto desarrollado abarca los procesos de:

- Creación del metadato y generación del índice.
- Protocolo de manejo de datos y de metadatos.
- Proceso de búsqueda.
- Proceso del geoportal.

Varios de estos procesos son la base para la realización de los restantes, por lo que el código fuente (Ver Anexo 1) es lo suficientemente modular para integrarlos en un futuro.

A continuación, se detalla el diseño técnico del STIIP-Piloto.

## 6.1 Campos del metadato

El metadato del STIIP está compuesto por diferentes campos. En esta sección se especificará los campos acerca de diferentes aspectos de los datos, a los cuales el metadato queda relacionado. Antes de indicar los campos del metadato es preciso definir el sistema cartográfico base.

### 6.1.1 El sistema cartográfico base

El sistema cartográfico del STIIP se refiere al conjunto de mapas, que servirá de base para el criterio territorial de búsqueda de la información y proporcionará la dimensión semántica de los datos contenidos en los metadatos del índice del STIIP.

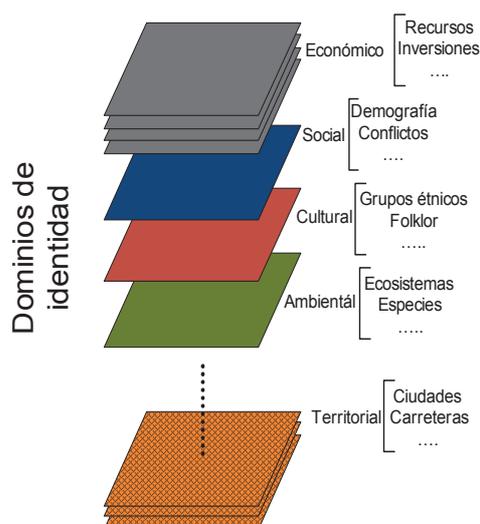


FIGURA 6-1 ELEMENTOS DEL SISTEMA CARTOGRÁFICO BASE DEL STIIP

El STIIP presentará cierto número de mapas base, que definirán el espacio de indexación territorial de la información. Los mapas base serán asimilables a las capas vectoriales de representación de los elementos geográficos del territorio (provincia, cantón, ciudad, etc.). Cada mapa contendrá elementos homogéneos desde el punto de vista semántico y serán parte integrante del dominio de identidades de la información territorial. Por ejemplo, el elemento territorial “*páramo*” es parte de un componente “*ecosistema*”, dentro del dominio de identidad “*ambiental*”, que a su vez es parte de un dominio de identidad “*territorial*”.

En la figura 6-1 [basado en 34] se representa la estructura de jerarquización de la cartografía base del STIIP. Los dominios de identidad, los componentes y los elementos relativos, serán especificados de forma más detallada en la sección de campos territoriales del metadato. Esta estructura jerárquica está diseñada para ser flexible: en cualquier momento un usuario administrador del sistema podrá definir nuevos dominios de identidad, nuevos componentes y actualizar componentes ya presentes en el sistema.

En el dominio “*territorial*” se incluirán varias capas predeterminadas, que faciliten presentar una imagen visual del territorio ecuatoriano, por ejemplo, capas de divisiones administrativas (límites de ciudades, parroquias, cantones, provincias y nacionales), elementos de infraestructura humana (carreteras, ferrocarriles, oleoductos), elementos de cartografía física (levantamientos topográficos, ecosistemas, ríos, lagunas). El número de capas de este dominio incluidas para la visualización inicial del mapa o resultados en el mapa, no será excesivamente grande, de manera que el mapa pueda desplegarse de forma rápida y permita un acceso ágil a las funciones del visor de mapas. Esto quiere decir que, el primer mapa a desplegarse contará con las mínimas capas requeridas (límites de provincias y del país) para una visualización territorial en una escala apropiada para mostrar todo el

país en la pantalla y, de esta manera evitar la saturación cognitiva si se mostrara todo el resto de capas territoriales (ríos, carreteras, cantones, ciudades, etc.). Se debe decir también que, a medida que se amplíe la escala de visualización del mapa en la pantalla, se irán visualizando las capas territoriales acordes, por ejemplo a nivel provincial se mostrarán los cantones de dicha provincia. El usuario podrá también, activar o desactivar en el mapa algunos de los componentes de los restantes dominios de identidad de la cartografía base del STIIP, con el fin de visualizar alguna información relevante para sus fines.

La cartografía base del STIIP es, entonces, representada por el conjunto de componentes o capas que se necesita para la integración de búsqueda de la información georreferenciada contenida en el índice del STIIP. Todos los datos (que provean metadatos georreferenciados) y los mapas que componen el sistema cartográfico base del geoportal, serán almacenados en la base de datos georreferenciada del STIIP.

### **6.1.2 Campos del Dublin Core**

En esta sección del metadato, se ingresará la información acerca de los tipos de contenidos de los datos, las personas responsables del mantenimiento, los formatos de almacenamiento y las posibilidades de conexión remota para el acceso a los datos. Además, se indicará la cobertura espacial y temporal, claves, idiomas, derechos y usos. Los campos del Dublin Core son:

## Descriptivos

- **Título:** el nombre dado a un recurso, habitualmente por el autor.
- **Claves:** los tópicos del recurso, típicamente expresará las claves o frases que describen el título o el contenido del recurso.
- **Descripción:** una descripción textual del recurso, que puede ser un resumen en el caso de un documento o una descripción del contenido en el caso de un documento multimedia.
- **Fuente:** secuencia de caracteres usados para identificar unívocamente un trabajo a partir del cual proviene el recurso actual.
- **Lengua:** lengua(s) del contenido intelectual del recurso.
- **Relación:** identificador de un segundo recurso y su relación con el recurso actual, que permite enlazar los recursos relacionados y las descripciones de los recursos.
- **Cobertura:** característica de cobertura espacial y/o temporal del contenido intelectual del recurso. La cobertura espacial se refiere a una región física, utilizando por ejemplo coordenadas. La cobertura temporal se refiere al espacio de tiempo para el cual fue creado el contenido del recurso, no a la fecha de creación.

## Propiedad Intelectual

- **Autor o creador:** persona u organización responsable de la creación del contenido intelectual del recurso, por ejemplo, los autores de documentos escritos, artistas o fotógrafos en el caso de recursos visuales.
- **Editor:** entidad responsable de que el recurso se encuentre disponible en la red en su formato actual.

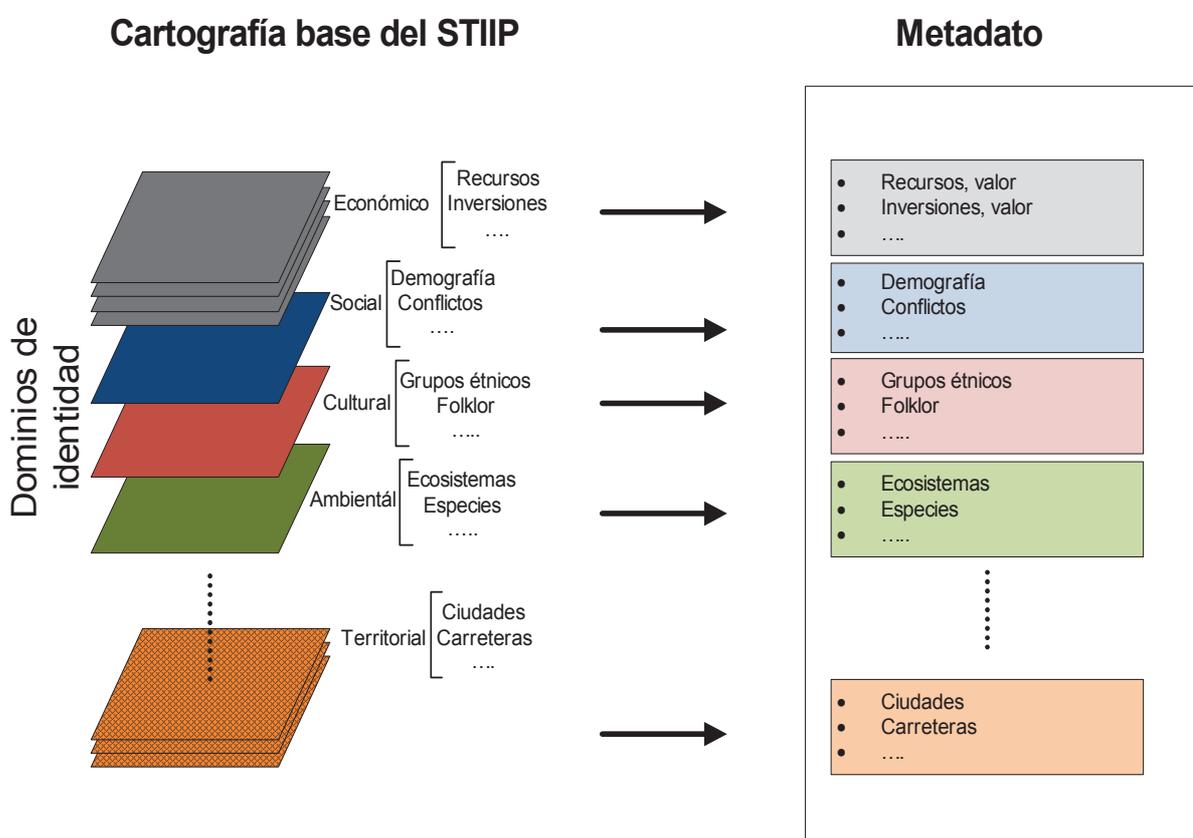
- **Otros colaboradores:** personas u organizaciones que hayan tenido una contribución intelectual significativa, pero que esta sea secundaria en comparación con las de las personas u organizaciones especificadas en el elemento autor o creador, por ejemplo, ilustrador o traductor.
- **Derechos:** referencia (por ejemplo, una URL) de vinculación hacia una nota sobre derechos de autor, un servicio de gestión de derechos o un servicio que dará información sobre términos y condiciones de acceso a un recurso.

### Instanciación

- **Fecha:** fecha de creación, en la cual el recurso se puso a disposición del usuario en su forma actual.
- **Tipo del recurso:** la categoría del recurso, por ejemplo, página personal, romance, poema, diccionario, etc.
- **Formato:** es el formato de datos de un recurso, usado para identificar el software y, posiblemente, el hardware que se necesitaría para mostrar el recurso, por ejemplo, pdf, jpg, doc, etc.
- **Identificador del recurso:** secuencia de caracteres utilizados para identificar unívocamente un recurso, depende del recurso para el tipo del identificador, por ejemplo, para recursos en línea pueden ser URLs, para otros recursos como libros puede ser usado el ISBN (International Standard Book Number).

### 6.1.3 Campos de dominio de identidad y territoriales

El metadato contendrá algunos campos, que se utilizarán para realizar una búsqueda territorial de la información recopilada por el STIIP. En particular, en los metadatos del STIIP serán definidos campos que correspondan a los componentes también definidos en la cartografía base del STIIP, como se muestra en la figura 6-2 [basado en 34].



**FIGURA 6-2 CORRESPONDENCIA ENTRE CAMPOS DE METADATOS Y COMPONENTES DE LA CARTOGRAFÍA BASE DEL STIIP**

Se ha visto que la cartografía base del STIIP, consiste en la jerarquía de componentes (capas) organizadas en dominios de identidad (ambiental, social, económico, etc.). En cada componente encontramos elementos que pertenecen o son relativos a las características del territorio de aquel. Por ejemplo, en el componente *ciudad* del dominio *territorial*, encontramos la lista de los nombres y características geográficas de las ciudades ecuatorianas, en el componente *especies* del dominio *ambiental*, encontramos las especies de seres vivos de los territorios ecuatorianos.

El metadato definido en el STIIP, será estructurado de tal manera que se presente campos correspondientes a todos los componentes incluidos en la cartografía de base del STIIP. Se debe mencionar que no todos los metadatos tendrán explícitamente llenos todos los campos: en los metadatos sobre datos de infraestructuras deportivas, en general, no presentarán explícitamente campos relativos por ejemplo a especies nativas. Varios de los campos del recurso que no tengan relación a un dominio, serán solamente considerados sin contenido o “no llenos”.

Los dominios y elementos de integración territorial, de acuerdo a la cartografía base del STIIP, se listan a continuación en la siguiente tabla:

TABLA 6-1 DOMINIOS DE IDENTIDAD Y CAMPOS DEL METADATO DEL STIIP

DOMINIO	CAMPO
Ambiental	<ol style="list-style-type: none"> <li>1. Especie, clasificación: clasificación taxonómica</li> <li>2. Especie, categoría: extinta, en peligro, vulnerable, rara.</li> <li>3. Ecosistema</li> <li>4. Geología</li> <li>5. Sismología</li> <li>6. Climatología</li> <li>7. Geografía botánica</li> <li>8. Geografía zoológica</li> <li>9. Agronomía</li> <li>10. Silvicultura</li> </ol>
Salud	<ol style="list-style-type: none"> <li>11. Sanidad pública</li> <li>12. Personal sanitario</li> <li>13. Recursos médicos</li> <li>14. Infraestructuras sanitarias</li> <li>15. Salud intercultural</li> </ol>
Cultural	<ol style="list-style-type: none"> <li>16. Artes en general</li> <li>17. Artes visuales</li> <li>18. Música</li> <li>19. Folklor</li> <li>20. Literatura</li> <li>21. Arte culinario</li> <li>22. Radiodifusión</li> <li>23. Historia</li> <li>24. Documentación en general, producción documental</li> <li>25. Centros de documentación</li> <li>26. Educación</li> <li>27. Fiestas populares</li> <li>28. Fiestas nacionales</li> <li>29. Fiestas religiosas</li> <li>30. Religiosidad</li> </ol>
Social	<ol style="list-style-type: none"> <li>31. Viviendas</li> <li>32. Etnias</li> <li>33. Idiomas</li> <li>34. Geografía social</li> <li>35. Infraestructura educativa</li> <li>36. Educación intercultural</li> <li>37. Personal educativo</li> </ol>

DOMINIO	CAMPO
Patrimonial	38. Patrimonio arqueológico 39. Patrimonio documental 40. Patrimonio inmaterial 41. Patrimonio inmueble 42. Patrimonio mueble 43. Patrimonio natural
Económico	44. Industrias 45. Minería
Deportivo	46. Infraestructura deportiva 47. Actividades deportivas 48. Deportivos 49. Excursiones, montañismo, camping 50. Oferta pública de actividades deportivas
Gubernamental	51. Proyectos y programas 52. Instituciones culturales 53. Instituciones ambientales
Territorial	54. Ciudad 55. Parroquia 56. Cantón 57. Provincia 58. Región 59. Vías de comunicación terrestres 60. Vías de comunicación marítimas 61. Vías de comunicación aéreas 62. Ingeniería hidráulica 63. Ingeniería civil 64. Cursos de agua naturales 65. Geografía física 66. Zonas de Riesgo

#### 6.1.4 Campos de otras secciones del metadato del STIIP

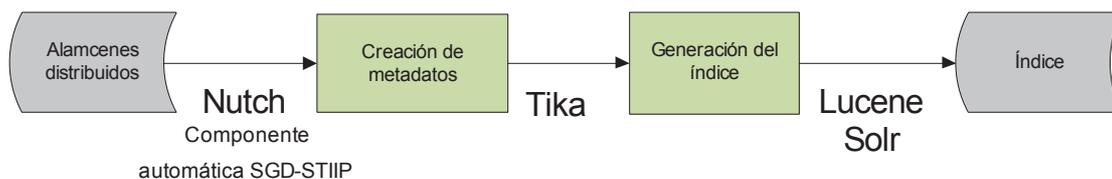
Otros campos importantes en la estructura del metadato del STIIP serán:

- El campo *texto*, donde se almacenarán integralmente todos los componentes textuales de cualquier documento recopilado por el STIIP.
- Identificadores del CSP.
- Los campos del perfil ecuatoriano de metadato geográfico.

## 6.2 Arquitectura del motor de búsqueda

Como se ha señalado en los numerales precedentes, el corazón del STIIP será un motor de búsqueda, en particular las herramientas tecnológicas que proporcionarán la creación, el manejo y la utilización del índice. Es muy importante entonces, determinar los programas de código abierto (según lo establecido en los requerimientos) que son más adecuados para ser utilizados como base del motor de búsqueda del STIIP-Piloto. Se ha seleccionado el software desarrollado por Apache Software Foundation [13] de gestión, creación y manejo del índice. Además, se ha desarrollado el software entero del motor de búsqueda del STIIP-Piloto en Java, debido a su alto desempeño y existencia de varias herramientas a usarse como las siguientes:

- Apache Lucene v.2.9
- Apache Solr v.1.4
- Apache Tika v.0.7
- Apache Nutch v.1.1



**FIGURA 6-3: APLICACIÓN DEL SOFTWARES EN LOS PROCESOS DEL STIIP**

El la Figura 6-3 [basado en 34] se representa el diagrama de aplicación del software en los procesos establecidos. A continuación se expone la descripción de lo que son y de sus funcionalidades.

### 6.2.1 Lucene: Motor de búsqueda textual

Apache Lucene [14] es una librería de motor de búsqueda y cuenta con completas funcionalidades para la búsqueda de texto, originalmente implementada en Java. Lucene está apoyado por la Apache Software Foundation y se distribuye bajo la Apache Software License (de código abierto). Lucene tiene también, versiones para otros lenguajes incluyendo Delphi, Perl, C#, C++, Python, Ruby y PHP.

Lucene es útil para cualquier aplicación que requiera indexado, búsqueda de texto completo o implementación de motores de búsquedas. Las ventajas de Lucene son su escalabilidad, alto desempeño, algoritmos de búsqueda poderosos y precisos y, multiplataforma al ser implementado en Java.

En la parte central de la arquitectura lógica de Lucene, se encuentra el concepto de documento que contiene campos de texto. Esta flexibilidad permite a Lucene ser independiente del formato del fichero. Textos que se encuentran en pdf, páginas HTML, documentos de Microsoft Word, así como muchos otros pueden ser indexados automáticamente. En definitiva, Lucene es el software usado para la construcción del índice, el corazón del STIIP.

### **6.2.2 Solr: Servidor empresarial de motor de búsqueda**

Solr [15] es servidor empresarial, para la utilización del motor de búsqueda de código abierto basado en la librería de programación Lucene. Sus principales características incluyen: APIs en XML/HTTP y JSON, resaltado de resultados, búsqueda facetada, caché y una interfaz web para su administración. Solr corre sobre un contenedor de servlets Java como puede ser Apache Tomcat.

