



# Universidad de Granada

E.T.S. Ingenierías Informática y Telecomunicaciones

Departamento de Ciencias de la Computación e  
Inteligencia Artificial

SISTEMA MULTIAGENTE PARA LA GESTIÓN  
INTELIGENTE DE COLECCIONES COMPLEJAS

TESIS DOCTORAL

Ramón Pérez Pérez

Granada, 8 de Junio de 2006





# Universidad de Granada

SISTEMA MULTIAGENTE PARA LA GESTIÓN  
INTELIGENTE DE COLECCIONES COMPLEJAS

Ramón Pérez Pérez

TESIS DOCTORAL

Directores: Dr. D. Miguel Delgado Calvo-Flores

Dr. D. Waldo Fajardo Contreras

Junio, 2006



**DECSAI**

**Departamento de Ciencias de la Computación e I.A.**

Universidad de Granada



La memoria **Sistema MultiAgente para la Gestión Inteligente de Colecciones Complejas**, que presenta Ramón Pérez Pérez, para optar al grado de Doctor en Informática, ha sido realizada en el Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada, bajo la dirección del Profesor Dr. D. Miguel Delgado Calvo-Flores, Catedrático de Universidad y del Profesor Dr. D. Waldo Fajardo Contreras, Profesor titular de Universidad.

Granada, Junio de 2006.

Fmdo: Miguel Delgado Calvo-Flores

Fmdo: Waldo Fajardo Contreras

Fmdo: Ramón Pérez Pérez



*A mi familia y a Raquel*

**SISTEMA MULTIAGENTE PARA LA GESTIÓN  
INTELIGENTE DE COLECCIONES COMPLEJAS**

Ramón Pérez Pérez





# Agradecimientos

Un sincero agradecimiento a mis directores, Miguel y Waldo, por su tiempo prestado, interés, apoyo y ánimo incondicional en la realización de este trabajo. Gracias a ambos por compartir mi entusiasmo e ilusión en estos años.

Además, he de nombrar también al Herbario de la Universidad de Granada, y en especial a Carmen, por desarrollar mi formación botánica e instruirme en que la naturaleza no está formada solamente de ceros y unos.

No me puedo olvidar de agradecer a todos los integrantes del grupo de investigación ARAI (*Approximated Reasoning and Artificial Intelligence*), del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada; en particular, a mis compañeros de Biblioteconomía, por la paciencia mostrada y los buenos momentos que he pasado con ellos.

Quiero hacer una mención especial para agradecer a todos los miembros del grupo de investigación de EVANNAI (*Evolutionary Algorithms & Neural Networks & Artificial Intelligence*) de la Universidad Carlos III de Madrid, por su cálida acogida y por facilitarme el desarrollo del presente trabajo.

Dicen que las pautas para escribir los agradecimientos es ser lo más breve posible, pero no se puede. Gracias a la Unidad de coordinación de GBIF España por su apoyo.

Todo esto nunca hubiera sido posible sin el apoyo incondicional de mis padres, Ramón y Filomena, mis hermanos, Antonio y Javier, y demás personas queridas.

Gracias a Raquel, mi inseparable compañera que tanto me ha ayudado y aguantado durante este largo y arduo camino, que en un día como hoy, 8 de Junio, comenzó nuestro viaje.



# Índice

<b>INTRODUCCIÓN GENERAL .....</b>	<b>1</b>
<b>CAPÍTULO I. APLICACIONES PARA LA BIOINFORMÁTICA Y ESTÁNDARES.....</b>	<b>7</b>
1. ÁREAS DE APLICACIÓN DE LA INFORMÁTICA PARA LA BIODIVERSIDAD .....	8
1.1. Identificación mediante caracteres .....	8
1.2. Gestión de los Nombres taxonómicos y clasificación.....	10
1.3. Gestión de los Datos sobre los especímenes .....	17
2. EL TRATAMIENTO DE LAS COLECCIONES. ....	19
2.1. Informatización y gestión de colecciones.....	24
2.1.1. Determinación de los especímenes.....	24
2.1.2. Informatización del Material.....	25
2.1.3. Conservación de los especímenes y su uso .....	26
2.1.4. Préstamos de Especímenes.....	26
2.1.5. Revisiones de los Especímenes .....	27
2.1.6. Resumen del tratamiento del espécimen .....	28
3. ESTÁNDARES INTERNACIONALES .....	29
3.1. Iniciativas para la estandarización .....	30
3.1.1. GBIF: Global Biodiversity Information Facility .....	30
3.1.2. BioCASE: Biological Collection Access Service .....	35
3.2. Esquemas de intercambio de especímenes u observaciones .....	38
3.2.1. Darwin Core (TDWG, 2005) .....	38
3.2.2. ABCD Schema (Taxonomic Databases Working Group, 2006) .....	40
3.3. Protocolos de Comunicación .....	42
3.3.1. DiGIR .....	43
3.3.2. Proyecto BioCASE .....	54
3.3.3. TAPIR: Protocolo de Acceso para la Recuperación de Información.....	56
<b>CAPÍTULO II. SISTEMAS SOFTWARE EXISTENTES .....</b>	<b>63</b>
1. EL GESTOR DE BASES DE DATOS DE BIODIVERSIDAD: BIOTA 2 .....	66
1.1. Características .....	66
1.2. Análisis de Características.....	68
2. SPECIFY .....	69
2.1. Características .....	70
2.2. Análisis de Características.....	71
3. BIOTICA .....	72

3.1.	<i>Características</i> .....	72
3.2.	<i>Análisis de Características</i> .....	74
4.	BIOLINK .....	74
4.1.	<i>Características</i> .....	75
4.2.	<i>Análisis de Características</i> .....	75
5.	HERBAR .....	76
5.1.	<i>Características</i> .....	77
5.2.	<i>Análisis de Características</i> .....	78
6.	SISTEMA DE GESTIÓN DE HERBARIOS E INVESTIGACIÓN BOTÁNICA: BRAHMS .....	79
6.1.	<i>Características</i> .....	79
6.2.	<i>Análisis de Características</i> .....	80
7.	SISTEMA PARA LA GESTIÓN DE HERBARIO MEDIANTE ACCESS (HERBARIUM ACCESS MANAGEMENT SYSTEM) .....	81
7.1.	<i>Características</i> .....	81
7.2.	<i>Análisis de Características</i> .....	83
8.	UN SISTEMA PARA LA GESTIÓN DE HERBARIOS: TRACY .....	83
8.1.	<i>Características</i> .....	84
8.2.	<i>Análisis de Características</i> .....	84
9.	PLABEL: HERBARIUM LABEL PROGRAM .....	85
9.1.	<i>Características</i> .....	85
9.2.	<i>Análisis de Características</i> .....	86
10.	HERBARIO VIRTUAL EXPRESS: VHE (VIRTUAL HERBARIUM EXPRESS).....	87
10.1.	<i>Características</i> .....	87
10.2.	<i>Análisis de Características</i> .....	88
11.	RESUMEN COMPARATIVO .....	89
11.1.	<i>Análisis de la situación actual</i> .....	90
11.2.	<i>El uso de los datos</i> .....	92
<b>CAPÍTULO III. BIOMEN: UN SISTEMA DE INFORMACIÓN PARA EL TRATAMIENTO DE COLECCIONES COMPLEJAS .....</b>		<b>95</b>
1.	DESARROLLO E IMPLEMENTACIÓN .....	96
1.1.	<i>Modelo de ciclo de vida</i> .....	96
1.2.	<i>Requerimientos del sistema</i> .....	98
1.2.1.	Requerimientos no funcionales.....	99
1.3.	<i>Arquitecturas y Software de desarrollo</i> .....	100
1.3.1.	<i>Software necesario para la solución del problema</i> .....	101
1.3.2.	Arquitectura de tres capas.....	106
2.	CARACTERÍSTICAS .....	110
2.1.	<i>Control de usuarios</i> .....	111
2.2.	<i>Gestión de especímenes</i> .....	115

2.3.	<i>Generación de Etiquetas</i> .....	117
2.4.	<i>Consultas</i> .....	121
2.5.	<i>Préstamos de material</i> .....	125
2.6.	<i>Información multimedia</i> .....	127
2.7.	<i>Referencias bibliográficas</i> .....	128
2.8.	<i>Información taxonómica</i> .....	130
2.9.	<i>Servicios Externos</i> .....	131
2.10.	<i>Exportación de información</i> .....	142
2.11.	<i>Distribución de Especímenes</i> .....	144
3.	COMPARACIÓN CON LOS SISTEMAS ACTUALES.....	148
<b>CAPÍTULO IV. BIOMEN: UN SISTEMA MULTIAGENTE .....</b>		<b>151</b>
1.	INTRODUCCIÓN .....	151
2.	SISTEMA MULTIAGENTE COOPERATIVO DISTRIBUIDO A TRAVÉS DE INTERNET .....	155
3.	CARACTERÍSTICAS .....	157
3.1.	<i>Personalización de la interfaz del usuario</i> .....	160
3.1.1.	Filtrado Cooperativo (CF).....	166
3.1.2.	Minería de Datos Web .....	167
3.1.3.	Aplicación de la Minería de Datos a la Web.....	169
3.1.4.	Aplicación de RSS como sistema de publicación de noticias.....	173
3.2.	<i>Georreferenciación de colecciones</i> .....	175
3.3.	<i>Comunicación con proyectos nomenclaturales</i> .....	181
3.3.1.	<i>ITIS</i> .....	182
3.3.2.	<i>Species 2000</i> .....	184
3.4.	<i>Estudio de la secuenciación nominal</i> .....	187
<b>CAPÍTULO V. RESUMEN Y CONCLUSIONES.....</b>		<b>199</b>
<b>TRABAJOS FUTUROS .....</b>		<b>203</b>
<b>GLOSARIO 205</b>		
<b>APÉNDICE A. ESTÁNDAR PARA LA CONSTRUCCIÓN DE SISTEMAS</b>		
<b>MULTIAGENTES: LA FIPA.....</b>		<b>213</b>
1.	ARQUITECTURA PROPUESTA .....	213
2.	FIPA SL CONTENT LANGUAGE SPECIFICATION .....	218
3.	FIPA AGENT MANAGEMENT SPECIFICATION.....	220
4.	FIPA REQUEST INTERACTION PROTOCOL SPECIFICATION .....	221
5.	FIPA QUERY INTERACTION PROTOCOL SPECIFICATION.....	223
6.	FIPA CONTRACT NET INTERACTION PROTOCOL SPECIFICATION .....	224
7.	FIPA ITERATED CONTRACT NET INTERACTION PROTOCOL SPECIFICATION .....	225
8.	FIPA BROKERING INTERACTION PROTOCOL SPECIFICATION .....	225

---

9.	FIPA RECUITING INTERACTION PROTOCOL SPECIFICATION.....	226
10.	FIPA COMMUNICATIVE ACT LIBRARY (CAL) SPECIFICATION.....	226
11.	FIPA AGENT MESSAGE TRANSPORT SERVICE SPECIFICATION.....	228
12.	FIPA AGENT MESSAGE TRANSPORT PROTOCOL FOR IIOP SPECIFICATION.....	230
13.	FIPA AGENT MESSAGE TRANSPORT PROTOCOL FOR HTTP SPECIFICATION.....	230
<b>APÉNDICE B. INTRODUCCIÓN AL LENGUAJE XML.....</b>		<b>233</b>
1.	¿QUÉ ES XML?.....	233
2.	HISTORIA DE XML.....	233
3.	XML BIEN FORMADO.....	233
4.	XML NAMESPACES.....	236
5.	XML Y DICCIONARIOS DE DATOS.....	237
6.	LECTURA Y MANIPULACIÓN DE UN DOCUMENTO XML.....	238