Aunque es un proyecto público desde hace relativamente poco, lo utilizan varias páginas de alto tráfico de Internet; es decir, es utilizado en motores de búsqueda en sitios web con un gran número de visitantes en sus páginas, por ejemplo Netflix, AOL, CNET, Zappos, etc. Solr es un proyecto activo de Apache que cuenta con muchas mejoras cada nueva versión, por ejemplo, las últimas actualizaciones, al momento de construir el prototipo de software del STIIP, fueron la posibilidad de realizar búsqueda distribuida, mejoras en el indexado, la búsqueda y el facetado, a más de nuevas funcionalidades como el procesamiento de documentos enriquecidos (PDF, Word, HTML), el clustering de resultados de búsqueda y

mejoras de rendimiento en general. La versión utilizada para el STIIP-Piloto es la 1.4 y las características de Solr utilizadas para el proyecto STIIP-Piloto son:

- Utiliza la librería Lucene para la búsqueda textual.
- Navegación y búsqueda facetada.
- Presentación de resultados.
- Interfaz de administración HTTP.
- Replicación en otros servidores Solr, permitiendo búsquedas eficientes con un gran número de documentos.

La característica principal de Solr es ser un servidor HTTP de Lucene, que servirá para el manejo de extracción de información del índice y todas las operaciones de comunicación remota entre servidores distribuidos y el servidor central del MCP. Se ha utilizado el API de Java de Solr, llamado SolrJ, que provee las funcionalidades embebidas para desarrollar el motor de búsqueda en Java del STIIP-Piloto. El API SolrJ es aquel que será usado para los procesos de búsqueda, protocolos de datos y de metadatos.

### **6.2.3 Tika: Análisis de contenido textual**

Tika [16] es un subproyecto de Apache Lucene y es un software que detecta y extrae metadatos y texto estructurado desde varios documentos, utilizando librerías de análisis sintáctico preexistentes. Tika soporta la lectura y subsecuente extracción de texto desde diferentes tipos de formatos, tales como:

- HyperText Markup Language (HTML, XHTML)
- XML y derivados
- Documentos de Microsoft Office (doc, ppt, xls, etc.)
- OpenDocument (formato de OpenOffice)
- Portable Document Format (pdf)
- Electronic Publication Format (formato de publicación electrónica usado por muchos libros electrónicos)
- Rich text format (rtf)
- Formato de texto (txt)
- Formatos de audio (mp3, midi, etc)
- Formatos de imagen (jpg, jpeg)
- Formatos de video (mp4, flash)
- Formato mbox (archivos de correo electrónico)

Tika será utilizado por la araña del sistema de información, para extraer la información de los documentos almacenados en los servidores periféricos, analizar el texto e ingresarlo en forma oportuna en un metadato. En definitiva, Tika tiene las herramientas y será usado para extraer automáticamente los campos del Dublin Core de archivos en formatos diversos, además de extraer todo el texto contenido en ellos.

#### **6.2.4 Nutch: Araña y motor de búsqueda**

Nutch [17] es proyecto de un motor de búsqueda web basado en Lucene. Nutch es software libre o código abierto. A diferencia de Lucene, Nutch ofrece las funcionalidades de rastreo o *crawling* de archivos para las funcionalidades de la araña del STIIP. Está desarrollado en

Java y, basa su arquitectura en la plataforma Hadoop de desarrollo de sistemas distribuidos. Su sistema es altamente flexible y extendible a través de plug-ins. Nutch será utilizado en el STIIP como araña (*crawler*), para recopilar las informaciones de los sitios web de las instituciones del CSP.

Las funciones que ejecutará la araña que se ha descrito en los numerales anteriores, deberán ser implementadas en el STIIP, a través de un código desarrollado expresamente para este fin. El programa presentará una interfaz simple y eficaz a través de la cual los usuarios administradores del STIIP podrán:

- Ingresar las direcciones de los almacenes distribuidos de cada institución coordinada.
- Ingresar las sentencias SQL, que servirán para definir las máscaras de extracción de datos de las bases de datos, distribuidas en los almacenes de las instituciones coordinadas.
- Ingresar las máscaras, que permitirán extraer datos de las hojas de cálculo.

### **6.3 Arquitectura del geoportal**

Las funciones del proceso geoportal son ofrecer al usuario la posibilidad de visualizar datos e información en forma de mapas dinámicos. En su estructura fundamental, el proceso de geoportal no es nada más que SIG web, es decir, una aplicación capaz de leer datos desde los servicios geográficos de un servidor geográfico, conectado con una base de

datos georreferenciados y producir un correspondiente *render* de un mapa. A esta serie de operaciones base, el proceso de geoportal del STIIP, deberá acoplar otras funciones que permitirán al visualizador de mapas, conectarse con otros procesos del STIIP. Lo que quiere decir que, por ejemplo el proceso de geoportal podrá leer los resultados de las acciones del proceso de búsqueda como entrada y presentarlos sobre un mapa dinámico. Lo mismo será posible con los procesos de gestión y análisis de indicadores. Para que esto sea factible es necesario utilizar software especializado en mapas para el STIIP-Piloto.

En esta línea, se deberá construir una arquitectura integrada de programas, que nos permitan realizar un sistema compatible con los estándares más difundidos en el campo de geoservidores. Precisamente en el caso de los geoservidores del STIIP, el prototipo desarrollado deberá respetar los estándares del *Open Geospatial Consortium (OGS)*. Exactamente como se referencia los estándares de distribución de datos geográficos establecidos para el OGS, se ha identificado la arquitectura de una estructura integrada de software interrelacionado, que es reconocido como idóneo en vista de la realización de Sistemas de Información Geográfica web. Dicha arquitectura para el STIIP-Piloto estará constituida por los siguientes programas:

- PostgreSQL v.8.3 [18] – Sistema de Gestión de Base de Datos (SGBD).
- PostGIS v.1.3 [19] – Extensión Espacial PostgreSQL.
- MapServer v.5.4.2 [20] – Servidor de Servicios Geográficos en conjunto  
p.Mapper v.4.2 [21] – Estructura integrada de visualización de mapas  
PHP/MapServer. GeoServer v.1.7 [22] – Servidor de Servicios Geográficos en

conjunto con OpenLayers v.2.9.1 [23] – librería JavaScript de visualización mapas dinámicos.

En las siguientes secciones, se detallará una rápida descripción de cada uno de los mencionados programas, pero antes es conveniente mostrar la estructura global integrada de todo el software de georreferencia.

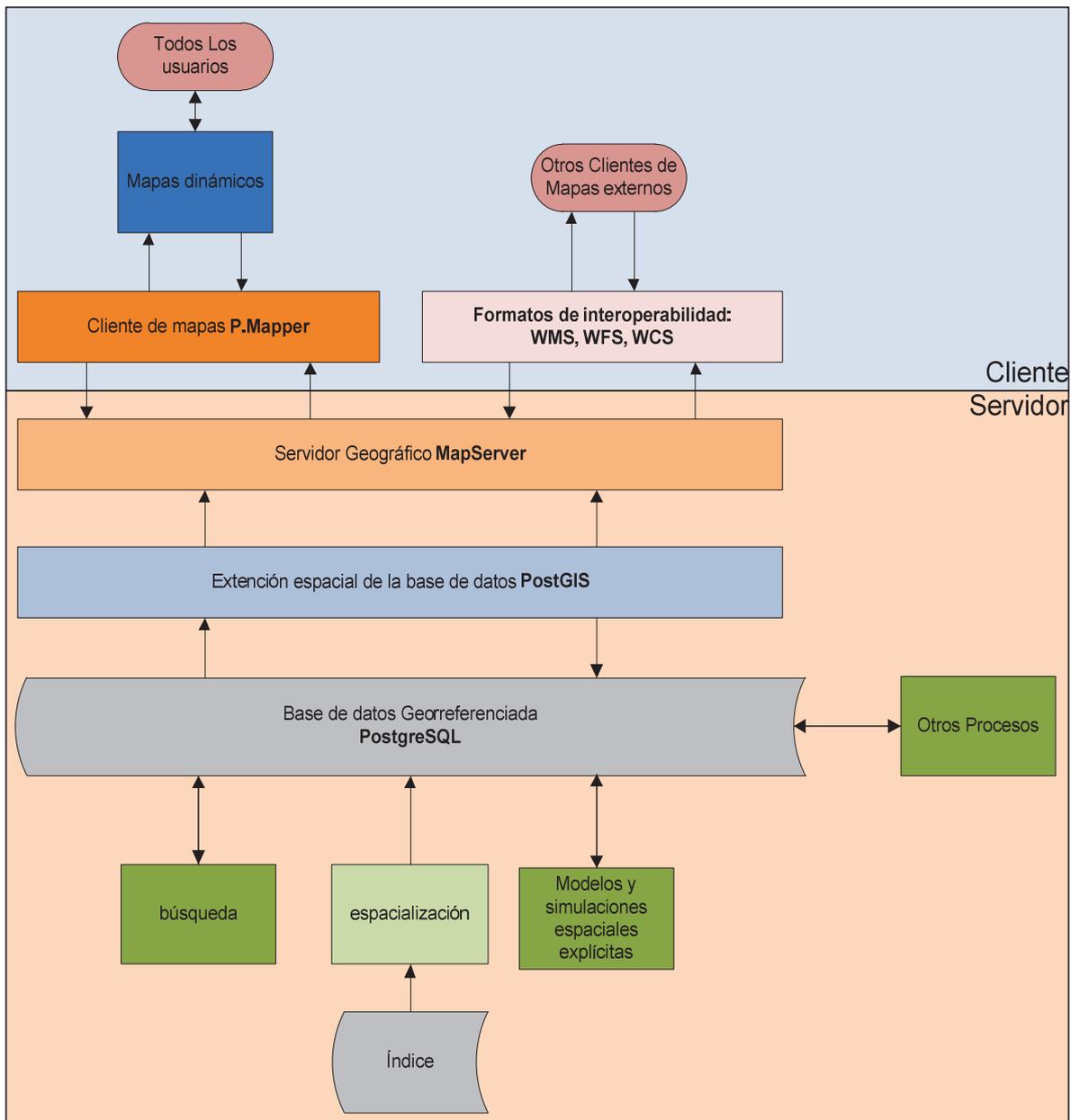
### **6.3.1 Estructura integrada de software del geoportal**

Como información de políticas de desarrollo del STIIP, podemos decir que la mejor opción para implementar el geoportal es utilizar el conjunto PostgreSQL, PostGIS y MapServer con p.Mapper. De hecho esta opción con el MapServer es la más indicada porque el MCP ya tiene una versión de geoportal que funciona con MapServer y p.Mapper. Esta será la opción a la que nos referiremos cuando hablemos de geoportal del STIIP.

Es claro que los elementos de software presentados en el numeral precedente no servirán exclusivamente para componer la interfaz del geoportal, sino que también serán parte de la estructura profunda de este proceso. Por ejemplo, para el usuario del STIIP, el uso de estas herramientas será transparente de acuerdo al modelo propuesto de desarrollo MVC; es decir, no se dará cuenta del hecho de estar utilizando el servidor de base de datos PostgreSQL y tampoco de estar utilizando el MapServer.

Los elementos de software introducidos en la sección anterior constituyen la base del software del proceso geoportal. Sin embargo, los programas seleccionados no servirían para la realización de este proceso si no se diseñara ni construyera una arquitectura integrada de relación entre ellos. El esquema de la arquitectura integrada del software que constituye el proceso geoportal está representado en la figura 6-4. La función del diseño de la arquitectura es definir la manera en la cual, los distintos software que componen el sistema informativo, considerados como unidades independientes, se interrelacionan entre ellos.

La arquitectura integrada del software corresponde a la realización del modelo físico del proceso geoportal y, en su representación esquemática en la figura 6-4; como siempre las flechas de conexión corresponden a los flujos de información a través del proceso. He aquí la importancia de seleccionar un conjunto de software, que tenga las características de ser interrelacionados e interoperables entre ellos.



**FIGURA 6-4 ARQUITECTURA INTEGRADA DE SOFTWARE  
DEL PROCESO GEOPORTAL DEL STIP**

Como se observa en la figura 6-4 [basado en 34], el proceso geoportales puede ser separado en dos partes: la parte servidor y la parte cliente, que corresponden en modo aproximado a la parte interna y de interfaz del proceso geoportales. El corazón de la parte servidor es por

supuesto el Sistema de Gestión de la Base de Datos, SGBD. A través del conjunto de software PostgreSQL-PostGIS, los datos georreferenciados se podrán crear, modificar y distribuir en forma de mapas. También en la parte servidor, se colocan las comunicaciones SGBD – MapServer. La parte cliente incluye las salidas de las aplicaciones de gestión de metadatos, las salidas de MapServer (en distintos formatos) y la producción de mapas a través del software cliente de mapas, MapServer-p.Mapper.

En la figura 6-4, podemos ver cómo otros procesos del STIIP están conectados con la base de datos georreferenciada. Además, cabe observar que la base de datos georreferenciada no será solo un lugar de almacenamiento por los resultados de análisis cumplidos en otros procesos del STIIP, pero también alojará en forma de capas, datos relativos a mapas temáticos que se presentarán en la interfaz del geoportal. Estas capas, presentarán datos de interés general acerca de aspectos geográficos de todo el Ecuador y también de aéreas específicas en particular. El usuario podrá construir también mapas dinámicos “personalizados”, en el sentido en que podrá, por ejemplo, junto a los resultados de sus búsquedas, sobreponer capas seleccionándolas en una lista de capas disponibles y cruzar así de forma muy sencilla datos en un mapa geográfico.

Antes de pasar a la descripción del software usado, es importante detallar algunas definiciones acerca de georreferencia e interoperabilidad.

### 6.3.2 Algunas definiciones y requerimientos de interoperabilidad

El acceso remoto a la visualización de información gráfica territorial, que elabora el consejo sectorial, constituye unos de los principales objetivos de la realización del STIIP para ofrecer la cartografía y el acceso a la información georreferenciada a entidades externas. La manera de implementar este acceso, es la de realizar una infraestructura que integre los estándares de interoperabilidad en los campos de intercambios de datos georreferenciados. Los estándares adoptados en el STIIP son los del OGC [24].

#### 6.3.2.1 Servicios de mapas (*Web Map Service, WMS*)

La publicación de cartografía y datos georreferenciados en el STIIP, se realiza mediante un servicio WMS, basado en las especificaciones OGC, habilitando las siguientes operaciones [25]:

1. Devolución de metadatos del servicio (*GetCapabilities*).
2. Devolución de un mapa con los parámetros geográficos y dimensiones, correctamente definidos en el interfaz de llamada WMS (*GetMap*).
3. Devolución de la información alfanumérica asociada a la entidad gráfica (*GetFeatureInfo*).

*“Se entiende por un servicio de mapas (WMS) aquél que proporciona una imagen de un mapa para un área determinada. Este es, sin duda, el servicio que más ha proliferado dentro de las organizaciones e instituciones cartográficas. [...] La especificación “Web Map Service (WMS) Implementation Specification v1.3” del Open Geospatial Consortium,*

*aprobada como ISO 19128 “Geographic Information – Web Map Server Interface” define las operaciones para obtener una descripción de los mapas ofrecidos por el servidor (GetCapabilities), obtener un mapa (GetMap) y consultar cierta información limitada sobre las entidades mostradas en el mapa (GetFeatureInfo).” [26]*

Una operación en un servidor geográfico, que tenga por ejemplo el servicio WMS activo, se realiza usando un navegador web estándar, mediante una dirección URL (*Uniform Resource Locator*), junto a una serie de parámetros, que son estandarizados por el OGC. Por ejemplo, el WMS del Estado de Massachusetts, Estados Unidos, llamado *MassGIS* [27]:

*http://giswebservices.massgis.state.ma.us/geoserver/wms?VERSION=1.1.1&REQUEST=GetMap&SERVICE=WMS&LAYERS=massgis:GISDATA.TOWNS\_POLYM,massgis:GISDATA.SCHOOLS\_PT&STYLES=Black\_Lines,&BBOX=232325.38526025353,898705.3447384972,238934.49648710093,903749.1401484597&SRS=EPSG:26986&WIDTH=570&HEIGHT=435&FORMAT=image/png*

En el servicio WMS, se publicarán los mapas que se generan en base a las entidades existentes en el STIIP y, que se engloban en varios ámbitos de datos y servicios.

### *6.3.2.2 Servicio de publicación de entidades (Web Feature Service, WFS)*

Dentro de un contexto de interoperabilidad cartográfica, en un ámbito heterogéneo de fuentes de datos georreferenciados, como con el que se podría encontrar el STIIP, es

necesario la implementación de un servicio WFS de publicación de objetos que permita recuperar entidades en formato vectorial junto a sus atributos, codificados en formato *Geography Markup Language, GML* [28]. El GML es otro estándar del OGC que permite el transporte y almacenamiento de información geográfica, en formato de lenguaje de XML.

El WFS, soporta las especificaciones OGC, *Web Feature Service Implementation Specification* versión 1.0.0 y 1.1.0. Las operaciones soportadas son:

1. Devolución de las características del servicio, tipos de fenómenos y operaciones soportadas (*GetCapabilities*).
2. Devolución de las estructuras de atributos y campos de los fenómenos que ofrece el servicio (*DescribeFeatureType*).
3. Devolución de los objetos en formato GML, en base a solicitud por atributos o consultas espaciales (*GetFeature*).

#### 6.3.2.3 Servicio de catálogo (*Web Catalog Service, WCS*)

El servicio de catálogo recoge las funcionalidades para la búsqueda de información y servicios a partir de sus metadatos, con filtros como por ejemplo, área, fecha, descripción, etc. Este servicio permite realizar búsquedas de los datos espaciales existentes y evaluar la validez y precisión de los mismos para un determinado propósito.

Los servicios WMS, WFS y CSW son los tres servicios básicos que definen el estándar OGC de interoperabilidad entre servidores geográficos. Estos estándares son los requeridos en el campo de interoperabilidad con la SENPLADES.

### **6.3.3 PostgreSQL: Sistema de gestión de base de datos**

PostgreSQL [29] es un sistema de gestión de base de datos (SGBD) relacional, orientada a objetos y de código abierto, publicado bajo la licencia BSD. PostgreSQL posee una aplicación de gestión y desarrollo llamada PgAdmin III, que ofrece la posibilidad de utilizar una interfaz gráfica, a través de la cual se puede efectuar la mayoría de las operaciones relacionadas a la gestión de PostgreSQL.