# Tabla de Ilustraciones

FIGURA 1.	PROCESO DE DETERMINACIÓN DEL <i>STATUS</i> DE UN NOMBRE CIENTÍFICO.....	15
TABLA 1.	ÁREAS DE INVESTIGACIÓN EN BIODIVERSIDAD COMPUTACIONAL (STOCKWELL, 1997).....	23
FIGURA 2.	PROCESO DE INFORMATIZACIÓN Y FLUJO DE INFORMACIÓN.....	29
FIGURA 3.	INTEGRACIÓN DE FUENTES HETEROGÉNEAS.....	34
FIGURA 4.	FLUJO DE DATOS EN BIOCASE. ....	37
TABLA 2.	ATRIBUTOS DE DARWIN CORE (UNIDAD DE COORDINACIÓN DE GBIF EN ESPAÑA, 2006B). (*: DENOTA EL CARÁCTER DE OBLIGATORIO)...	40
TABLA 3.	ESQUEMA DE INTERCAMBIO DE INFORMACIÓN ABCD SCHEMA ....	41
FIGURA 5.	DIAGRAMA DE ACTUACIÓN DE DIGIR. (DIGIR, 2006).....	44
FIGURA 6.	PROVEEDOR DE DATOS DIGIR (DIGIR, 2006) .....	45
FIGURA 7.	DIAGRAMA DEL REGISTRO DE PROVEEDORES (DIGIR, 2006).....	46
FIGURA 8.	DIAGRAMA DEL PORTAL DE DIGIR (DIGIR, 2006) .....	46
FIGURA 9.	COMPONENTES DE UNA CABECERA DE UN MENSAJE EN EL PROTOCOLO DIGIR.....	48
FIGURA 10.	SENTENCIA REQUEST DE DIGIR.....	48
FIGURA 11.	OPERACIÓN DE BÚSQUEDA DE INFORMACIÓN. PROTOCOLO DIGIR	49
FIGURA 12.	ESTRUCTURA PETICIÓN DE INVENTARIADO. PROTOCOLO DIGIR....	50
FIGURA 13.	ESTRUCTURA DE LOS METADATOS. PROTOCOLO DIGIR.....	51
FIGURA 14.	RESPUESTA A LA PETICIÓN DE UN MENSAJE. PROTOCOLO DIGIR...52	
FIGURA 15.	CÓDIGO XML QUE IMPLEMENTA LA PETICIÓN DE METADATOS DIGIR 52	
FIGURA 16.	CÓDIGO XML DE RESPUESTA A PETICIÓN DE METADATOS EN DIGIR 53	
FIGURA 17.	ARQUITECTURA DEL PROVEEDOR DE DATOS .....	54
FIGURA 18.	PETICIÓN DE INFORMACIÓN A UN PROVEEDOR BIOCASE. ....	55

FIGURA 19.	OPERACIÓN <i>CAPABILITIES</i> EN EL PROTOCOLO BIOCASE .....	56
FIGURA 20.	DIAGRAMA DEL PROTOCOLO DE COMUNICACIÓN TAPIR.....	57
FIGURA 21.	TAPIR: ESTRUCTURA DEL MENSAJE DE PETICIÓN.....	58
FIGURA 22.	TAPIR: ESTRUCTURA DEL MENSAJE <i>CAPABILITIES</i> .....	59
FIGURA 23.	TAPIR: OPERACIÓN DE PETICIÓN DE INVENTARIADO.....	60
FIGURA 24.	TAPIR: ESTRUCTURA DE UNA VISTA .....	61
FIGURA 25.	TAPIR: PETICIÓN DE BÚSQUEDA (“ <i>SEARCH</i> ”).....	61
FIGURA 26.	TAPIR: ESTRUCTURA DEL MENSAJE DE RESPUESTA.....	62
FIGURA 27.	CAPTURA DE PANTALLA DE BIOTA .....	68
FIGURA 28.	FORMULARIO DE ENTRADA DE DATOS SOBRE ESPECÍMENES EN SPECIFY .....	71
FIGURA 29.	MENÚ PRINCIPAL DE BIÓTICA.....	72
FIGURA 30.	PANTALLA PRINCIPAL DE HERBAR .....	78
FIGURA 31.	CAPTURA PANTALLA PRINCIPAL BRAHMS .....	80
FIGURA 32.	ENTRADA DE DATOS DE ESPECÍMENES EN HERBARIUM ACCESS....	82
FIGURA 33.	PLANTILLA DE DATOS SOBRE ESPECÍMENES .....	84
FIGURA 34.	ENTRADA DE DATOS DE ESPECÍMENES DE PLABEL .....	85
FIGURA 35.	PANTALLA PRINCIPAL DE VHE .....	88
TABLA 4.	TABLA COMPARATIVA DE LOS SISTEMAS <i>SOFTWARE</i> EXISTENTES (VER LEYENDA EN SIGUIENTE TABLA) .....	89
TABLA 5.	LEYENDAS DE TABLA COMPARATIVA.....	90
FIGURA 36.	MODELO DE CICLO DE VIDA EVOLUTIVO .....	97
FIGURA 37.	ESTRUCTURA CLIENTE/SERVIDOR DE BIOMEN.....	100
FIGURA 38.	COMPARATIVA DE USO DE LOS DISTINTOS SERVIDORES WEB (WWS, 2005) 103	
FIGURA 39.	ARQUITECTURA DE TRES CAPAS.....	107
FIGURA 40.	GENERACIÓN DINÁMICA MEDIANTE XML Y XSLT.....	108
FIGURA 41.	SALIDA DE LA INTERFAZ EN HOJA DE CÁLCULO EXCEL.....	109
FIGURA 42.	FRAGMENTO DE LA ONTOLOGÍA QUE DEFINE EL DOMINIO DE LA INTERFAZ DE USUARIO.....	110