PostgreSQL es el software que, en el proceso del geoportal se ocupa de definir, almacenar construir, distribuir y manipular datos georreferenciados. Esto se realiza mediante su extensión espacial PostGIS, razón por la cual se decidió usar esta base de datos y no otras del mercado. A continuación, se describe los detalles de PostGIS.

### **6.3.4 PostGIS: Extensión espacial de PostgreSQL**

PostGIS [30] es un módulo o extensión, con licencia GNU, que añade soporte de objetos geográfico a PostgreSQL, convirtiéndola en una base de datos espacial o georreferenciada, para su utilización en Sistemas de Información Geográficas. PostGIS ha sido desarrollado por la empresa canadiense *Refraction Research*.

PostGIS ha sido certificado en el 2006 por el *Open Geospatial Consortium*, lo que garantiza la interoperabilidad con otros sistemas OGC. PostGIS almacena la información geográfica en una columna del tipo GEOMETRY, donde se puede almacenar la geometría en formato WKB (*Well-Known Binary*). Cada vez son más las aplicaciones que utilizan PostGIS como soporte geográfico a su base de datos.

PostGIS se utilizará en el STIIP-Piloto, para asociar extensiones espaciales a los datos georreferenciados. Eso permite la generación, el almacenamiento, y análisis de elementos geográficos. Los análisis se realizan a través de consultas. Para clarificar su uso, se puede decir en general que las consultas en un sistema de información espacial son de tres categorías fundamentales [31]:

- Consultas exclusivamente de propiedades espaciales, por ejemplo “buscar todos los pueblos que no están más lejos de 10 km del mar”.
- Consultas sobre propiedades no espaciales, por ejemplo “número de personas que viven en Riobamba”.
- Consultas que combinan propiedades espaciales con no espaciales, por ejemplo "determinar el número de personas que viven en unas cuadras o manzanas localizadas de Riobamba"

En conjunto PostgreSQL y la extensión PostGIS, han sido seleccionadas para la realización de la base de datos georreferenciada del STIIP-Piloto. Otra opción que se pensó usar fue GeoServer para ser implementado como geoportal en el STIIP con su aplicación OpenLayers, para visualizar mapas en un navegador web. GeoServer es un servidor

gratuito escrito en Java, que permite a los usuarios compartir y revisar los datos geoespaciales. Diseñado para ser un software de alta interoperabilidad, GeoServer publica datos originados de cualquier fuente de datos espaciales. Se decidió el uso de PostgreSQL, porque la cartografía base ecuatoriana ya había sido desarrollada en esta plataforma, por un miembro del equipo STIIP, Francesco Pizzitutti. Aclaramos que la recolección y creación de una base de datos de cartografía de un país entero, es un proyecto bastante grande que sale del alcance del presente proyecto de tesis; por lo que, se presentará los resultados del uso de dicha base de datos, sin entrar en detalle de su creación ni su diseño, sino en la aplicación que se le dio en el prototipo STIIP-Piloto.

### 6.3.5 MapServer: Servidor de servicios geográficos

MapServer [32] es un servidor de código abierto y gratuito escrito en C, que permite a los usuarios compartir y revisar los datos geoespaciales. MapServer permite crear “mapas de imágenes geográficas”, es decir, mapas que puede dirigir a los usuarios al contenido. Esta herramienta puede ser usada para la creación de mapas temáticos del STIIP.

**p.Mapper** es una estructura de integración entre PHP y MapServer, que provee diferentes aplicativos y herramientas para la visualización y la manipulación de mapas dinámicos. El conjunto de MapServer y p.Mapper está desplegado (*deployed*) en un servidor Apache. Esta herramienta será utilizada en el STIIP-Piloto, para poder utilizarlo en un ambiente web, es decir en un navegador de internet, en el cual también será desarrollado el motor de búsqueda pero con tecnología Java para internet, los JSF.

## 7 DESARROLLO DEL STIIP-PILOTO

En este capítulo se describe la etapa de desarrollo del STIIP-Piloto y todas sus funcionalidades logradas. El STIIP-Piloto como se mencionó en el capítulo anterior, está desarrollado en dos plataformas. El motor de búsqueda está desarrollado enteramente en Java; debido a la utilización de las librerías, APIs y demás herramientas tecnológicas que proveen soporte para la creación de un motor de búsqueda: Lucene, Solr, Tika. El visualizador de mapas, sin embargo, está desarrollado en PHP, igualmente por la selección del software para este componente: MapServer y p.Mapper; con la utilización de una base de datos con soporte a georreferencia: PostgreSQL con su extensión PostGIS. Para unir estas dos plataformas se utilizan archivos (de texto) que intermedian la información generada o recibida entre ellos. En los siguientes numerales se describe el proceso de funcionamiento del último prototipo desarrollado del STIIP-Piloto.

Este prototipo fue desarrollado usando la metodología RAD (desarrollo rápido de aplicaciones), que es útil cuando no se cuenta al inicio con la suficiente información para un análisis exhaustivo y definitivo, sino mediante el análisis y diseño constante tras la realización de varios prototipos. Se realizaron aproximadamente 20 prototipos en el plazo establecido para el proyecto. Cada nuevo prototipo incluyó codificación de nuevas funcionalidades, corrección de errores de programación, cambios en el diseño, integración de nuevas especificaciones, etc. En el presente capítulo se tratará la funcionalidad del último prototipo, es decir, el prototipo final que incluye todos los avances indicados anteriormente.

## **7.1 Especificaciones técnicas para el uso o desarrollo del STIIP-Piloto**

A continuación se presenta una lista detallada de los requisitos para utilizar, ejecutar o modificar el prototipo, STIIP-Piloto.

### **7.1.1 Para el usuario**

El usuario del prototipo como su nombre lo indica, es aquel que usa la aplicación sin necesidad de conocer su lógica interna. Los usuarios se definieron en el capítulo de diseño y son: la ciudadanía, los técnicos, los tomadores de decisiones y las entidades externas al CSP. Como se estableció en el modelo MVC, el usuario es aquel que está en la parte de la Vista y los requisitos para dicho usuario es tener instalado lo siguiente:

Para usar la página general:

- Navegador de Internet

Para utilizar el motor de búsqueda:

- JDK 5 y JRE.

Para utilizar el indexador de metadatos para una base de datos ajena al prototipo:

- Conexión a una base de datos local o remota: MySQL, PostgreSQL, etc.

### 7.1.2 Para el desarrollador

El desarrollador es la persona que modifica el software creado, en este caso, el autor del presente documento de tesis. El desarrollador es el que crea la parte del Modelo y Controlador del MVC. Para cualquier modificación del código del prototipo, primero se debe realizar los comandos del Inicio rápido (Servidor Apache MapServer, Servidor Glassfish, Apache Solr, PostgreSQL) y, cumplir con todos los requisitos para el usuario. Luego, los requisitos técnicos para el desarrollo del prototipo STIIP-Piloto son:

Para modificar STIIP\_PILOTO\_MOTOR\_BUSQUEDA desarrollado en Java:

- Java EE 5
- Netbeans 6.7.1
- JDK 1.6
- Plugins extras de Netbeans 6.7.1
  - Visual JSF
  - Visual JSF Runtime
  - Compatibility Web Pack
- Librerías para Java web:
  - JSF 1.1/1.2 Support:
    - jsfcl.jar
    - appbase.jar
    - dataprovider.jar
    - sqlx.jar
  - Web UI Componentes:

- webui-jsf.jar
  - commons-fileupload-1.0.jar
  - json-2.jar
  - jsf-extensions-common-0.1.jar
  - jsf-extensions-dynamic-faces-0.1.jar
- Web UI Default Theme
  - web-jsf-suntheme.jar
- Librerías extras para Solr y Tika
  - commons-codec-1.3.jar
  - commons-httpclient-3.1.jar
  - commons-io-1.4.jar
  - commons-logging-1.0.4.jar
  - errorhandler.jar
  - geronimo-stax-api\_1.0\_spec-1.0.1.jar
  - jcl-over-slf4j-1.5.5.jar
  - jxl.jar
  - mysql-connector-java-5.1.6-bin.jar
  - postgresql-8.4-701.jdbc4.jar
  - slf4j-api-1.5.5.jar
  - slf4j-simple-1.5.5.jar
  - solr-solrj-1.4.0.jar
  - stax-api-1.0.1.jar
  - tika-app-0.7.jar
  - tika-bundle-0.7.jar
  - tika-core-0.7.jar

- tika-parsers-0.7.jar
- wstx-asl-3.2.7.jar

Para modificar el sitio web del geoportal/mapas desarrollado en PHP:

- Editor de textos.

Para modificar la base de datos georreferenciada PostgreSQL:

- Administrador de bases de datos, por ejemplo pgAdmin III.

## 7.2 Motor de búsquedas y geoportal

El motor de búsqueda está desarrollado enteramente en Java, bajo los principios de programación de manejo de sesiones, *beans*, JavaServer Faces (JSF), entre otros.

Se debe recalcar que el motor de búsqueda de Lucene contiene internamente las funciones necesarias para un óptimo proceso de búsqueda; es decir, cuenta con funciones automáticas de conteo invertido de frecuencias de términos sobre el total de documentos y cálculo de pesos de cada metadato para la presentación en los resultados de búsqueda. Si bien el alcance del presente documento de tesis no contempla el estudio de algoritmo de búsqueda interno del Lucene, en el capítulo de resultados se presentará el rendimiento del uso de este.

## 7.2.1 Archivos de configuración

### 7.2.1.1 Definición del *schema.xml*

El archivo *schema.xml* es aquel que contiene las especificaciones de los metadatos para su utilización en el índice; es decir, los meta-metadatos. En este archivo se incluyen los campos del metadato especificados en la fase de diseño, aplicados al índice en Lucene/Solr. Este archivo contiene las especificaciones de todos los campos del metadato base del STIIP en las etiquetas `<field>`, por ejemplo para el campo 'cultural\_musica':

```
<field name="cultural_musica" type="text" indexed="true" stored="true"
multiValued="true"/>
```

También se incluyen los campos dinámicos `<dynamicField>`, para la indexación de campos por defecto. Además, se incluye el esquema de funcionamiento de la generación del índice. Por ejemplo, existe un filtro para tratar los espacios en blanco:

```
<tokenizer class="solr.WhitespaceTokenizerFactory"/>;
```

Otro filtro para tratar mayúsculas y minúsculas sin diferenciación:

```
<filter class="solr.LowerCaseFilterFactory"/>;
```

Palabras comunes (de, por, el, la, etc.) que no deben ser indexadas e indicadas en el archivo *stopwords.txt*:

```
<filter class="solr.StopFilterFactory" ignoreCase="true" words="stopwords.txt"
enablePositionIncrements="true"/>
```

En definitiva, el archivo de configuración de Solr es el que maneja la lógica de la generación de metadatos para el índice. El CSP proveyó una muestra de 50 documentos de patrimonio ecuatoriano en diferentes formatos (xls, doc, pdf) y una base de datos, para la generación de un índice de muestra para realizar pruebas del STIIP-Piloto. Su ubicación en el CD es:

```
\\Instaladores\Solr\apache-solr-1.4.0\apache-solr-1.4.0\example\solr\conf\schema.xml
```

#### 7.2.1.2 Definición del manejador de búsqueda “DisMax”

El archivo de configuración *solrconfig.xml* del Solr, especifica el funcionamiento del motor de búsqueda. Se utiliza el manejador llamado *DisMaxRequestHandler*, DisMax, que es uno de los algoritmos de búsqueda que provee nativamente Lucene. Este manejador incluye las funcionalidades comunes de un motor de búsqueda moderno, por ejemplo, la utilización de comillas para agrupar frases, utilización de +/- para denotar cláusulas mandatorias u opcionales, entre otras [33]. Este algoritmo de búsqueda también contiene funciones para potenciar búsquedas o resultados específicos; por ejemplo, si se desea que los archivos del Ministerio del Ambiente queden clasificados en la lista de resultados, por encima de los archivos de otro ministerio, se deberá modificar el manejador DisMax, dándole un peso más significativo a los primeros. En el algoritmo DisMax usado para el prototipo STIIP-

Piloto se definió que, todas las instituciones del CSP tendrán el mismo peso en la lista de resultados tras una búsqueda. Su ubicación en el CD es:

```
\Instaladores\Solr\apache-solr-1.4.0\apache-solr-1.4.0\example\solr\conf\solrconfig.xml
```

### 7.2.2 Integración del motor de búsqueda con el geoportal

Se debe mencionar que la integración de las dos plataformas Java y PHP se realiza en este punto. Como se mencionó anteriormente, el motor de búsqueda está desarrollado con Lucene y SolrJ (API de Solr para Java) y el visualizador de mapas con MapServer y p.Mapper (PHP). Cuando se realiza una búsqueda, el software STIIP-Piloto utiliza primero SolrJ para realizar la búsqueda en el índice y presentar los resultados en un navegador web, como se ha visto en los diferentes tipos de búsquedas. Al terminar la búsqueda, se crea un archivo temporal XML, que contiene los resultados con todos los archivos que tengan algún campo territorial o georreferenciado, para su visualización en el mapa. El archivo XML será leído por el MapServer, para dibujar las correspondientes capas de resultados sobre la cartografía base del STIIP. Este archivo temporal se guardará en la carpeta temporal del sistema operativo (en el caso de Windows: C:\Users\USER\_NAME\AppData\Local\Temp) con el nombre *stiiip\_map.xml* por cada nueva búsqueda realizada y, tendrá un formato como el siguiente:

```

<map>
  <doc>
    <query>PALABRAS_CLAVES_DE_LA_BÚSQUEA</query>
    <titulo>TÍTULO_DEL_DOCUMENTO</titulo>
    <icon_type>NÚMERO_DEL_TIPO_DE_ÍCONO</icon_type>
    <url>UBICACIÓN_DEL_DOCUMENTO</url>
    <territorial_xxx1>VALOR_TERRITORIAL</territorial_xxx1>
    <territorial_xxx2>VALOR_TERRITORIAL</territorial_xxx2>
    <territorial_xxx3>VALOR_TERRITORIAL</territorial_xxx3>
    ...
  </doc>
  <doc>
    ...
  </doc>
  ...
</map>

```

Este es el proceso de espacialización mencionado con las tablas fantasmas en la fase de diseño, concretizado con este archivo temporal XML.

## **8 RESULTADOS**

### **8.1 Rendimiento del motor de búsqueda**

El uso del API de Apache Solr, hizo más eficiente el uso de búsquedas que un motor de búsquedas que contiene una base de datos tradicional (MySQL, PostgreSQL, SQL Server). El modelo de indexación del Solr es mediante el uso de tablas de dispersión con índices variables.

Los resultados de búsqueda obtenidos con el uso del motor de búsqueda Lucene, están en el orden de magnitud de 10 microsegundos por número de documentos indexados. El mayor tiempo de espera (tiempo del motor de búsqueda en una consulta, más tiempo de despliegue de los resultados en el navegador) registrado para cualquier búsqueda con el índice de muestra fue de aproximadamente 0.5 segundos, con un total de documentos indexados de 56757 (proveídos por el CSP), los cuales registran más de 1.2 millones de términos únicos. Debido a la escalabilidad de Apache Lucene, la variación en la que cambia la velocidad de búsqueda, se incrementa a medida que aumenta el número de documentos y, se encuentran más términos frecuentes a los indexados, esto por el uso interno de tablas de dispersión del Lucene.

### **8.2 Implementación posterior del STIIP**

Como se mencionó en la introducción del presente documento de tesis, la consultoría STIIP del MCP tenía como objetivo sentar las bases (diseño conceptual y prototipo) para

un futuro proyecto del mismo ministerio, que contemple la implementación total de la consultoría antes mencionada. Este nuevo proyecto se estimaba, a la fecha de finalización de la consultoría de la CTT-USFQ, que durase aproximadamente dos años, divididos entre la implementación completa del geoportal (en base al prototipo STIIP-Piloto realizado) y la recolección de datos georreferenciados de los bienes patrimoniales de todo el país. La labor de mayor duración del plazo sería la recolección de datos patrimoniales de todo el país, pues se necesitaría contratar grupos de personas oriundas del lugar o regiones del Ecuador para que obtengan informaciones de acuerdo al diseño conceptual del STIIP, para luego poder procesarlas en el motor de búsqueda e índice. A la fecha de redacción del presente documento, está funcionando el geoportal desarrollado en la consultoría posterior a la de la CTT-USFQ, en la página web del MCP: [www.geopatrimonio.gob.ec/visor/index.html](http://www.geopatrimonio.gob.ec/visor/index.html).

Este geoportal entre otros aspectos, cuenta con datos del patrimonio cultural material (mueble, inmueble, natural), inmaterial (tradiciones orales, artes del espectáculo) y arqueológico (yacimientos, colecciones); aún los datos de varias entidades, como lenguas o deportes del país, por ejemplo, falta recopilar o incluirlos dentro del geoportal (ver Anexo 5). La continuación del proyecto STIIP, demuestra que el diseño conceptual y el prototipo de la consultoría de la CTT-USFQ (razón del presente documento de tesis), fueron un éxito y sirvieron de guía tanto para el desarrollo completo del proyecto, así como también para la puesta en producción efectiva del STIIP en la página web del MCP.

## 9 CONCLUSIONES

- Del desarrollo del STIIP, se concluye que la organización de la información en un sistema de información es de vital importancia para el mejoramiento y control en la administración de todas las entidades públicas y privadas. En el caso del STIIP, beneficiará al óptimo desempeño del Ministerio Coordinador de Patrimonio del Ecuador, sus entidades coordinadas en el Consejo Sectorial del Patrimonio, otras entidades externas y a la ciudadanía en general, por ejemplo, en las tareas de toma de decisiones, transparencia de la información, participación ciudadana, etc.
- El STIIP no pretende imponer un protocolo único y rígido del manejo de datos en todas las instituciones del CSP, de tal manera que invalide cualquier sistema que se esté usando. Además, el MCP por su mismo nombre es una entidad coordinadora, no regidora, por lo que el proyecto STIIP busca adaptarse a los sistemas de organización preexistentes en cada institución e integrarlos, de manera que se genere información valiosa tras la unión de todos ellos. El diseño conceptual y el prototipo desarrollado son los cimientos para esta integración.
- Es necesario definir dentro de cada institución coordinada, un protocolo estándar de manejo de datos, lo cual implica que se deberá decidir de forma definitiva los lugares y las maneras en que se irán almacenando los datos relevantes al propósito integrador del patrimonio del STIIP.

- Como se obtuvo en los resultados, el uso de Apache Solr/Lucene benefició al prototipo de software para aumentar la eficacia de búsqueda. Con respecto a la parte técnica, el uso de herramientas de software libre fue ideal para el desarrollo de calidad y específico para las necesidades del usuario, en este caso el MCP.
- La elaboración del diseño conceptual y especialmente la realización del prototipo de software STIIP-Piloto, servirán de base para la planeación del proyecto completo de implementación del Sistema Territorial Integrado, con una duración aproximada de dos años. El mencionado proyecto incluiría la recolección de datos en digital, de todo archivo relacionado con el patrimonio del país. El prototipo de software desarrollado permitirá, principalmente, la difusión del proyecto en esta fase posterior por la materialización del diseño conceptual en un producto funcional que se puede mostrar o probar.

## 10 RECOMENDACIONES

- Los resultados del STIIP se verían opacados si no se implementan principios de cultura organizacional en las diferentes instituciones del Consejo Sectorial del Patrimonio. Por lo tanto, se recomienda que con la introducción del software, se creen los incentivos necesarios para la utilización del mismo, ya que el STIIP, no es un sistema destructor de otros sistemas de información en producción, la introducción de aquel no es un cambio brusco para los usuarios del CSP, pero sí exigirá un modelo de organización que permita con el tiempo, la creación de estándares para la integración.
- Dentro del alcance del presente documento de tesis, el prototipo final desarrollado toma en cuenta las funcionalidades básicas del proyecto STIIP. Las proyecciones futuras del software a ser desarrolladas incluirán módulos para que cualquier persona pueda ingresar al portal web y añadir archivos, fotos, multimedia de patrimonio del país. Esto para cumplir el aspecto de participación ciudadana, llamado Patrimonio participativo.
- Otro aspecto de mejoras al software del STIIP, es el Procesamiento del Lenguaje Natural (PLN). En el caso que se quiera extender el alcance del SGD-STIIP parte automática, se podría añadir al algoritmo de la araña, un procedimiento de análisis textual basado en las más avanzadas técnicas de minería de datos y minería de textos. Estas técnicas permiten según varias orientaciones, clasificar un documento en base a la distribución o frecuencia de palabras dentro del texto contenido en el documento. La ventaja de esto es el poder catalogar textos de forma automática o semiautomática. En una fase futura de implementación del sistema, es posible diseñar algoritmos más refinados

en fineza, que permitan un análisis textual profundo, de tal manera que sea posible llenar algunos campos más específicos del metadato de forma completamente automatizada. Las herramientas de PLN a la fecha, no son muy precisas o desarrolladas, por lo que también se recomienda la mejora en la creación de máscaras automáticas para la extracción de la información en el STIIP.

- En la misma línea de recomendaciones para las mejoras del software, está el análisis de imágenes. Este punto es necesario en el CSP, pues hay muchos archivos que se encuentran en formato físico (papel) solamente y, su digitalización implicaría una amplia variedad de resultados y precisión en las búsquedas del sistema de información territorial propuesto. Esto serviría en el caso en que se encuentren almacenes de datos en los cuales haya una gran cantidad de archivos con textos en forma de imágenes, por ejemplo, aquellos documentos en papel en proceso de escaneo y digitalización con alto contenido en texto.
- El desarrollo de la herramienta de la araña Nutch, quedó en sus etapas iniciales en el último prototipo del STIIP-Piloto, por lo que se recomienda la continuación del desarrollo de las siguientes operaciones de la araña:
  - ¿Con qué frecuencia la araña deberá visitar cada uno de los almacenes de datos de las instituciones coordinadas?
  - ¿Cuándo la araña deberá visitar cada uno de los almacenes de datos de las instituciones coordinadas (en la noche, en los fines de semana)?
  - ¿Qué hacer con los datos recuperados: crear copias de seguridad o no, almacenar en series históricas, etc.?

- La última recomendación se enfoca en el diseño técnico del motor de búsqueda. Como se especificó en el capítulo del desarrollo del STIIP-Piloto, la utilización del manejador DisMax controla el algoritmo de búsqueda y presentación de resultados. Se mantuvo un algoritmo que dota del mismo peso a todas las instituciones del CSP, pero se recomienda reestructurar dicho manejador, DisMax, para dar más relevancia a áreas de mayor prioridad del CSP. En el algoritmo del DisMax usado para el prototipo STIIP-Piloto, se definió que todas las instituciones del CSP, tendrán el mismo peso en la lista de resultados tras una búsqueda, pero de acuerdo a la Agenda Sectorial del MCP, se debería adecuarlo para que los resultados del sistema informativo vayan acorde con esta agenda.

## 11 BIBLOGRAFÍA

- [1] Wikipedia. *Modelo Vista Controlador*. Obtenido en línea el 26-04-2012. Disponible en:  
<[http://es.wikipedia.org/wiki/Modelo\\_Vista\\_Controlador](http://es.wikipedia.org/wiki/Modelo_Vista_Controlador)>
- [2] Wikipedia. *Dato*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://es.wikipedia.org/wiki/Dato>>
- [3] Wikipedia. *Información*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://es.wikipedia.org/wiki/Información>>
- [4] gis.com. *What is GIS?* Obtenido en línea el 26-04-2012. Disponible en:  
<<http://www.gis.com/content/what-gis>>
- [5] Ministerio de Cultura del Ecuador. *Información general*. Obtenido en línea el 26-04-2012. Disponible en:  
<[http://www.ministeriodecultura.gob.ec/publicaciones/cat\\_view/123-nueva-prueba.html](http://www.ministeriodecultura.gob.ec/publicaciones/cat_view/123-nueva-prueba.html)>
- [6] Ministerio del Ambiente. *Ministerio*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://www.ambiente.gob.ec/?q=node/11>>
- [7] Ministerio del Deporte. *¿Quiénes somos?* Obtenido en línea el 26-04-2012. Disponible en:  
<<http://deporte.gob.ec/quienes-somos>>
- [8] VYAVHARE, Amol. *Top 3 Internet Search Engines – A Comparison*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://www.buzzle.com/articles/comparing-the-top-internet-search-engines.html>>
- [9] Wikipedia. *Gestión Documental*. Obtenido en línea el 26-04-2012. Disponible en:  
<[http://es.wikipedia.org/wiki/Gesti%C3%B3n\\_documental](http://es.wikipedia.org/wiki/Gesti%C3%B3n_documental)>
- [10] Wikipedia. *Diagrama de Flujo de Datos*. Obtenido en línea el 26-04-2012. Disponible en:  
<[http://es.wikipedia.org/wiki/Diagrama\\_de\\_Flujo\\_de\\_Datos](http://es.wikipedia.org/wiki/Diagrama_de_Flujo_de_Datos)>
- [11] Wikipedia. *Metadato*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://es.wikipedia.org/wiki/Metadato>>
- [12] Dublin Core Metadata Initiative. *Dublin Core Metadata Element Set, Version 1.1*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://dublincore.org/documents/dces/>>
- [13] The Apache Software Foundation. *Foundation Project*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://apache.org/foundation/>>

- [14] Apache Lucene. *Welcome to Apache Lucene*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://lucene.apache.org/index.html>>
- [15] Apache Solr. *Apache Solr*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://lucene.apache.org/solr/>>
- [16] Apache Tika. *Apache Tika – a content analysis toolkit*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://tika.apache.org/>>
- [17] Apache Nutch. *Welcome to Apache Nutch*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://nutch.apache.org/>>
- [18] PostgreSQL. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://www.postgresql.org/>>
- [19] PostGIS Refrations. *What is PostGIS?* Obtenido en línea el 26-04-2012. Disponible en:  
<<http://postgis.refrations.net/>>
- [20] MapServer. *Welcome to Mapserver*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://mapserver.org/>>
- [21] P.Mapper. *A MapServer PHP/MapScript Framework*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://www.pmapper.net/>>
- [22] GeoServer. *Welcome*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://geoserver.org/display/GEOS/Welcome>>
- [23] OpenLayers. *OpenLayers: Free Maps for the Web*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://openlayers.org/>>
- [24] Open Geospatial Consortium. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://www.opengeospatial.org/>>
- [25] BOLÍVAR, Miguel. *Desarrollo e implantación de un Geoportal y de servicios de Infraestructura de Datos Espaciales en el Ayuntamiento de Barcelona*. IDEE: Barcelona, Obtenido en línea el 26-04-2012. Disponible en:  
<[http://www.idee.es/resources/presentaciones/JIDEE08/ARTICULOS\\_JIDEE2008/articulo9.pdf](http://www.idee.es/resources/presentaciones/JIDEE08/ARTICULOS_JIDEE2008/articulo9.pdf)>

- [26] ROMERO, Emilio y RODRÍGUEZ, Antonio (2006). *Recomendaciones para la creación y configuración de servicios de mapas*. IDEE. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://www.01.idee.es/resources/recomendacionesCSG/RecomendacionServicioMapas.pdf>>
- [27] Office of Geographic Information (MassGIS). Obtenido en línea el 26-04-2012. Disponible en:  
<<http://www.mass.gov/mgis/massgis.htm>>
- [28] Wikipedia. *Geography Markup Language*. Obtenido en línea el 26-04-2012. Disponible en:  
<[http://es.wikipedia.org/wiki/Geography\\_Markup\\_Language](http://es.wikipedia.org/wiki/Geography_Markup_Language)>
- [29] PostgreSQL-es. *Sobre PostgreSQL*. Obtenido en línea el 26-04-2012. Disponible en:  
<[http://www.postgresql.org.es/sobre\\_postgresql](http://www.postgresql.org.es/sobre_postgresql)>
- [30] Wikipedia. *PostGIS*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://es.wikipedia.org/wiki/PostGIS>>
- [31] Wikipedia. *Base de datos espacial*. Obtenido en línea el 26-04-2012. Disponible en:  
<[http://es.wikipedia.org/wiki/Base\\_de\\_datos\\_espacial](http://es.wikipedia.org/wiki/Base_de_datos_espacial)>
- [32] apServer. *Bienvenidos a MapServer*. Obtenido en línea el 26-04-2012. Disponible en:  
<<http://mapserver.org/es/index.html>>
- [33] Lucid Imagination (2010). *What's a "DisMax"?* Obtenido en línea el 26-04-2012. Disponible en:  
<<http://www.lucidimagination.com/blog/2010/05/23/whats-a-dismax/>>
- [34] Ministerio Coordinador de Patrimonio (MCP) del Ecuador. *Diseño conceptual y estrategias de desarrollo e implementación del STIIP*. Reporte final de consultoría del MCP elaborado por el Centro de Transferencias y desarrollo de Tecnología (Carlos Mena, Francesco Pizzitutti, Citlali Galindez y Erick Bastidas) de la Universidad San Francisco de Quito. 31-05-2010

## **12 GLOSARIO DE TÉRMINOS MÁS USADOS**

CSP: Consejo Sectorial del Patrimonio.

DFD: Diagrama de Flujo de Datos.

MCP: Ministerio Coordinador de Patrimonio.

SENPLADES: Secretaría Nacional de Planificación y Desarrollo.

SGD: Sistema de Gestión Documental.

SI: Sistema de Información.

SIG: Sistema de Información Geográfica.

SNI: Sistema Nacional de Información.

STIIP: Sistema Territorial de Información Integrada del Patrimonio.

STIIP-Piloto: Prototipo de software (motor de búsqueda y geoportal) desarrollado dentro del proyecto STIIP.

## 13 ANEXOS

### Anexo 1. Código fuente

#### 1. Código del motor de búsqueda (en java)

La Estructura del código, de acuerdo al modelo MVC desarrollado, se explicará brevemente las clases de la aplicación que aproximadamente cuenta con más de 7000 líneas de código. Su ubicación en el CD es:

```
\Instaladores\Glassfish v3 prelude\STIIP_PILOTO_MOTOR_BUSQUEDA\
```

Clases del Modelo:

- **ApplicationBean1.java:** Es un *Application Enterprise JavaBean (EJB)* que se encarga de manejar las variables globales de la aplicación, es decir, las variables usadas por cualquier usuario de la aplicación concurrentemente. Aquí se especifica la dirección URL del Solr, la dirección local (en el servidor central) de los archivos del Solr y se inicializa la variable *CommonsHttpSolrServer*, que es aquella que se iniciará al momento de correr la aplicación y contiene el servidor de Solr.
- **FacetSessionBean.java:** Es un *Session EJB* que se encarga de manejar las listas facetadas de los resultados de búsqueda. Se especifica aquí la dirección del archivo

*schema.xml*, para luego extraer los campos del metadato base definidos en este archivo, almacenarlos en un árbol de java (*DefaultMutableTreeNode*) y su posterior utilización o transversación en las clases que se manejen los resultados de búsqueda asociados a una lista facetadas de categorías

- **RequestBean1.java:** Es un *Request EJB* que se encarga de procesar los pedidos entre la Vista y el Modelo, en MVC, es decir, por cada pedido GET/POST de un usuario, este JavaBean se encarga de hacer el llamando a los métodos de las clases del Controlador, empaquetar la información en un JavaBean y enviarlo a una página JSP de la Vista.
- **SessionBean1.java:** Es un *Session EJB* que se encarga de manejar las sesiones de cada usuario (o navegador) que utilicen la aplicación de manera concurrente. En esta clase se inicializa una instancia del servidor de Solr (*CommonsHttpSolrServer*) por cada usuario que se encuentre utilizando el buscador. Aquí se realizan por lo tanto, las tareas de indexación (*commitImmediately*) de archivos en directorios, hojas de cálculo o bases de datos en el servidor de Solr, de cada usuario. Además cuenta con métodos generales de ayuda para el funcionamiento de la aplicación, por ejemplo crear un archivo XML de un buffer, útil en la integración del motor de búsqueda y los mapas (visto en la sección 7.2).
- **SolrBean.java:** Es un *Session EJB*, extendido el *SessionBean1*, que realiza la lógica de todas las funciones del motor de búsqueda Solr. Esta clase tiene métodos para exportar la información a las clases del Controlador del motor de búsqueda (*Search.java*) e ir

presentando en las diferentes clases JSP de las Vistas. En concreto se tiene el método de extracción de los resultados de una búsqueda en vectores, manejar los árboles de las listas facetadas, manejar los campos de los metadatos utilizados en la búsqueda, se ejecutan los queries de una búsqueda de un usuario en el servidor Solr (*query*, *filterQuery*). Finalmente, en esta clase se crea el archivo xml en la carpeta temporal del sistema operativo usado, con los datos de los resultados de una búsqueda, a ser leídos por el geoportal para la visualización de los resultados georreferenciados en el mapa.

- **TikaBean.java:** Es un *Session EJB*, extendido el *SessionBean1*, que tiene la lógica de todas las funciones de Tika. Aquí se analizan los todos los archivos en directorios (que no sean hojas de cálculo ni bases de datos) para su indexación, mediante las librerías de Tika. Más en concreto, se extraen los campos del Dublin Core, el campo texto y se cuenta con los métodos de agregación al índice de Lucene, usados en el Controlador en el sitio de Preferencias.

Clases del Controlador:

- **CSVView.java:** En esta clase se tiene el código para el manejo de la indexación de hojas de cálculo. Primero, se crea una tabla de visualización con los datos de la hoja de cálculo y se pueblan las listas *drop-down* con los campos del metadato base del STIIP, para cada columna de la hoja de cálculo a indexar. Segundo, maneja los metadatos con los campos de identificación de una hoja de cálculo: nombre, descripción, tipo/ícono. Por último, se tiene el botón de “Añadir CSV” que es aquel que finalmente, agrega los

datos identificadores de la hoja de cálculo más las columnas a indexar dependiendo de la selección hecha en las listas *drop-down*; esta clase además se encarga de indexar la dirección URL del archivo o la dirección local.

- **CachedView.java:** Esta clase se encarga de poblar en una tabla los datos de un documento del índice que se seleccionó de la lista de resultados de una búsqueda. Se diferencia los tipos de documentos en archivos generales, bases de datos u hojas de cálculo. Para el primer tipo, se despliega el cache del campo *texto* del metadato del documento seleccionado; para los dos últimos tipos se despliega el cache del texto de las tablas de las hojas de cálculo o bases de datos. Esta página se despliega después de hacer clic en el link “Cached” en algún resultado de una búsqueda.
- **ConvertCSV.java:** Es una clase de ayuda que simplemente se encarga de convertir un archivo de formato xls, en un buffer de texto con formato csv; para su procesamiento en las clases de agregación de hojas de cálculo en el índice (*CSVView.java*).
- **DatabaseView.java:** Esta clase es similar a la clase que maneja las hojas de cálculo, pero ahora para bases de datos, ya que ambas trabajan con tablas (columnas, filas). A diferencia de añadir la totalidad de una hoja de cálculo, esta clase tiene la opción de ingresar sentencias SQL para crear vistas o máscaras de una base de datos seleccionada. Con la opción de realizar un query SQL, se crea una tabla de visualización con los resultados de este query y se pueblan las listas *drop-down* con los campos del metadato base del STIIP, para cada columna de la tabla de resultado de la query a indexar. Luego, se maneja los metadatos con los campos de identificación de una base de datos: nombre, descripción, tipo/ícono. Por último, se tiene el botón de

“Añadir Vista” que es aquel que finalmente, agrega los datos identificadores de la vista de la base de datos más las columnas a indexar, dependiendo de la selección hecha en las listas *drop-down*. Esta clase además se encarga de indexar la dirección `JDBC_URL` de la base de datos, sea local o remota; y también tiene la opción de eliminar del índice las vistas previamente indexadas.