FIGURA 43.	LISTA DE PERMISOS EN BIOMEN .....	113
FIGURA 44.	TIPOS DE INFORMACIÓN EN EL MÓDULO DE ESPÉCIMENES. ....	115
FIGURA 45.	FILTRADO DE DATOS EN BIOMEN .....	116
FIGURA 46.	EXTRACTO DE INFORMACIÓN PARA UN ESPÉCIMEN .....	116
FIGURA 47.	DATOS PARA UN NUEVO CENTRO.....	117
TABLA 6.	CÓDIGOS DE BARRAS DE UNA DIMENSIÓN MÁS REPRESENTATIVOS. 118	
TABLA 7.	CÓDIGOS DE BARRAS MULTIDIMENSIONALES .....	119
FIGURA 48.	ETIQUETA CON CÓDIGO DE BARRAS EN BIOMEN .....	120
FIGURA 49.	EQUIVALENCIA ENTRE DARWIN CORE Y CÓDIGO BIDIMENSIONAL PDF417 EN BIOMEN .....	121
FIGURA 50.	TRES SUBMÓDULOS DE CONSULTA EN BIOMEN.....	121
FIGURA 51.	FRAGMENTO DE CONSULTA POR MUESTRA .....	122
FIGURA 52.	CONSULTA SIMPLE EN BIOMEN.....	123
FIGURA 53.	PASO 1: ESTABLECIENDO CONDICIONES DE CONSULTA .....	123
FIGURA 54.	PASO 2: SELECCIÓN DE COLUMNAS DE RESULTADO.....	123
FIGURA 55.	RESULTADO DE CONSULTA AVANZADA .....	124
FIGURA 56.	ESTRUCTURA DE BÚSQUEDA Y REFINAMIENTO DE LA CONSULTA 125	
FIGURA 57.	PROCESO DE ALTA DE UN PRÉSTAMO. ....	126
FIGURA 58.	ASOCIACIÓN DE INFORMACIÓN MULTIMEDIA A UNA MUESTRA ..	128
FIGURA 59.	ACCESO A LA INFORMACIÓN MULTIMEDIA DESDE EL MÓDULO DE CONSULTA .....	128
FIGURA 60.	ASIGNACIÓN DE BIBLIOGRAFÍA A UNA MUESTRA.....	129
FIGURA 61.	CAPTURA DEL MÓDULO DE BIBLIOGRAFÍA DE BIOMEN .....	130
FIGURA 62.	PLANTILLA PARA UN NUEVO GÉNERO EN BIOMEN .....	130
FIGURA 63.	ESQUEMA DE PETICIÓN HTTP A TRAVÉS DEL NAVEGADOR WEB..	133
FIGURA 64.	ESQUEMA PROPIO PARA LAS CONSULTAS EXTERNAS.....	133
FIGURA 65.	RESPUESTA A CONSULTA EXTERNA PARA LISTAR ESPECIES .....	135
FIGURA 66.	DIAGRAMA DE FLUJO DE DATOS EN UN SERVICIO WEB .....	137

FIGURA 67.	MENSAJE WSDL PARA EL SERVICIO WEB CONSULTA DE ESPECÍMENES. ....	140
FIGURA 68.	COMUNICACIÓN DESDE JAVA CON LOS SERVICIOS WEB DE BIOMEN 141	
FIGURA 69.	RESULTADO DE LAS LLAMADAS A LOS SERVICIOS WEB DE BIOMEN 141	
FIGURA 70.	LISTADO DE VISTAS DE EXPORTACIONES EN BIOMEN .....	142
FIGURA 71.	DATOS DINÁMICOS PARA LAS VISTAS DE EXPORTACIONES EN BIOMEN .....	143
FIGURA 72.	EXTRACTO DE LA DEFINICIÓN DE LOS CAMPOS A EXPORTAR EN BIOMEN .....	144
FIGURA 73.	EXTRACTO DE LA DEFINICIÓN DE CONDICIONES PARA LA EXPORTACIÓN EN BIOMEN .....	144
FIGURA 74.	RESPUESTA DE SOLICITUD DE VISUALIZACIÓN DE MAPA EN MAPQUEST .....	146
FIGURA 75.	FICHERO KML GENERADO POR BIOMEN PARA GOOGLE EARTH ....	147
FIGURA 76.	VISUALIZACIÓN CON GOOGLE EARTH DE DATOS GEOESPACIALES DE BIOMEN .....	148
TABLA 8.	TABLA COMPARATIVA DE LOS SISTEMAS <i>SOFTWARE</i> EXISTENTES (VER LEYENDA EN SIGUIENTE TABLA) CON BIOMEN .....	149
TABLA 9.	LEYENDAS DE TABLA COMPARATIVA.....	149
FIGURA 77.	VISIÓN ABSTRACTA DE LA ARQUITECTURA DEL SISTEMA MULTIAGENTE .....	156
FIGURA 78.	ONTOLOGÍA DE IDIOMA PARA BIOMEN.....	161
FIGURA 79.	PORTE QUE PERSONALIZA LA PÁGINA WEB .....	162
FIGURA 80.	PARTES PERSONALIZABLES DE BIOMEN.....	162
FIGURA 81.	CONSULTA DE LOS AGENTES QUE TRADUCEN TEXTO (TRANSLATE- LANGUAGE).....	164
FIGURA 82.	ALGORITMO PARA LA NEGOCIACIÓN DE LA TRADUCCIÓN DE UN TEXTO.....	164
FIGURA 83.	ÁREAS DE APLICACIÓN DE MINERÍA WEB (ARAYA, SILVA M., 2004) 168	
FIGURA 84.	SISTEMA DE PUBLICACIÓN DE NOTICIAS PERSONALIZADAS .....	175

FIGURA 85.	EJEMPLO DE COMUNICACIÓN CON BIOGEOMANCER.....	178
FIGURA 86.	RESPUESTA DE BIOGEOMANCER .....	178
FIGURA 87.	DATOS ESPACIALES OBTENIDOS POR EL AGENTE LOCALIDAD .....	179
FIGURA 88.	VISUALIZACIÓN DE LA GEORREFERENCIACIÓN PARA UNA MUESTRA.....	180
FIGURA 89.	PÁGINA WEB Y CÓDIGO FUENTE DE CONSULTAR A ITIS .....	183
FIGURA 90.	INSTANCIA DE CONSULTA A SPECIES 2000 Y <i>STATUS</i> DEL NOMBRE 184	
FIGURA 91.	INSTANCIA DE SPECIES 2000 DONDE SE INDICA EL NOMBRE ACEPTADO PARA UNA CONSULTA .....	185
FIGURA 92.	RESPUESTA DE BIOMEN INTEGRANDO FUENTES EXTERNAS.....	186
FIGURA 93.	ONTOLOGÍA GENERADA CON PROTÉGÉ PARA LA COMUNICACIÓN ENTRE LOS AGENTES NOMENCLATURALES.....	187
FIGURA 94.	EJEMPLO DE SINONIMIA Y SECUENCIACIÓN NOMINAL.....	191
FIGURA 95.	RESULTADO DE LOS AGENTES TRAS ANALIZAR LAS MUESTRAS DEL EJEMPLO.....	194
FIGURA 96.	EJEMPLO REAL PARA ANALIZAR SU SECUENCIA NOMINAL .....	194
FIGURA 97.	REVISIÓN POR PARTE DEL EXPERTO DE LA SECUENCIACIÓN NOMINAL OBTENIDA POR LOS AGENTES. ....	196
FIGURA 98.	ARQUITECTURA PROPUESTA POR LA FIPA.....	214
TABLA 10.	CONCEPTOS DE LA FIPA.....	218
FIGURA 99.	COMUNICACIÓN ENTRE PLATAFORMAS.....	221
FIGURA 100.	PROTOCOLO DE PETICIÓN DE ACCIONES.....	222
FIGURA 101.	PROTOCOLO DE CANCELACIÓN. ....	222
FIGURA 102.	PROTOCOLO DE CONSULTA.....	223
FIGURA 103.	PROTOCOLO DE NEGOCIACIÓN (FOUNDATION FOR INTELLIGENT PHYSICAL AGENTS, 2001F) .....	224
FIGURA 104.	MENSAJE ENTRE AGENTES. ....	229
FIGURA 105.	EJEMPLO BÁSICO DE XML.....	234
TABLA 11.	ENTIDADES XML.....	235