- **DatabaseExtractor.java:** Es una clase de ayuda que se encarga de realizar las tareas de la conexión con una base de datos, ejecución de una sentencia de SQL, envío de datos de una vista de una base de datos a la instancia del servidor de Solr (*CommonsHttpSolrServer*) para su indexación.
- **Download.java:** Esta clase se encarga únicamente de hacer un llamado al escritorio de la máquina, para abrir la aplicación por defecto y apropiada para desplegar un documento descargado del servidor de motor de búsqueda. Por ejemplo, si se descargó un archivo xls, se debe abrir MS Excel; o si se descargó un archivo pdf se debe abrir el Acrobat Reader y ver los documentos originales descargados del servidor central.
- **MetadatadaView.java:** Esta clase se encarga de desplegar en una tabla, todos los metadatos de un documento en la lista de resultados de una búsqueda. Esta página se despliega después de hacer clic en el link “Metadato” en algún resultado de una búsqueda.
- **Preference.java:** Esta clase contiene todo el código para controlar la pantalla de preferencias o sitio de administración. Se divide en tres: 1) La sección Directorios, en

donde se ingresa la dirección desde la raíz del disco duro de una carpeta, para realizar la indexación de todos los archivos de un directorio y sus subcarpetas mediante el manejador TikaBean; 2) la sección Bases de Datos, en donde se ingresan las credenciales para la conexión a una base de datos, para posteriormente pasar a la pantalla de DatabaseView para realizar los procesos de indexación de vistas de bases de datos ya mencionados y; 3) la sección de Hojas de Cálculo, en donde se ingresa la dirección desde la raíz del disco duro de una hoja de cálculo (xls o csv), posteriormente pasar a la pantalla de CSVView para realizar los procesos de indexación de hojas de cálculo ya mencionados.

- **Search.java:** Esta clase contiene la mayor parte del controlador del motor de búsqueda Solr. Mediante el modelo SolrBean, esta clase extrae la información de los queries hechos en el servidor Solr para presentar los resultados en el JSP de la vista Search.jsp. aquí se realiza la población de los resultados en las listas de resultados de una búsqueda, como también de las listas de categorías facet. Este controlador se encarga además de distinguir entre los tipos de búsqueda generales, específicas, avanzadas o facetadas (ver Anexo 8), que se puede realizar para el procesamiento apropiado al algoritmo de búsqueda.
- **faces-config.xml:** Este es un archivo de configuración en donde se especifica las reglas de navegación entre las páginas JSP y además, se especifica el nombre, la clase y el alcance de cada JavaBean.

- **sun-web.xml**: Descriptor para el despliegue de la aplicación en el servidor Glassfish (*Sun Deployment Descriptor*).
- **web.xml**: Descriptor para el despliegue de una aplicación web (*Web Application Deployment Descriptor*), configuraciones del servlet, sesiones, errores de páginas.

#### Clases de la Vista:

- **CSVView.jsp**: Pantalla de vista de la configuración de indexación de una hoja de cálculo. Vista controlada por la clase *CSVView.java*.
- **CachedView.jsp**: Pantalla de vista de los datos en cache de un resultado de una búsqueda. Vista controlada por la clase *CachedView.java*.
- **DatabaseView.jsp**: Pantalla de vista de la creación de vistas o máscaras en una base de datos. Vista controlada por la clase *DatabaseExtractor.java*.
- **Download.jsp**: Pantalla de vista cuando se ha descargado un documento de un resultado de una búsqueda. Esta pantalla es vacía en el navegador, ya que se debe abrir la aplicación del archivo, por ejemplo si es un archivo .xls, se debe abrir MS Excel. Vista controlada por la clase *Download.java*.

- **MetadatadaView.jsp**: Pantalla de vista de los metadatos de un resultado de una búsqueda. Vista controlada por la clase *MetadatadaView.java*.
- **Preference.jsp**: Pantalla de vista del sitio de administración o preferencias. Vista controlada por la clase *Preference.java*.
- **Search.jsp**: Pantalla de vista del buscador. Vista controlada por la clase *Search.java*.

## 2. Código del geoportal (en php)

El código del geoportal se encuentra fusionado con los archivos del Apache MapServer y el p.Mapper, además de varios archivos que únicamente contienen el HTML para el despliegue de la aplicación web (mas no la lógica del geoportal), por lo que se mencionará los archivos relevantes desarrollados que realizan las funciones de georreferencia. La ubicación del código del geoportal en el CD es:

\Instaladores\MapServer y Apache\STIIP\

- **C:\ms4w\Apache\htdocs\STIIP\busqueda\_pages\initialize\_map.php**: En este archivo se realiza la integración de la búsqueda con la visualización de los resultados en el mapa. Aquí se lee el archivo xml creado por la aplicación de motor de búsqueda (en java), para posteriormente realizar una búsqueda en la base de datos PostgreSQL

que contiene los dominios de territorio, para comprobar si existen tales campos territoriales resultados del motor de búsqueda. Se busca en la tabla “integracion\_territorial” de la base de datos geo\_bienesc, los campos de provincia, canton y poblados. Si se encuentran resultados compatibles, se crea una tabla fantasma para crear una capa de resultados en el mapa. Es necesario aclarar que, una búsqueda del índice que produzca algún campo territorial, no necesariamente puede producir una capa de resultados en el mapa, pues la base de datos de territorios, no cuenta con todos los datos de los territorios del país sino con una muestra (recogidos fuera del alcance de esta tesis).

- **C:\ms4w\Apache\htdocs\STIIP\busqueda\_pages\central\_home.php:** En este archivo se tiene que especificar la ubicación URL de la aplicación del motor de búsqueda para empotrarlo en un iframe. Esta ubicación dependerá de la instalación del Glassfish, especialmente del puerto de este, por ejemplo si se instaló en el puerto 3232, el atributo src del iframe será:

[http://localhost:3232/STIIP\\_PILOTO/faces/Search.jsp](http://localhost:3232/STIIP_PILOTO/faces/Search.jsp)

- **C:\postgres\_connection\_geo\_bienesc.php:** Se refiere a los datos de conexión a la base de datos georreferenciada PostgreSQL.
- **C:\ms4w\Apache\htdocs\STIIP\php\_scripts\functions.php:** Contiene las funciones para realizar la conexión a la base de datos PostgreSQL y ejecutar sentencias SQL.

- **C:\ms4w\Apache\htdocs\STIIP\php\_scripts\classes.php**: Variables globales a la aplicación del geoportal.

## Anexo 2. Capturas de pantalla del proyecto SIPE

The screenshot displays the SIPE (Sistema Integrado de Información del Patrimonio Ecuatoriano) web application. The interface includes a navigation menu with categories like 'Inicio', 'Interculturalidad', 'Deportes', 'Cultura', 'Ambiente', 'Patrimonio Cultural', and 'Consultas'. The main content area is titled 'Bienes Arqueológicos' and features a table with the following data:

Lugar	# de Bienes Arqueológicos - Colecciones 2009	# de Bienes Arqueológicos - Yacimientos 2009
AZUAY	0.00	342.00
BOLIVAR	79.00	261.00
CAÑAR	29.00	383.00
CARCHI	19.00	122.00
COTOPAXI	7.00	51.00
CHIMBORAZO	63.00	222.00
EL ORO	5.00	341.00
ESMERALDAS	16.00	353.00
GUAYAS	30.00	458.00
IMBABURA	20.00	458.00
LOJA	38.00	681.00
LOS RIOS	7.00	373.00
MANABI	31.00	735.00
MORONA SANTIAGO	28.00	378.00
SUCUMBIOS	2.00	66.00
ORELLANA	5.00	110.00
SANTO DOMINGO DE LOS TSÁCHILAS	1.00	8.00
SANTA ELENA	18.00	331.00
ZONA NO DELIMITADA	s/n	s/n
<b>Total</b>	<b>0.00</b>	<b>0.00</b>

Below the table, there is a bar chart titled '# de Bienes Arqueológicos - Yacimientos (todo) 2009'. The chart shows the number of archaeological sites for each province, with the highest counts for Manabí and Loja.

© 2009 - all rights reserved - MINISTERIO DE COORDINACIÓN Y PATRIMONIO - ECUADOR

FIGURA 13-1 PANTALLA DE SIPE

## Anexo 3. Capturas de pantalla de Portal de promoción de la cultura de España

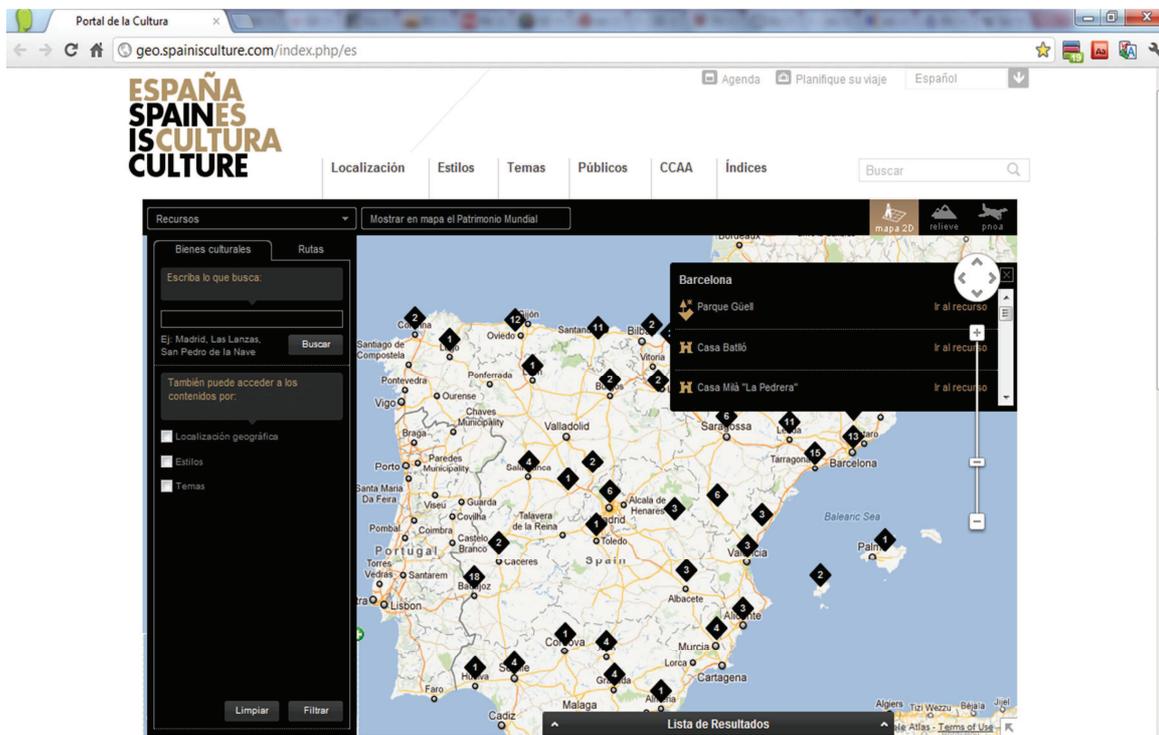


FIGURA 13-2 PANTALLA DEL PORTAL DE PROMOCIÓN DE LA CULTURA DE ESPAÑA

## Anexo 4. Captura de pantalla de los resultados de la búsqueda de la palabra "riobamba" en el Google

El panel de la izquierda nos muestra la lista de categorías o búsqueda facetada (Todo, Imágenes, Videos, Noticias) y en el panel de la derecha los vínculos a las páginas encontradas de la búsqueda.

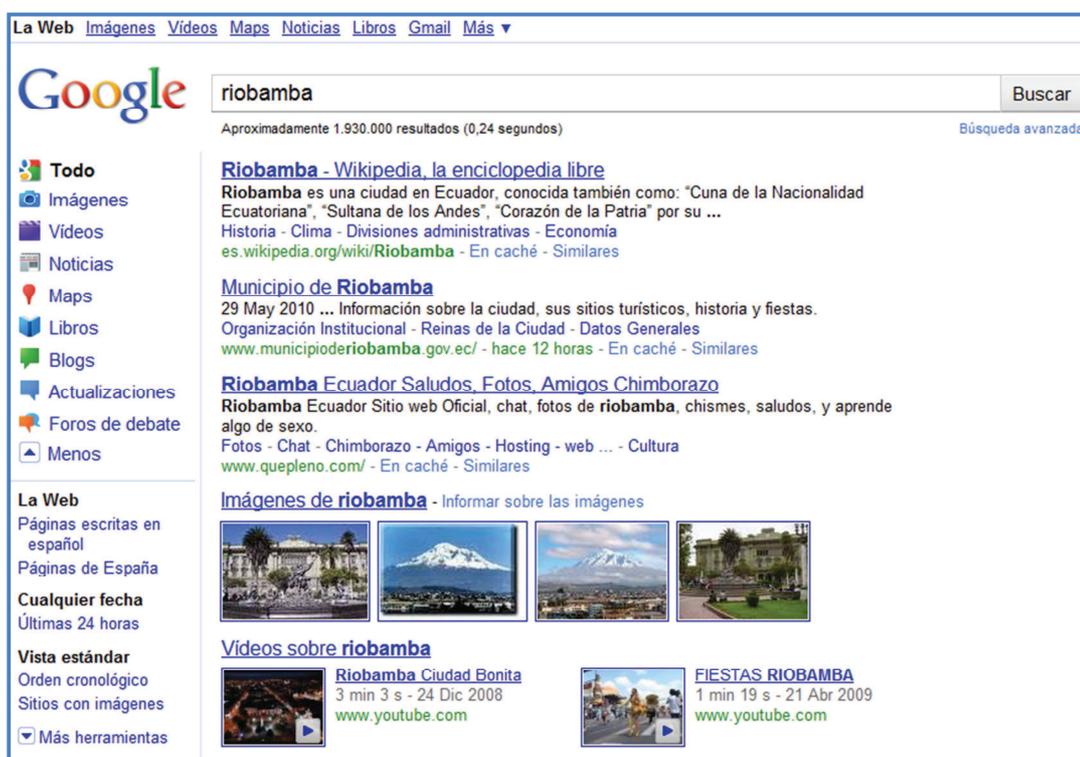


FIGURA 13-3 PANTALLA DE LA BÚSQUEDA DE "RIOBAMBA" EN GOOGLE

## Anexo 5. Captura de pantalla del geoportal STIP en la página del MCP y acta de entrega del proyecto

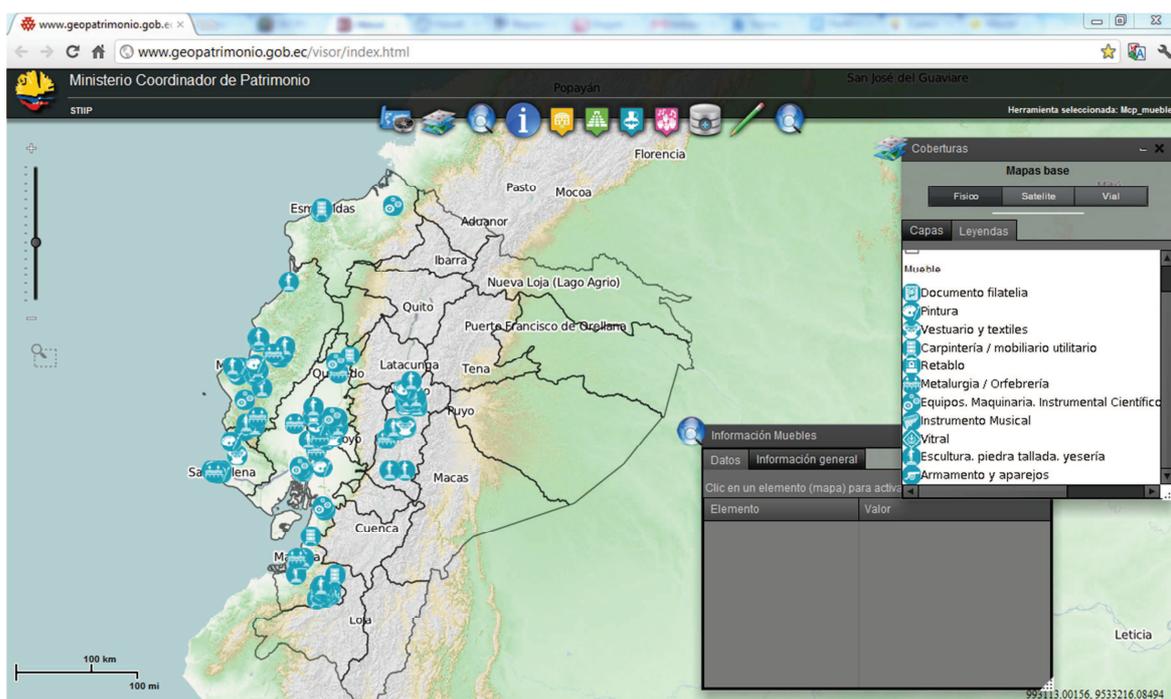


FIGURA 13-4 PANTALLA DEL GEOPORTAL STIP ACTUAL EN LA PÁGINA WEB DEL MCP

Fuente: <http://www.geopatrimonio.gob.ec/visor/index.html>

Acta de entrega del MCP**ACTA DE ENTREGA RECEPCION UNICA Y DEFINITIVA**

A los 22 días del mes de septiembre del 2010, en la ciudad de Quito, suscriben la presente acta de entrega recepción, por una parte, en representación del Ministerio Coordinador de Patrimonio el señor Pool Segarra, Subsecretario de Análisis e Información y por el Centro de Transferencias y Desarrollo de Tecnologías de la Universidad San Francisco de Quito, el señor Victor Viteri; en su calidad de Director Administrativo, a quienes para efectos de este instrumento se le podrá denominar como el "MINISTERIO" y el "CONSULTOR" respectivamente. Los comparecientes por sus propios derechos y los que representan suscriben la siguiente acta de entrega recepción al tenor de las siguientes cláusulas:

**CLAUSULA PRIMERA.- ANTECEDENTES:**

Con fecha 10 de Mayo del 2010, el MINISTERIO celebró un contrato No. CDC-MCP-003-2010 con el CONSULTOR con el fin de realizar la "Consultoría – Diseño para desarrollar un Sistema Territorial de Información Integrada de los Patrimonios como parte del Sistema Nacional de Información que permita planificar y gestionar los territorios del país – STIIP".

El 1 de junio del 2010, el CONSULTOR mediante comunicado dirigido al Director Administrativo Financiero del MINISTERIO, solicita una extensión del plazo de entrega del producto final, programado inicialmente 60 días después de la firma del contrato llevado a cabo el 10 de mayo del 2010, para 30 días calendario adicionales. Dicho pedido es aceptado por el MINISTERIO mediante resolución No. 053-CP-MCP-2010, de fecha 11 de junio del 2010.

El 15 de junio del 2010, el CONSULTOR entregó al MINISTERIO el producto 1 de la consultoría, el mismo que fue aceptado mediante un informe técnico y se procedió al pago del 20% del valor total del contrato.

En cumplimiento con el contrato antes mencionado, el CONSULTOR entregó los productos 2, 3, 4 y 5 objetos del contrato el 6 de agosto del 2010, los cuales fueron aceptados a satisfacción del MINISTERIO mediante un informe técnico final.

En vista de que se cumplió con la totalidad del objeto del contrato, dentro del término previsto, se procede a la suscripción del presente instrumento Acta de Entrega Recepción definitiva.

**PRODUCTOS ENTREGADOS:**

De conformidad con el objeto de la adquisición acordada entre las partes comparecientes y a los antecedentes expuestos, se entrega a completa satisfacción del "MINISTERIO", dentro del plazo estipulado para tal efecto de común acuerdo entre las partes, los siguientes productos:

Producto	Detalle de los Productos Objetos de la Consultoría	Entregado a Satisfacción
1	Construir una propuesta de sistema de información integrado al SNI, que visualice los requerimientos de MCP y que contenga una estrategia de Desarrollo e Implementación del STIIP, como plataforma de almacenamiento, consulta institucional y de la ciudadanía en general, integración de la información territorial y visualización de la información importante para los diversos entes coordinados por el MCP así como a la población en general	SI
2	Generar el diseño conceptual del STIIP, en base al proyecto aprobado por el MCP, así como sus interrelaciones internas y externas, en correspondencia con las necesidades del Ministerio Coordinador de Patrimonio Natural y bajo el marco de la Constitución de la República del Ecuador, y con el Plan Nacional del Buen Vivir a través del S.N.I	SI
3	Establecer la viabilidad técnica para la implementación, operatividad y ajustes del sistema con alta capacidad de adaptación a diversos entornos institucionales y herramientas para la visualización y consulta de la información intersectorial; así como el fortalecimiento de los procesos de generación de registros administrativos en el sector	SI
4	Recopilar los datos de al menos dos sectores involucrados en el proceso, los mismos que deberán servir como ejemplo en el prototipo diseñado	SI
5	Diseñar un prototipo (maqueta) que refleje la funcionalidad básica de la estructura del STIIP, sobre la base del diseño conceptual, con la finalidad que sea un insumo para la construcción del sistema en mención y adicionalmente sea una herramienta que permita visualizar el proyecto y de esta manera difundir el mismo.	SI

**CLAUSULA SEGUNDA.- LIQUIDACION PLAZOS:**

Por cuanto el trabajo objeto del contrato ha sido entregado dentro del plazo constante en la adjudicación, se deja expresa constancia de que no existe mora en el cumplimiento del objeto del mismo.

Ministerio Coordinador  
de Patrimonio



El Plazo de entrega de la "Consultoría – Diseño para desarrollar un Sistema Territorial de Información Integrada de los Patrimonios como parte del Sistema Nacional de Información que permita planificar y gestionar los territorios del país – STIIP", incluida la extensión de plazo, es de **90 días** (noventa días) calendario contados a partir de la notificación de la firma del contrato, mismo que se efectuó el día 10 de mayo del 2010 de acuerdo al reporte del área Administrativo-Financiera del Ministerio, cumpliendo así dentro del plazo estipulado.

**CLAUSULA TERCERA.- LIQUIDACION ECONOMICA.-**

Con estos antecedentes, "EL MINISTERIO" se comprometió a cancelar la suma de VEINTE Y CINCO MIL 00/100 DÓLARES DE LOS ESTADOS UNIDOS DE AMÉRICA (25.000,00). De conformidad con lo estipulado en el contrato, "EL MINISTERIO" entregó el 20% del monto del contrato, es decir, la suma de CINCO MIL 00/100 DÓLARES DE LOS ESTADOS UNIDOS DE AMÉRICA (5.000,00), en calidad de pago por la entrega del primer producto objeto del contrato. El saldo correspondiente al 80% del monto de la compra, es decir, la cantidad de VEINTE MIL 00/100 DÓLARES DE LOS ESTADOS UNIDOS DE AMÉRICA (20.000,00), lo cancelará "EL MINISTERIO" a la suscripción de la presente Acta.

**CLAUSULA CUARTA.- CONSTANCIA DE LA RECEPCION.-**

Por lo expuesto, por medio de la presente acta de Entrega Recepción Única que tiene efectos de definitiva las partes dejan expresa constancia de la entrega a entera satisfacción del Ministerio Coordinador del Patrimonio:

Ing. Pool Segarra

Subsecretario de Análisis e Información  
Ministerio Coordinador de Patrimonio

Ph.D Carlos Mena  
Coordinador de la Consultoría  
Centro de Transferencias y  
Desarrollo de Tecnologías de la  
Universidad San Francisco de Quito

## **Anexo 6. Lista de categorías para la búsqueda avanzada**

### Dublin Core

- autor
- claves
- cobertura
- derechos
- descripcion
- editor
- fecha
- formato
- fuente
- identificador
- lengua
- otros\_colaboradores
- relacion
- tipo
- titulo

### Identificadores del CSP

- institucion
- institucion\_consejo\_patrimonio

### Dominios de identidad

- ambiental\_especie\_clasificacion
- ambiental\_especie\_categoria
- ambiental\_ecosistema
- ambiental\_geologia
- ambiental\_sismologia
- ambiental\_climatologia
- ambiental\_geografia\_botanica

- ambiental\_geografia\_zoologica
- ambiental\_agronomia
- ambiental\_silvicultura
- ambiental\_infraestructura
- salud\_sanidad\_publica
- salud\_personal\_sanitario
- salud\_recursos\_medicos
- salud\_infraestructura
- salud\_intercultural
- cultural\_artes\_en\_general
- cultural\_artes\_visuales
- cultural\_musica
- cultural\_folklor
- cultural\_literatura
- cultural\_arte\_culinario
- cultural\_radiodifusion
- cultural\_historia
- cultural\_documentacion\_en\_general
- cultural\_produccion\_documental
- cultural\_centros\_de\_documentacion
- cultural\_educacion
- cultural\_fiestas\_populares
- cultural\_fiestas\_nacionales
- cultural\_fiestas\_religiosas
- cultural\_religiosidad
- cultural\_infraestructura
- social\_viviendas
- social\_etnias
- social\_idiomas
- social\_geografia\_social
- social\_infraestructura\_educativa
- social\_educacion\_intercultural
- social\_personal\_educativo

- patrimonio\_arqueologico
- patrimonio\_documental
- patrimonio\_inmaterial
- patrimonio\_inmueble
- patrimonio\_mueble
- patrimonio\_natural
- economico\_industrias
- economico\_mineria
- deportivo\_infraestructura
- deportivo\_actividades\_deportistas
- deportivo\_deportivos
- deportivo\_excursiones\_montañismo\_camping
- deportivo\_oferta\_publica\_de\_actividades\_deportistas
- gubernamental\_proyectos\_y\_programas
- gubernamental\_instituciones\_culturales
- gubernamental\_instituciones\_ambientales
- gubernamental\_infraestructura

### Territoriales

- territorial\_ciudad
- territorial\_parroquia
- territorial\_canton
- territorial\_provincia
- territorial\_region
- territorial\_vias\_de\_comunicacion\_terrestres
- territorial\_vias\_de\_comunicacion\_maritimas
- territorial\_vias\_de\_comunicacion\_aereas
- territorial\_ingenieria\_hidraulica
- territorial\_ingenieria\_civil
- territorial\_cursos\_de\_agua\_naturales
- territorial\_geografia\_fisica
- territorial\_zonas\_de\_riesgo

## Anexo 7. Manual de operación: Inicio rápido

En primera instancia, se debe arrancar todos los servicios de software mencionados anteriormente. Este proceso de inicio rápido se lo realiza en el servidor central, en este caso, el servidor del MCP. Esto quiere decir que, para la instalación en el lado del servidor del prototipo se debe ejecutar los siguientes comandos<sup>1</sup> (no necesariamente en el orden establecido):

1. Instalar e iniciar el servidor Apache MapServer v.2.3.1 (el cual incluye las librerías de PHP) en el puerto 80 (es necesario que ningún otro proceso se encuentre corriendo en el mismo puerto). Incluir el sitio web de geoportal/mapas (aquel desarrollado en PHP) en la carpeta *htdocs* del servidor Apache.
2. Instalar e iniciar el servidor Glassfish v3 Prelude en el puerto 8080 (es necesario que ningún otro proceso se encuentre corriendo en el mismo puerto). Añadir el sitio web “STIIP\_PILOTO\_MOTOR\_BUSQUEDA” al servidor Glassfish, que es aquel que soporta JavaServer Faces (JSF), en lo cual está desarrollado el proceso de búsqueda y la indexación.
3. Iniciar Apache Solr v.1.4.0. Esto se lo puede realizar con un archivo ejecutable Jar. Ir al directorio *solr/example/* y ejecutar el comando:

```
java -jar start.jar
```

Dos archivos de la carpeta de Solr son importantes. Lo primero, es cargar el archivo “*schema.xml*”, en la carpeta *solr/example/solr/conf*; este archivo es el que contiene las especificaciones de los metadatos en el índice o las secciones previas: los meta-

---

<sup>1</sup> En el CD adjunto al presente documento, se encuentran todos los instaladores y archivos necesarios para ejecutar el STIIP-Piloto en un servidor.

metadatos. Más adelante se detallará el contenido de este archivo. Lo segundo, es ubicar los archivos del índice en la carpeta *solr/example/solr/data/index*. En el CD de instaladores, consta un ejemplo de índice que es el resultado de la indexación de varios documentos (+50 archivos y una base de datos) dotados por el CSP, para las pruebas de funcionamiento del STIIP-Piloto; lo cual produjo más de 50 mil documentos del índice en Lucene y más de 1.2 millones de campos-valores de metadatos en los cuales se realizan las búsquedas (ver Figura 13-5). En el caso de querer iniciar con un índice vacío, por ejemplo, en la fase de puesta en producción (fuera del alcance del presente proyecto de tesis), no se deberá cargar ningún archivo en la carpeta del índice “*index*”.

4. Instalar e iniciar el PostgreSQL. Cargar la base de datos que se provee en el CD de instaladores: “*geo\_bienesc.backup*” y “*stiip\_db.backup*”. Incluir en la raíz del sistema o en el disco C: los archivos “*postgres\_connection\_geo\_bienesc.php*” y “*postgres\_connection.php*”
5. Iniciar conexiones a bases de datos (remotas o locales) a indexarse, si es aplicable.

Este proceso de inicio, se lo debe realizar una sola vez en un servidor que siempre está en funcionamiento. También se proporciona el archivo *stiip-start.bat*, para iniciar automáticamente los servicios de los diferentes servidores (ejecutar el archivo con privilegios de administrador): Apache, Glassfish, Solr si previamente se ha realizado la ubicación de los diferentes archivos necesarios en las carpetas mencionadas: el sitio web de *geoportal/mapas*, el sitio web del motor de búsqueda *STIIP\_PILOTO\_MOTOR\_BUSQUEDA*, el archivo de configuración de metadatos “*schema.xml*”, los archivos del índice y cargadas las bases de datos georreferenciados en PostgreSQL.

Solr Admin (example)

Erick-PC:8983  
 cwd=C:\Users\Erick\Documents\CTT-USFQ (Solr)\STIIP\_PILOTO CD  
 final\Instaladores\Solr\apache-solr-1.4.0\apache-solr-1.4.0\example SolrHome=solr/

Apache Solr

Schema Browser | See [RAW SCHEMA.XML](#)

HOME  
 FIELDS  
 DYNAMIC FIELDS  
 FIELD TYPES

### Schema Information

Unique Key: [URL](#)

Default Search Field: [TEXT](#)

numDocs: 56616

maxDoc: 60963

numTerms: 1265096

version: 1274746578784

optimized: false

current: true

hasDeletions: true

directory:  
 org.apache.lucene.store.SimpleFSDirectory:org.apache.lucene.store.SimpleFSDirectory@  
 C:\Users\Erick\Documents\CTT-USFQ (Solr)\STIIP\_PILOTO CD  
 final\Instaladores\Solr\apache-solr-1.4.0\apache-solr-1.4.0\example\solr\data\index

lastModified: 2012-01-23T20:02:51.831Z

FIGURA 13-5 ESTADÍSTICAS DEL ÍNDICE DE MUESTRA USADO EN EL STIIP-PILOTO

## Anexo 8. Manual de usuario

### 1. Mapa del sitio

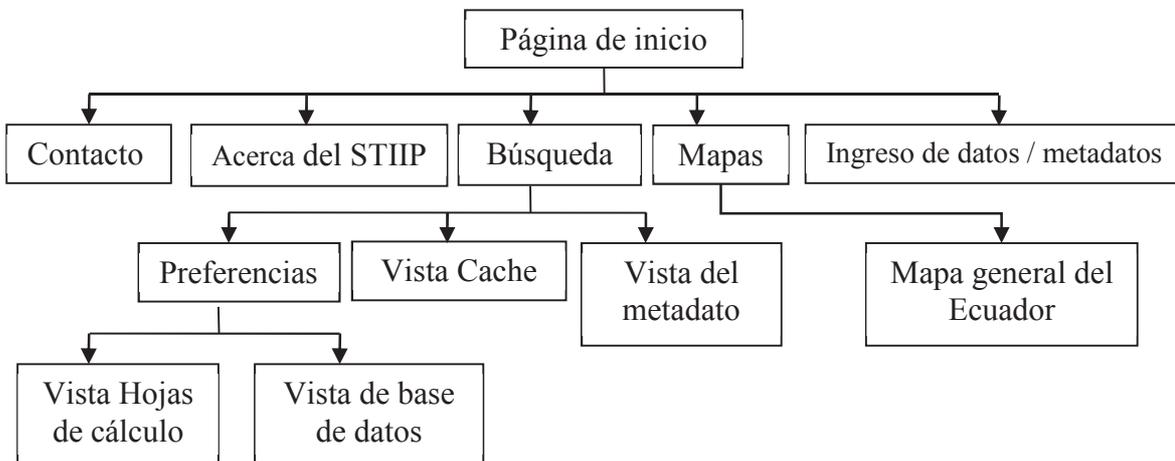


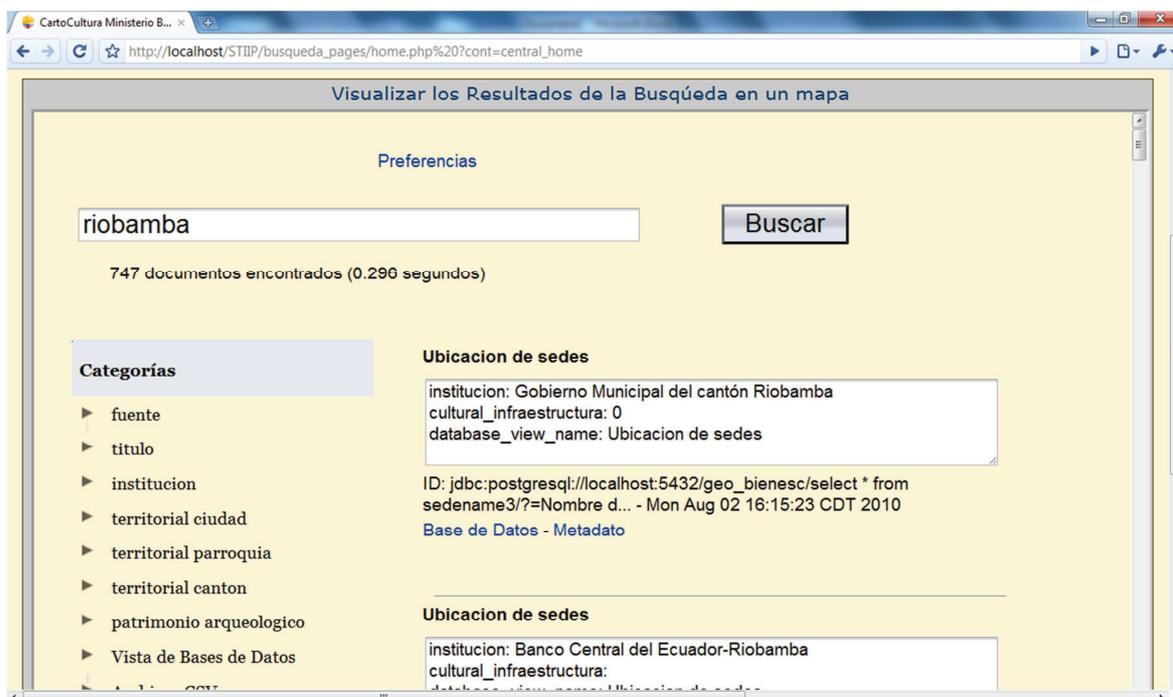
FIGURA 13-6 MAPA DEL SITIO



FIGURA 13-7 PÁGINA INICIO

## 2. Tipos de búsquedas

El proceso de búsqueda por palabra clave es muy sencillo desde el punto de vista de utilización. Este proceso permite al usuario del STIIP ingresar una palabra clave en un campo de texto dentro de una página de un navegador de Internet y, obtener como resultado una lista de documentos relacionados a la palabra o al grupo de palabras ingresadas. Como se ha indicado anteriormente, este procedimiento es en todo similar a lo que es la utilización de cualquier motor de búsqueda en internet como Google, Yahoo!, etc.



**FIGURA 13-8 PANTALLA DE LA BÚSQUEDA POR PALABRA CLAVE "RIOBAMBA"**

El proceso de búsqueda se realizará en todos los campos del metadato especificado en la fase de diseño del metadato.

## 2.1 Búsquedas generales

Para realizar una búsqueda se ingresará la(s) palabra(s) clave(s) que se desea buscar dentro de la caja de texto. Los resultados encontrados dentro del índice tendrán el siguiente formato a la vista del usuario:

Archivos generales:

- Nombre / Título del documento (con link para abrir el archivo localmente).
- Descripción o resumen del documento (si está disponible o si fue posible su extracción mediante Tika).
- Link al texto del documento en Cache (visualizado en el navegador) - Link al metadato completo del documento (a ser visualizado en el navegador).