FIGURA 106.	EJEMPLO DE DOCUMENTO XML COMPLETO.....	236
FIGURA 107.	EJEMPLO DE DOCUMENTO XML CON NAMESPACES.....	236
FIGURA 108.	EJEMPLO DE DTD.....	237
FIGURA 109.	EJEMPLO DE ESQUEMA XML.....	238

## Introducción General

El uso de los computadores para el apoyo y la automatización de las tareas desarrolladas por los biólogos no es muy reciente. Concretamente, en lo que a la Taxonomía se refiere, podemos encontrar referencias ya en los años 70, como las de Morse, Pankhurst o Dallwitz. (Dallwitz, 1974; Morse, 1970; Pankhurst, 1970; Pankhurst, 1991).

Según Stockwell (Stockwell, 1997) “La Biodiversidad es la ciencia que trata del estudio de tendencias históricas y actuales en la riqueza biológica de los ambientes y, aún estando enlazada con la Biología evolutiva, la Taxonomía y la Ecología se centra fundamentalmente en la recolección y análisis de información útil para el manejo científico de recursos naturales y la conservación”.

Las colecciones biológicas son la principal fuente de información para los estudios de biodiversidad (E. Mata Montero, 2006; Krishtalka and Humphrey, 2000; Suarez and N.D. Tsutsui, 2004). Se calcula que existen hoy en día más de tres billones (James L. Edwards, Meredith A. Lane, 2000; Robert Guralnick and David Neufeld, 2005) de especímenes almacenados en Museos de Historia Natural y Herbarios.

En la actualidad, la protección del medioambiente, así como la evaluación de las amenazas al mismo, son temas pujantes. Sin temor a equivocarnos, podemos decir que el objetivo principal es, de forma más o menos directa, preservar la biodiversidad. Es por esto que los estudios relativos a la biodiversidad se han constituido en un tema de gran relevancia y, en la actualidad, la informática se ha mostrado como herramienta imprescindible para la realización de los mismos, debido al potencial de gestión, procesamiento y distribución del conocimiento que lleva asociada (M. Delgado, W. Fajardo, 2006). Gran cantidad de resultados de particular relevancia, aparecidos como fruto de la aplicación de sofisticadas herramientas informáticas, en combinación con las técnicas de estudio

clásicas, no habrían sido posibles hace unos años cuando no era patente esta interdisciplinaridad.

La interacción entre ambos campos ha sido de tal importancia que ha dado lugar a una nueva disciplina: la bioinformática (Govindaraj, 2001). Esta se enuncia en el sentido amplio de la denominación, y no en el restrictivo que habitualmente la restringe al campo de la biología molecular (Lesk, 2002; Nicholas M Luscombe, Dov Greenbaum, 2001).

Dado el elevado potencial descubierto a partir de la aplicación de técnicas computacionales avanzadas en los estudios de biodiversidad, se ha sobrepasado la etapa en la que el estudio de los problemas podía hacerse con un *software* sin más objetivos que la consulta de presencia-ausencia o algún estudio estadístico elemental. En la actualidad, debido al que el medioambiente y la preservación de la biodiversidad son problemas globales que como tales deben de ser abordados, el *software* necesario es obligatoriamente complejo, por la cantidad de datos a procesar para obtener respuesta y soluciones a problemas de enorme magnitud espacio-temporal.

Es por esto, que es necesario introducir en los sistemas herramientas que permitan, gestionar de forma adecuada un gran volumen de datos, incorporar mecanismos de intercambio y cooperación con proyectos internacionales, integrar fuentes heterogéneas, la extracción de información, el acceso a los datos, etc. Se justifica así la necesidad de las Ciencias de la Computación y la aplicación de técnicas de inteligencia artificial, que permitan la gestión de las colecciones, y la extracción y movilización del conocimiento asociada a las mismas.

Podemos establecer las siguientes tres líneas de trabajo:

1. La informatización de las colecciones: Las colecciones biológicas son los datos primarios que testifican el pasado y el presente de la biodiversidad.

2. Proyectos nomenclaturales y taxonómicos: proyectos encargados de recopilar los nombres científicos aceptados para cada especie. La recopilación de los nombres científicos y las descripciones taxonómicas son herramientas básicas para los taxónomos.
3. La realización de estándares de intercambio: la necesidad de los investigadores de estudiar las especies descritas implica el compartir información de qué existe en los centros y de qué información se dispone.

Hasta el momento, todas las propuestas y/o soluciones se centran en alguna de las líneas sin llegar a integrarlas (Berendsohn, 2006; Biótica., 2005; Colwell, 2004; Commonwealth and Organisation, 2005; DiGIR, 2006; Döring, 2005; Oliver, Pik, 2000).

El objetivo de este trabajo es desarrollar un Sistema Multiagente que permita gestionar de forma inteligente las colecciones complejas, integrando la gestión de las colecciones, el tratamiento de los nombres, el intercambio de información y la integración de fuentes heterogéneas.

Para esto comenzaremos introduciendo las diferentes áreas de aplicación de la informática a la biología, como colección compleja, centrándonos principalmente en: la gestión de las colecciones (Bedathur and Haritsa, 2000; Berendsohn, 2006), los proyectos nomenclaturales (Index, 2004; Integrated Taxonomic Information System, 2006; Species 2000, 2005) y los estándares de intercambio de información (Berendsohn, Döring, 2002; DiGIR, 2006; Döring, 2005; Döring, Giovanni, 2004).

Una vez establecida la base de nuestro trabajo, realizamos un estudio del estado del arte de los sistemas existentes para la gestión de colecciones (Biótica., 2005; Bridson and Forman, 1992; Colwell, 2004; Commonwealth and Organisation, 2005; John D. and Catherine T. MacArthur Foundation, 1996; Minnigerode, 1997; Pando, 1991; University of California Davis, 2004; University of Florida Herbarium, 2000; University of Oxford, 1985). Para ello, se

han analizado las características de los diez sistemas *software* más relevantes. Tras éste análisis, se ha comprobado la inexistencia de un sistema que integre todas las necesidades para una gestión adecuada y que permita, además, la realización de estudios de biodiversidad de forma totalmente integrada.

Las colecciones complejas requieren de un sistema que permita gestionar tanto la información existente, como los procesos asociados a las mismas. Es importante que el sistema intercambie información de forma estructurada y estandarizada con los proyectos existentes. Para ello, se han incorporado técnicas de inteligencia artificial que permiten que el sistema no actúe únicamente como un repositorio de datos, proporcionando al usuario la extracción de conocimiento existente en las colecciones biológicas.

En concreto esta memoria se organiza del siguiente modo:

- En el Capítulo I se realiza un estudio de las áreas de aplicación de la informática en la Biología, centrándonos en:
  - La informatización de las colecciones biológicas.
  - Los proyectos nomenclaturales existentes para el tratamiento de los nombres aceptados y sinónimos existentes.
  - La realización de estándares que permiten el intercambio de datos biológicos.
- En el Capítulo II analizamos los sistemas *software* existentes que permiten la gestión de colecciones biológicas.
- En el Capítulo III se presentan las bases del sistema realizado que permite gestionar las colecciones teniendo en cuenta, tanto las necesidades de gestión de los centros como las necesidades de investigación de los usuarios del centro.



- En el Capítulo IV se exponen las capacidades que aporta la utilización del Sistema Multiagente para el tratamiento y extracción de los datos incorporando técnicas de Web Mining, extracción y movilización del conocimiento.
- A continuación se exponen las conclusiones finales que se derivan de esta memoria y los principales caminos de investigación que han surgido durante el desarrollo de la misma.
- La memoria concluye con la Bibliografía empleada y dos apéndices:
  - El Apéndice A trata sobre los estándares, establecidos por la FIPA, para la realización de sistemas MultiAgentes.
  - El Apéndice B se realiza una introducción al lenguaje XML.



## Capítulo I. Aplicaciones para la bioinformática y estándares.

Las colecciones biológicas se caracterizan por contener un gran volumen de datos. El tratamiento masivo de datos, la extracción de información, el acceso a los datos, etc. justifican la necesidad de las Ciencias de la Computación y la aplicación de técnicas de inteligencia artificial, que permitan la gestión de las colecciones, y la extracción y movilización del conocimiento asociada a las mismas.

Existen múltiples definiciones del término **Bioinformática**, la mayoría de ellas hacen referencia del uso de la informática a aplicaciones relacionadas con el procesamiento del genoma humano (Lesk, 2002; Nicholas M Luscombe, Dov Greenbaum, 2001; Soberón and Peterson, 2004), pero este es un uso estrictamente restrictivo del término. La Bioinformática es un área interdisciplinar entre la biología, la computación y las ciencias de la información necesaria para la gestión, procesamiento y análisis de los datos generados por las entidades biológicas, independientemente de la escala de estudio (Govindaraj, 2001). Actualmente se está acuñando un término más específico, “Informática para la Biodiversidad” (*Biodiversity Informatics*) (Brisby, 2000; Soberón and Peterson, 2004) que se aplica a la sinergia entre la computación y la biología.

En el presente capítulo se van a tratar las distintas áreas de aplicación de la Informática para la Biodiversidad y se presentará el estado del arte atendiendo a las necesidades más urgentes de los sistemas actuales. La incorporación de las nuevas tecnologías al campo de la biodiversidad abre la posibilidad de intercambiar la información asociada a las colecciones biológicas. Esto ha hecho necesario el desarrollo de estándares de comunicación e intercambio que hagan posible el uso cooperativo de información.

Un paso previo a compartir información, es la localización en formato digital de la misma. Por ello, debemos hablar del tratamiento informático de las

colecciones, el cual posibilita que los datos estén en formato digital. A continuación se van a analizar las distintas áreas implicadas en el tratamiento de las colecciones, su interrelación, y los principales proyectos existentes en las mismas.

## **1. Áreas de aplicación de la Informática para la Biodiversidad**

En un estudio realizado por Stanley titulado: “*An overview of Biodiversity Informatics*” (Stanley D. Blum, 2000) se establecen las principales áreas de aplicación de la Informática para la Biodiversidad:

1. Identificación mediante caracteres.
2. Gestión de los Nombres taxonómicos y clasificación.
3. Gestión de los Datos de Especímenes.

Las tres áreas descritas por Stanley definen, a *grosso* modo, el conjunto de tareas necesarias para llevar a cabo la gestión de los especímenes. En un primer paso, se identifica el espécimen por sus caracteres para, de esta manera, asignarle el nombre científico aceptado. Una vez que se tiene toda la información necesaria sobre el espécimen, éste debe ser conservado e informatizado para posteriores estudios.

### **1.1. Identificación mediante caracteres**

Después de recolectar los especímenes comienza la labor de identificación. En el proceso de identificación, el taxónomo realiza un análisis de los atributos del espécimen. Este estudio tiene como objetivo determinar los caracteres que permiten asignarlo a un *taxon* determinado.

### ***Descripción y claves de identificación***

La descripción de una nueva especie es consecuencia de tener uno o más especímenes con una combinación de caracteres desconocida. Para ello, es necesaria una gran labor de investigación y de documentación que garantice que se trata de una nueva especie para la ciencia, y no de un organismo ya descrito con anterioridad, lo que daría lugar a confusión, sinonimias, etc.