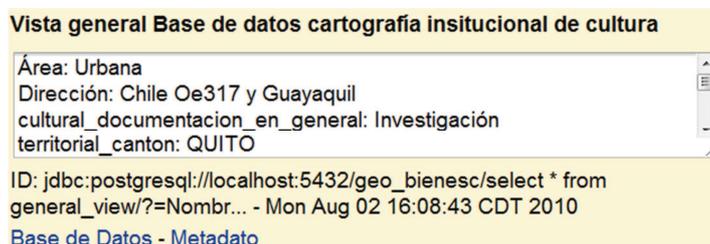
[Patrimonio 2007.doc](#)  
Descripción no disponible  
[Cached - Metadato](#)

**FIGURA 13-9 EJEMPLO DE FORMATO DE PRESENTACIÓN DE ARCHIVOS GENERALES EN LOS RESULTADOS DE UNA BÚSQUEDA**

Cada documento de la red del CSP, será guardado como un documento en el índice de Lucene con sus respectivos metadatos, a excepción de las bases de datos y archivos Excel o de formato csv.

Archivos de base de datos:

- Nombre de la vista guardada (predefinida en las máscaras de extracción de información), asociada a una base de datos.
- Descripción de la vista (si está disponible).
- Lista de campo-valor del registro (fila) de la tabla de la base de datos asociada al resultado de la búsqueda.
- Identificador del documento (ubicación local o remota de la base de datos + query de la máscara) - Fecha de indexación.
- Link a la vista completa (máscara) de la base de datos - Link al metadato completo del documento (a ser visualizado en el navegador).

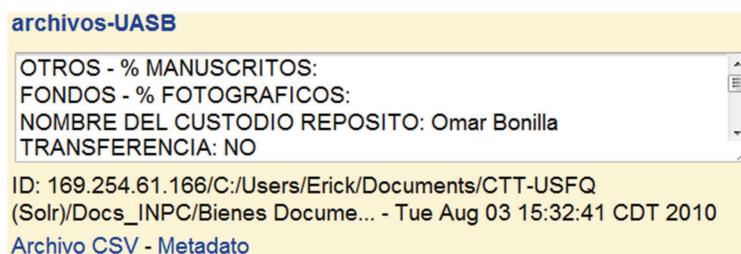


**FIGURA 13-10 EJEMPLO DE FORMATO DE PRESENTACIÓN DE BASES DE DATOS COMO ARCHIVOS DEL ÍNDICE EN EL RESULTADO DE UNA BÚSQUEDA**

Cada registro (fila) de una tabla, de una base de datos cualquiera, es almacenado como un archivo único en el índice Lucene, con los metadatos respectivos en las columnas de la tabla (o vista) de la base de datos.

Archivos de hojas de cálculo (en formatos xls o csv):

- Nombre de la hoja de cálculo indexada (con link para abrir el archivo localmente).
- Descripción de la hoja de cálculo (si está disponible).
- Lista de campo-valor del registro (fila) de la hoja de cálculo asociada al resultado de la búsqueda.
- Identificador del documento (dirección IP del documento en la red del CSP + ruta local del archivo (*filepath*) + nombre del archivo) - Fecha de indexación.
- Link a la hoja de cálculo en Cache – Link y al metadato completo del documento (a ser visualizado en el navegador).



**FIGURA 13-11 EJEMPLO DE FORMATO DE PRESENTACIÓN DE HOJAS DE CÁLCULO COMO ARCHIVOS DEL ÍNDICE EN EL RESULTADO DE UNA BÚSQUEDA**

Así como las bases de datos, cada registro (fila) de una hoja de cálculo cualquiera es almacenado como un archivo único en el índice Lucene, con los metadatos respectivos en las columnas de la tabla de la hoja de cálculo.

## 2.2 Búsquedas específicas

Utilizar la misma sintaxis de las búsquedas generales, usando comillas (“...”) para encontrar frases textuales. Por ejemplo, si se busca “*museo municipal*”, producirá una lista diferente de resultados si se busca solo *museo* o *municipal*.

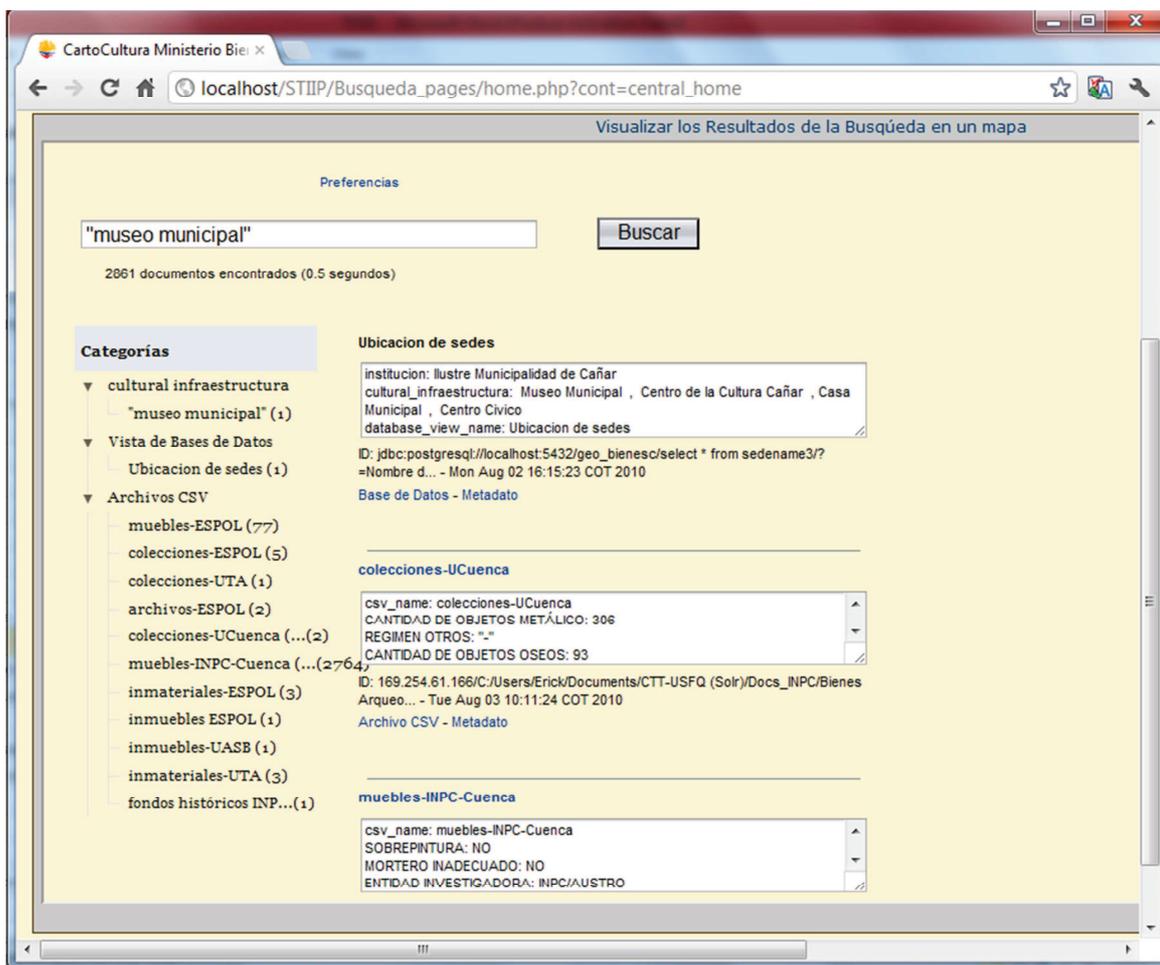


FIGURA 13-12 PANTALLA DE LA BÚSQUEDA ESPECÍFICA DE "MUSEOS MUNICIPALES"

### 2.3 Búsqueda avanzada

Para realizar directamente una búsqueda avanzada, es decir, por alguna categoría, se debe utilizar la siguiente sintaxis dentro de la caja de texto de búsqueda:

*categoría: clave*

Por ejemplo:

territorial\_ciudad: Manta

autor: “Julio Jaramillo”

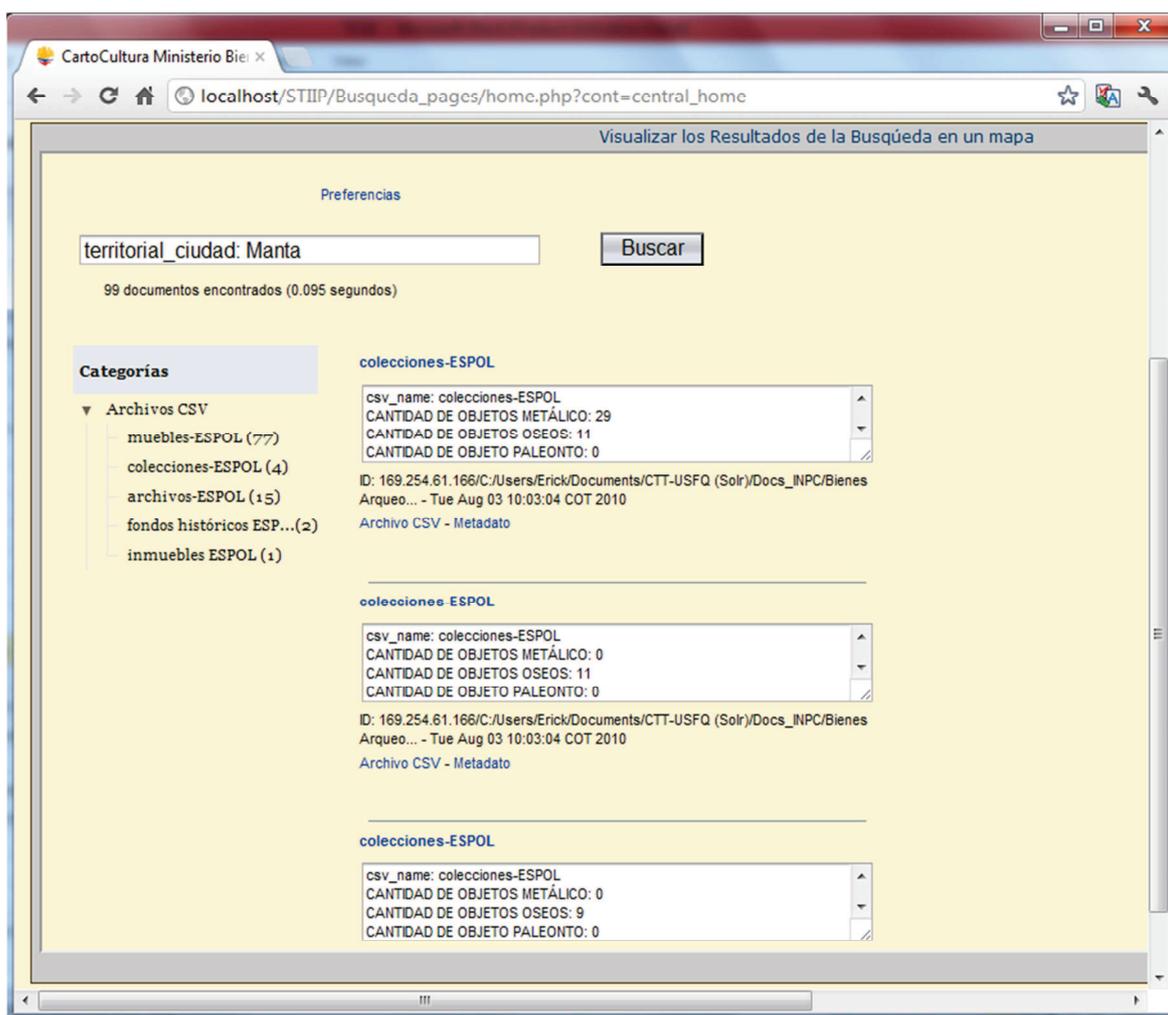


FIGURA 13-13 PANTALLA DE BÚSQUEDA AVANZADA O POR CATEGORÍA

Las categorías implementadas son las especificadas en el diseño; es decir, las del Dublin Core, los dominios de identidad, los identificadores del CSP y las territoriales. En el Anexo 6, se encuentra la lista completa de las categorías que se pueden usar en la búsqueda avanzada, que en total son 87 categorías.

#### *2.4 Búsquedas facetada o desde el árbol de categorías*

La función de búsqueda facetada o desde el árbol de categorías permite organizar automáticamente los resultados de una búsqueda, según un criterio de integración territorial. Cada vez que un usuario realiza una búsqueda cualquiera (general, específica o avanzada), los resultados serán presentados en la columna de búsqueda facetada, divididos en base a las coincidencias del texto buscado, con los campos de integración textual de los metadatos presentes en el índice. Cada documento encontrado será, por lo tanto, presentado en la columna de resultados según su puntaje de búsqueda y también en la columna de búsqueda facetada a lado de la columna de resultados. En la columna de búsqueda facetada, el mismo resultado podrá presentarse más de una vez, siendo posible que sean más que una las coincidencia de los campos de integración territorial del metadato con las palabras ingresadas en fase de búsqueda. Por ejemplo la búsqueda de *bosque alfaró* se puede ver en la figura 13-14.

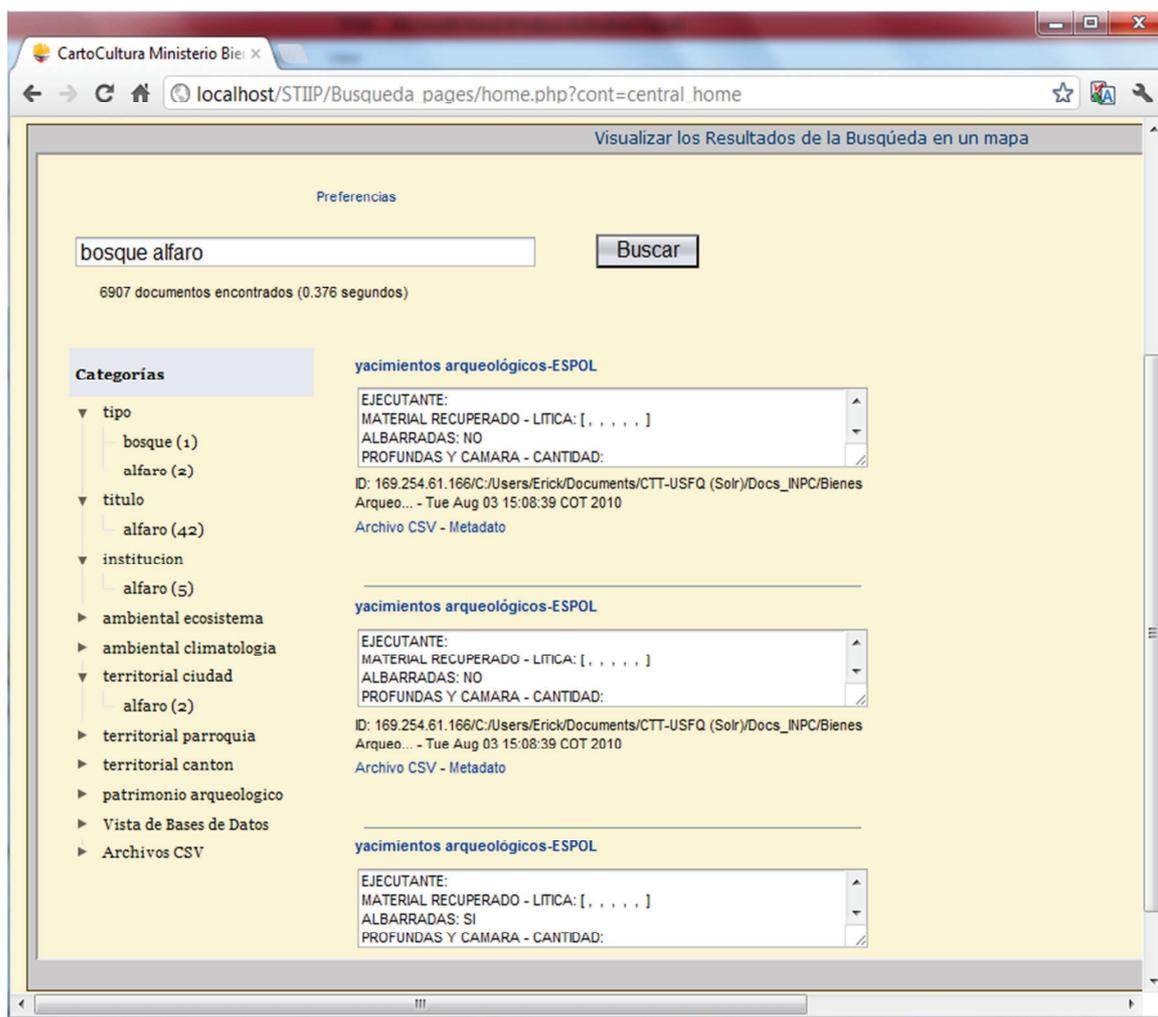


FIGURA 13-14 PANTALLA DE BÚSQUEDA DE *BOSQUE ALFARO* Y EL ÁRBOL DE CATEGORÍAS

Como se puede ver en la figura 13-14, la palabra clave *bosque* aparece en la categoría tipo, pero la palabra *alfarero* aparece en las categorías tipo, título, institución y territorial\_ciudad. Además, se ve entre paréntesis el número de documentos encontrados por palabra clave en las diferentes categorías, por ejemplo, en el índice existen 42 documentos con metadatos de campo-valor: titulo-alfarero, o un documento con el campo-valor: tipo-bosque.

El árbol de categorías muestra las categorías encontradas en los documentos que contengan la búsqueda realizada. Dentro de una categoría se muestra el número de documentos

encontrados con cada palabra clave del total de documentos encontrados. Cabe recalcar que, en el momento de la indexación de bases de datos y hojas de cálculo, cada fila representa un documento independiente, que en sus metadatos incluye el valor asignado a cada columna de la tabla, además de otros metadatos para el control y funcionamiento del resto del motor de búsqueda. Estos registros de los documentos de hojas de cálculo y bases de datos, representarán una unidad en el conteo de número total de documentos en cada categoría del árbol facetado. Si se encontraron documentos dentro de una base de datos o de alguna hoja de cálculo, se muestra el nombre del documento dentro de las categorías “Vista de Bases de Datos” y “Archivos CSV”, respectivamente.