En el proceso de descripción de una nueva especie, el taxónomo tiene que:

1. Describir el nuevo *taxon* de forma inequívoca para que otros investigadores puedan identificarlo.
2. Publicar ésta primera descripción, llamada descripción diagnóstica, y asignarle al nuevo organismo un nombre correcto. Ambos hechos han de ajustarse a las normas del Código Internacional de Nomenclatura.

Una descripción taxonómica enumera las características de un grupo de organismos. Una buena descripción es difícil de elaborar ya que debe de incluir los caracteres que según Stanley (Stanley D. Blum, 2000):

2. Son compartidos por los individuos del *taxon*.
3. Distingan el *taxon* de *taxa* similares.
4. Varíen dentro del *taxon* pero que no tengan importancia taxonómica.
5. Infieran la posición del *taxon* en la clasificación taxonómica.

### ***Gestión de la descripción.***

Las descripciones y las claves son las principales herramientas para poder realizar una identificación correcta de los especímenes. El uso de esta información es fuente de investigaciones para el desarrollo de sistemas de generación de claves automáticas (Gibaja, 2004) y para la gestión de las descripciones existentes. La aplicación de técnicas de Inteligencia Artificial permite la construcción de

sistemas Expertos (Gibaja, 2004) que ayudan a la identificación del espécimen teniendo en cuenta los caracteres que este presenta.

Las descripciones y las claves se componen del mismo tipo de información. Ambas podrían ser almacenadas en una matriz (Fermanian and Michalski, 1989), donde la primera dimensión hace referencia al *taxon* y las otras a sus características. En la matriz, cada celda almacena una descripción del estado observado para cada característica del correspondiente *taxon*. Muy poca información referente a descripciones taxonómicas está estructurada y adecuadamente informatizada mediante algún *software*.

El grupo internacional de trabajo en bases de datos taxonómicas, conocido como T.D.W.G. (*Taxonomic Databases Working Group*), ha abordado el desarrollo de estándares para representar información taxonómica en bases de datos biológicas. En su seno, el subgrupo de trabajo de estructuras de datos descriptivos, conocido como S.D.D. (*Structure of Descriptive Data*) comenzó a trabajar en 1998 y en su encuentro de 1999 celebrado en Harvard, concluyó que el subgrupo debía analizar los requerimientos para un nuevo estándar interoperativo de información descriptiva, este se denominaría: estándar SDD.

## **1.2. Gestión de los Nombres taxonómicos y clasificación.**

El nombre científico asignado a un espécimen indica la especie a la que pertenece. Por ello, hay que identificar la muestra. El taxónomo juega un papel muy importante en este proceso, puesto que es la persona con capacidad para ello.

### ***¿Qué es la taxonomía?***

Lawrence en la introducción a su obra (Lawrence G.H.M., 1951) (en (Bridson and Forman, 1992)) define la taxonomía como la ciencia que se ocupa de la identificación, nomenclatura y la clasificación de objetos, restringidos usualmente a objetos de origen biológico.

Para completar la definición propuesta por Lawrence hay que describir los términos de identificación, nomenclatura y clasificación:

1. **Identificación:** es la adjudicación del espécimen a un *taxon* por ser idéntico a, o similar a, otro que ya es conocido. Si esto no es posible, podría tratarse de un nuevo *taxon* por no estar aun descrito.
2. **Nomenclatura:** es la asignación del nombre científico aceptado de acuerdo al sistema nomenclatural. Por ejemplo, en el caso de la botánica, existe el Código Internacional Botánico de Nomenclatura (CIBN) que regula el *status* de los nombres científicos.
3. **Clasificación:** Es la organización de los organismos en grupos o *taxa* en categorías jerarquizadas. Por ejemplo, la pertenencia a una especie implica la pertenencia a un determinado género, la de un género a una familia, y así hasta los niveles más altos de la jerarquía del sistema de clasificación empleado.

La taxonomía adquiere cada vez más importancia ya que con más urgencia necesitamos respuesta a preguntas (Bridson and Forman, 1992) como las que siguen:

1. ¿Cómo se puede reconocer una especie? (**Identificación**)
2. ¿En qué orden debería ser nombrada una especie para que pueda ser tratada sin ambigüedad? (**Nomenclatura**)
3. ¿ Cuáles son sus parientes más cercanos?¿Hay otras especies que tengan propiedades similares o sistemas genéticos compatibles?  
(**Clasificación**)
4. ¿Dónde pueden crecer? (**Distribución**)
5. ¿En qué clase de hábitat pueden vivir? (**Ecología**)
6. ¿Tiene algún uso o propiedad? (**Uso**)

El papel de la taxonomía reside en poder responder de forma rápida y precisa a estas preguntas. La capacidad de respuesta depende en gran medida de la calidad y lo completas que estén las colecciones biológicas existentes en los Centros que las gestionan.

El nombre científico de un espécimen se establece como fuente para su posterior referencia y catalogación. Dada la complejidad existente en el proceso de nomenclatura, la existencia de sinónimos en las bases de datos biológicas complica la cuantificación de las muestras existentes para una especie.

Según Krishtalka y Humphrey, en la actualidad se estima que hay entre 13 y 14 millones de especies en general, de las cuales, conocemos de entre 1.5 a 2 millones (Krishtalka and Humphrey, 2000). Además, se calcula que existe aproximadamente entre uno y dos sinónimos por cada uno de los nombres aceptados (Alroy, 2002). Por ello, recopilar una lista de los nombres para un grupo de especies nos llevaría años de esfuerzo.

Todo este proceso se complica aun más cuando se da un nombre atendiendo a un sistema de clasificación taxonómica. Existen diferentes sistemas de clasificación y dentro de cada sistema de clasificación el nombre será el aceptado o no con respecto a otros sistemas taxonómicos.

Existen distintas estandarizaciones internacionales, dependiendo de la disciplina, para establecer las características que debe de satisfacer un nombre para que sea el nombre aceptado. Por ejemplo, el Código Internacional de Nomenclatura Botánico (CINB), el Código Internacional de Nomenclatura de Zoología (CINZ), el Código Internacional de Nomenclatura Bacteriológica (ICNB), etc.