**FIGURA 13-15 PANTALLA DE LA BÚSQUEDA DE AZUAY QUECHUA Y EL ÁRBOL DE CATEGORÍAS (VISTA DE BASES DE DATOS Y ARCHIVOS CSV)**

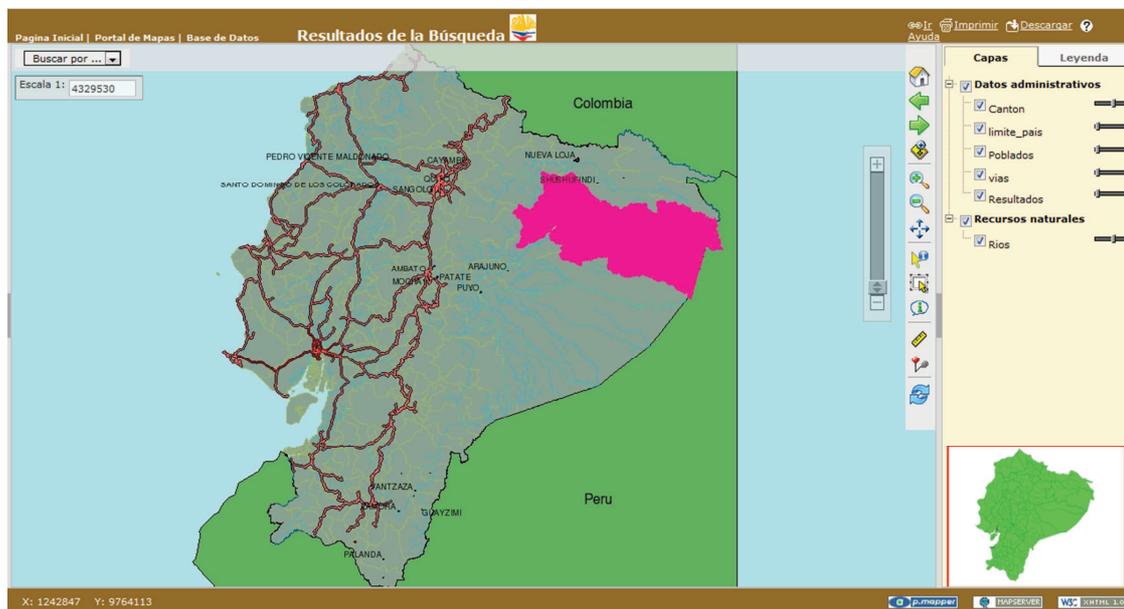
### 3. Geoportal y el visualizador de resultado en el visor de mapas

El proceso de recuperación de la información integrada, a través del mapa principal del geoportal del STIIP, luego de la definición de las capas de la cartografía base que serán visualizadas en el mapa, se basa sencillamente en la utilización de las herramientas típicas de cualquier visualizador de mapas. Se podrán utilizar instrumentos como el arrastre del mapa con el mouse, el zoom, activación o desactivación de capas en el mapa, entre otros.

Para ver el mapa de los resultados, hacer clic en el link “Visualizar los Resultados de la Búsqueda en un mapa” en la parte superior de la pantalla. Para poder ver algún resultado en el mapa, es necesario que la búsqueda realizada produzca una lista con algún documento georreferenciado, es decir, que tenga alguno de los siguientes campos lleno:

- territorial\_ciudad (activo en el prototipo)
- territorial\_parroquia
- territorial\_canton (activo en el prototipo)
- territorial\_provincia (activo en el prototipo)
- territorial\_region
- territorial\_vias\_de\_comunicacion\_terrestres
- territorial\_vias\_de\_comunicacion\_maritimas
- territorial\_vias\_de\_comunicacion\_aereas
- territorial\_ingenieria\_hidraulica
- territorial\_ingenieria\_civil
- territorial\_cursos\_de\_agua\_naturales
- territorial\_geografia\_fisica

- territorial\_zonas\_de\_riesgo



**FIGURA 13-16 PANTALLA DEL VISUALIZADOR DE RESULTADOS DE LA PALABRA CLAVE  
*ORELLANA***

#### **4. Preferencias o sitio de administración**

El sitio de administración permite a los técnicos o tomadores de decisiones, realizar las funciones destinadas para ellos. Entre las funciones desarrolladas en el STIIP-Piloto, se encuentra la posibilidad de indexar los archivos de un directorio completo (carpeta raíz más carpetas internas), a manera del algoritmo de la araña establecido en el diseño, indexar bases de datos creando las máscaras dentro de la misma aplicación e indexar hojas de cálculo. A continuación se detalla el funcionamiento de estas herramientas del sitio de administración.

#### 4.1 Directorios



**FIGURA 13-17 PANTALLA DE LA HERRAMIENTA PARA INDEXAR DIRECTORIOS COMPLETOS**

Para añadir los archivos de un directorio (y sus carpetas internas) al índice, simplemente se ingresa la dirección del directorio local y automáticamente se indexarán todos los archivos en este directorio y sus subdirectorios. La extracción de metadatos de archivos se realiza por medio de Apache Tika. Este extractor de metadatos soporta los siguientes formatos que serán los que podrán ser analizados e indexados por él:

- HTML
- XML
- MS Office (doc, xls, ppt)
- ODF
- PDF
- EPUB
- RTF
- TXT

- ZIP
- TAR
- MP3
- WAV
- MIDI
- JPEG
- FLV
- MBOX

Los campos extraídos por TIKa, son los campos del Dublin Core:

- Título
- Claves
- Descripción
- Fuente
- Lengua
- Relación
- Cobertura
- Autor
- Último autor
- Editor
- Otros colaboradores
- Derechos
- Fecha

- Tipo
- Formato
- Identificador
- Tamaño del archivo

Tika también tiene la posibilidad de extraer todo el contenido textual del documento y, este será archivado en el campo del metadato *texto* a ser presentado en el link “cached” de cada archivo cuando se realiza uno de los tipos de búsquedas.

Finalmente, se debe notar que Tika extrae esta información de los datos que podemos ver en las propiedades del documento (al hacer clic derecho en Propiedades y ver sus detalles), o los metadatos del documento. Por esta razón, es importante que las propiedades se las complete al momento de ser creado cada archivo a ser indexado para el CSP, de manera que Tika pueda realizar la extracción automática de esta información.

#### *4.2 Bases de Datos*

Esta herramienta de la página de preferencias, permite indexar bases de datos locales o remotas. Para añadir una base de datos, primero se debe ingresar la siguiente información:

- Tipo de driver (actualmente existe soporte solo para MySQL y PostgreSQL)
- Host
- Puerto

- Nombre de la base de datos
- Usuario
- Contraseña

Lo antes mencionado solo guarda los datos de conexión a una base de datos, mas aún no se realiza la conexión. Para añadir los datos de una nueva base de datos, no es necesario estar conectado a la base de datos todavía.

The screenshot shows a web interface titled "Bases de Datos" with a light yellow background. It is divided into two main sections: "Añadir una nueva base de datos o usuario:" on the left and "Editar Vistas:" on the right. The left section contains several input fields: "Tipo" (a dropdown menu set to "MySQL"), "Host:", "Puerto:", "Base de Datos:", "Username:", and "Password:". Below these fields are two buttons: "Añadir" and "Validar Usuario". The right section, "Editar Vistas:", features a text area containing the connection string "jdbc:postgresql://localhost:5432/geo\_bienesca" and a scroll bar. Below this text area are two more input fields: "Usuario:" and "Password:", followed by a "Validar Usuario" button. At the bottom center of the interface is an "Eliminar" button.

**FIGURA 13-18 PANTALLA DE LA HERRAMIENTA PARA AÑADIR LOS DATOS DE CONEXIÓN A UNA BASE DE DATOS**

Después de añadir los datos de conexión de una base de datos, se muestra a la derecha en “Editar Vistas”, una lista de todas las bases de datos añadidas. Para realizar efectivamente la indexación de una de ellas se debe seleccionar en la lista una e ingresar el usuario y la contraseña. Se realizó de esta forma ya que el ingreso de los datos de la base de datos se debe ejecutar solo una vez, pero la indexación dependerá en el tiempo con las nuevas tablas de la base de datos, es decir, con la creación de máscaras o vistas de la base de datos que serán múltiples. Una vez seleccionada una base de datos e ingresado el usuario y

contraseña, debe aplastar el botón “Validar usuario” que comprueba que el usuario y contraseña sean correctos. Después hacer clic en el botón que aparecerá “Editar” para la creación de vistas o máscaras.

The screenshot shows a web-based interface for creating database views. It features a 'Regresar' button at the top left. The main form has the following fields:

- Ingresar comando SQL:** A text area containing the SQL query: `select * from general_view`. A 'query' button is next to it.
- Nombre de la Vista:** A text box containing 'Vista general geo\_bienesec'.
- Descripción de la vista:** A text box containing 'Detalles generales de bienes'.
- Tipo / Icono:** A dropdown menu currently set to 'Cultural'. An 'Añadir Vista' button is below it.

On the left side, there is a 'Vistas Indexadas' section with a scrollable list of SQL queries. At the bottom, there is a table with three columns: 'Nombre de la institución', 'Maxima Autoridad', and 'Depende de Otra Institución'. The table contains 15 rows of data.

Nombre de la institución	Maxima Autoridad	Depende de Otra Institución
Dirección Provincial de Cultura de Tung...	Tania MarisoNavarrete Cárdenas	Si
Casa de la Cultura Ecuatoriana - núcl...	María JudithPachano Holguin	No
CONSEJO PROVINCIAL DE TUNGURAHUA	FernandoNaranjo Lalama	No
Dirección Departamento de Cultura del M...	FernandoCallejas Barona	Si
Casa de Montalvo	Mario GerardoMoro Nieto	No
Dirección Provincial de Cultura de Chim...	Nancy LeonorHaro Pontón	Si
Casa de la Cultura Ecuatoriana núcleo C...	Gabriel FranciscoCisneros Abedravo	No
Gobierno Provincial de Chimborazo		No
Banco Central del Ecuador-Riobamba	Silvia JimenaBarahona Castillo	Si
Museo de Arte Religioso de la Concepció...	MauricioDilon Dominguez	No
Gobierno Municipal del cantón Riobamba	Angel IgnacioYáñez Cabrera	No
Instituto Nacional de Patrimonio Cultura...	Iván EduardoCastro Vaca	Si
Dirección Provincial de Cultura de Bolí...	Manuel PatricioBallesteros Trujillo	Si
Casa de la Cultura - núcleo Bolívar	María AliciaOsorio Castillo	No

**FIGURA 13-19 PANTALLA DE CREACIÓN DE VISTAS O MÁSCARAS EN UNA BASE DE DATOS**

En la figura 13-19, se muestra la herramienta de creación de máscaras. Se puede crear varias máscaras o vistas o eliminarlas también. En esta pantalla debe ingresar algún comando SQL (que será la vista o máscara de una o varias tablas de la base de datos), el nombre de la vista, la descripción y el tipo de patrimonio (cultural, ambiental, instrucción, deportivo, salud, infraestructura). Este último, el tipo de patrimonio, servirá para mostrar un ícono diferente en el mapa de resultados de una búsqueda. Después de ingresar el comando SQL y aplastar el botón “Query”, se cargará en la pantalla los resultados del query realizado. Luego, para cada columna debe ingresarse el tipo de dato (los campos definidos en el metadato base del STIIP) que corresponda a los valores de la columna. Si se selecciona la opción de “ignorar”, la columna será indexada de todos modos con un nombre por defecto (el cual incluye el nombre definido en la columna de la vista

realizada). El formato de campo en el metadato de las columnas ignoradas será el siguiente:

NOMBRE DE LA COLUMNA\_T

Por ejemplo, según la figura 13-19, la segunda columna será almacenada en el índice con el nombre del campo del metadato:

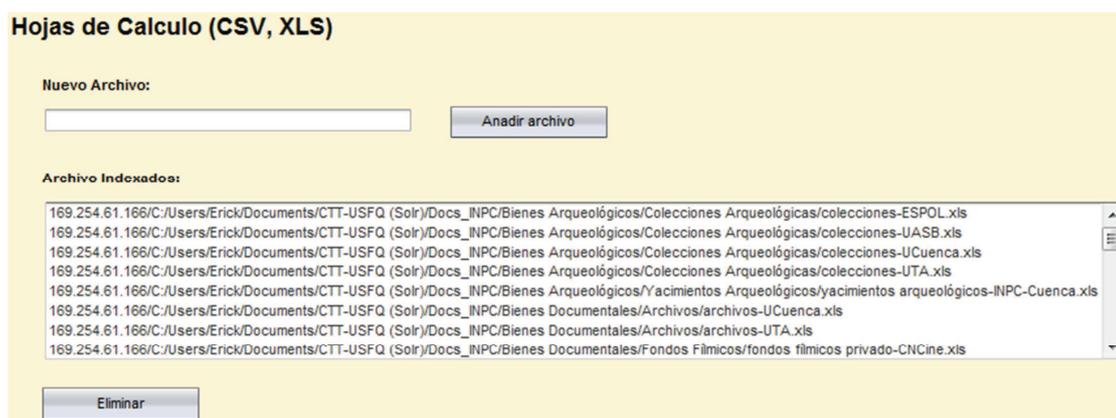
Maxima Autoridad\_T

Las columnas ignoradas podrán ser buscables pero no se incluirán en el árbol de categorías; es decir, que todos los campos de metadatos que incluyen una \_T al final, serán aquellas que no fueron definidas en el diseño del metadato base del STIIP, pero se incluyen en el índice a manera de campos de texto. Después de seleccionar el tipo de dato de cada columna, finalmente debe darse clic en el botón “Añadir Vista” y, quedarán indexadas en el índice de Lucene. Se puede crear varias vistas o máscaras para una base de datos en particular y las vistas añadidas se muestran en la parte izquierda de la pantalla, bajo el título “Vistas Indexadas”, en donde existe la posibilidad de eliminarlas también.

#### *4.3 Hojas de cálculo*

Esta herramienta es parecida a la herramienta de ingreso de bases de datos. Ingresar la dirección completa de la hoja de cálculo (en un directorio local). Los formatos soportados por el sistema son CSV y XLS. Para la lectura correcta de las hojas de cálculo en formato

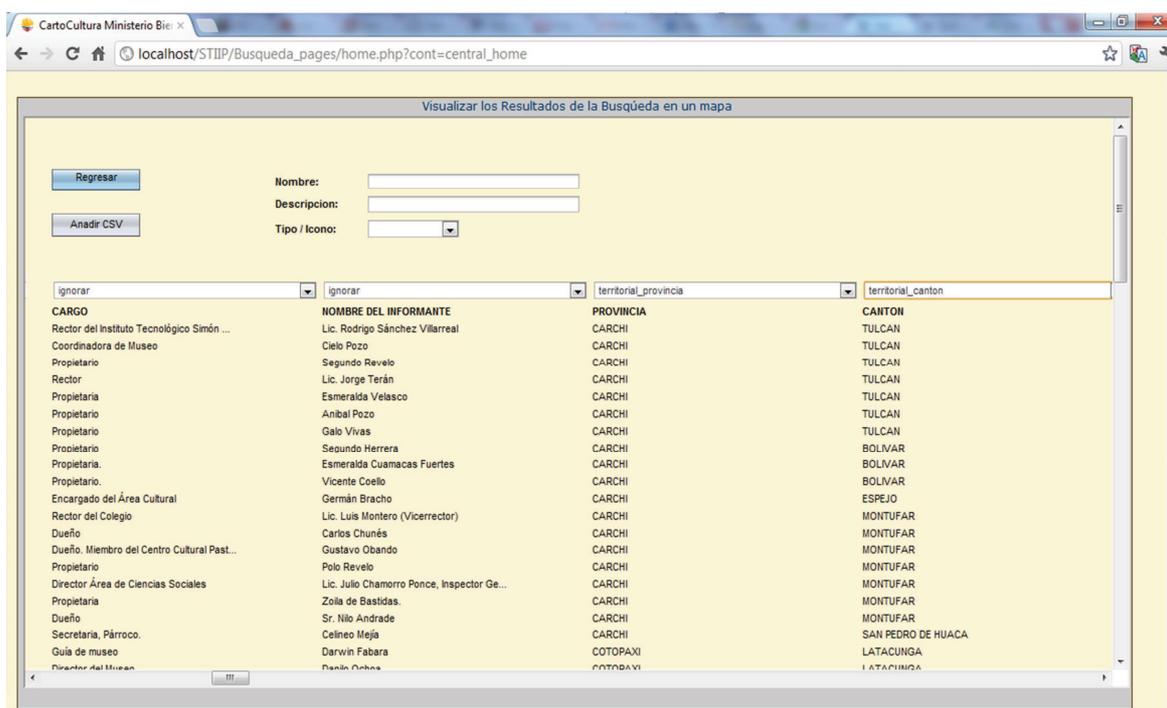
xls es muy importante que, la primera fila de la hoja de cálculo sea la línea en donde se encuentran los títulos y que todas las columnas utilizadas tengan algún título (no vacío), caso contrario se producirá un error en el extractor de datos para formatos xls y será necesario guardarlo como archivo con formato csv para su posterior indexación en el STIIP.



**FIGURA 13-20 PANTALLA DE LA HERRAMIENTA PARA AÑADIR UN ARCHIVO DE FORMATO XLS O CSV**

Igualmente que la herramienta de base de datos, después de añadir la ruta local del documento (ubicación y nombre del documento), se registrará automáticamente la IP de la computadora en la cual se añadió el archivo (para diferenciar archivos en servidores periféricos) y se añadirá la hoja de cálculo a la lista de Archivos indexados para la creación de máscaras. En la figura 13-21, se muestra la pantalla de configuración de la indexación de una hoja de cálculo. Es similar a la pantalla de creación de vistas de bases de datos, ya que se debe añadir un nombre de la máscara de la hoja de cálculo, descripción, tipo de ícono y modificar las columnas de acuerdo al tipo de campo del metadato. En la figura también se muestra que existe una columna PROVINCIA y CANTON, por lo que en este

caso podemos asociar a estas columnas el campo *territorial\_provincia* y *territorial\_canton*, respectivamente, de los campos del metadato base. Este proceso de designar a cada columna de la hoja de cálculo o de una base de datos con un tipo de campo del metadato STIIP, se lo realizó porque en el CSP existen varios archivos con formatos diferentes de creación. Por ejemplo, si un archivo contiene una columna llamada “canton-posicion”, “lugar de cantón” o “cantones”, estos datos de esta columna no serán agrupados bajo el campo *territorial\_canton*, estándar del metadato base del STIIP y, no podrán ser utilizados para las búsquedas facetadas o por el árbol de navegación, lo cual producirá a la larga un índice con muchos campos de metadatos creados a posteriori (aquellos que terminan en \_T) y una baja precisión en la búsqueda de documentos.



**FIGURA 13-21 PANTALLA DE CONFIGURACIÓN DE INDEXACIÓN DE UNA HOJA DE CÁLCULO**