Veamos las características que tiene que reunir un nombre para obtener el *status* de aceptado. Para ello, y para poder concretar, veamos por ejemplo el Código Internacional de Nomenclatura Botánico (CINB). Este nos indica, mediante un conjunto de reglas y principios, las características que tiene que cumplir un nombre para que sea el aceptado.



## ***Código Internacional de Nomenclatura Botánica***

Todas las normas que controlan la creación de nombres científicos para las plantas y categorías taxonómicas están contenidas en este código.

El sistema de nomenclatura botánica ha sido estandarizado en un código para que las comunicaciones sobre plantas estén basadas en nombres mundialmente aceptados.

El Código Internacional de Nomenclatura Botánica nos dice que el nombre científico es el símbolo nominal de la planta o un grupo de plantas. Además es una manera de indicar su rango o categoría taxonómica.

El Código Internacional de Nomenclatura Botánica está dividido en tres partes (Tormo Molina, 2006):

- Principios básicos del sistema de nomenclatura botánica.
- Reglas para poner en orden la nomenclatura antigua.
- Recomendaciones para conseguir uniformidad y claridad en la nomenclatura actual.

En el Código Internacional de Nomenclatura Botánica se han establecido seis principios:

1. La independencia de la nomenclatura zoológica y bacteriológica.
2. La aplicación de nombres a los grupos taxonómicos (*taxa*) de categoría de familia o inferior es determinada por medio de los tipos nomenclaturales.
3. La nomenclatura de un grupo taxonómico se funda en la prioridad de publicación.

4. Cada grupo taxonómico, de delimitación, posición y rango dados, no puede tener más que un nombre correcto, es decir, el más antiguo de conformidad con la reglas, salvo las excepciones especificadas.
5. Los nombres científicos de los grupos taxonómicos se expresan en latín, cualquiera que sea su categoría.
6. Las reglas de nomenclatura tienen efecto retroactivo, salvo indicación en contra.

En el Código Internacional de Nomenclatura Botánica se establecen las reglas que tiene que cumplir un nombre científico (ver Figura 1) para que sea aceptado mundialmente. A continuación se muestra un conjunto reducido de las mismas:

1. Estar publicado en formato papel y disponible a todo el público o por lo menos a las instituciones botánicas.
2. Publicación en un formato específico y acompañado de una descripción o de una referencia a la descripción del *taxon* (ser una publicación válida).
3. Descripción legítima: en concordancia con todas las reglas del código.
4. Cualquier nombre científico previo, que esté publicado, sea válido y legítimo, debe ser elegido como el aceptado de acuerdo a la regla de la prioridad. Los demás nombres serán **sinónimos**.

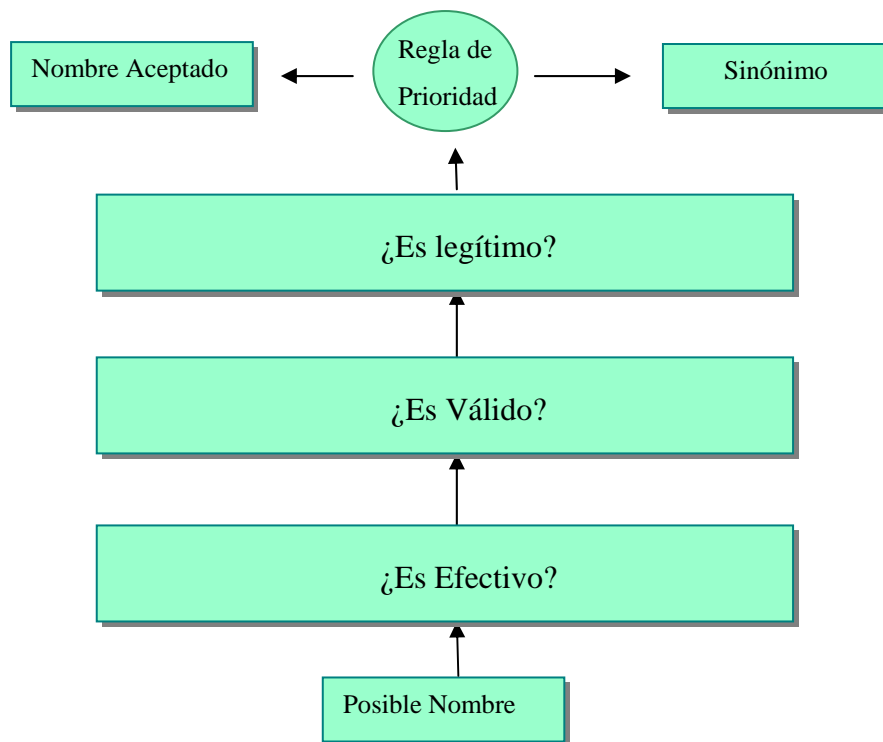


Figura 1. Proceso de determinación del *status* de un nombre científico

Esta es una versión reducida del código pero es suficiente para mostrar la complejidad del proceso a seguir para generar nombres válidos. Sin una estructura centralizada encargada de gestionar los nombres existentes y de establecer las prioridades, el uso de un nombre aceptado estaría condicionado al criterio del taxónomo y a la documentación consultada.

Dada la cantidad de nombres existentes y los problemas derivados de la aplicación del código de nomenclatura, surgen distintas iniciativas para realizar recopilaciones de los nombres científicos existentes. Podemos clasificarlos en dos grupos de proyectos:

- Los encargados de realizar una lista de nombres aceptados y sus correspondientes sinónimos (*checklist*)
- Los encargados de realizar una recopilación de los nombres usados, sin presentar opiniones referentes a la validez o aceptación del *taxon* correspondiente (*nomenclaturales*).

### Proyectos de tipo checklist

- **I.T.I.S.: Integrated Taxonomic Information System** (Sistema de Información para la Integración Taxonómica). Creado con la colaboración de cinco agencias gubernamentales de Estados Unidos: el Centro Geológico (U.S.G.S.: *U.S. Geological Survey*), el departamento de agricultura (U.S.D.A.: *U.S. Department of Agriculture*), la Administración Nacional oceánica y atmosférica (N.O.A.A.: *National Oceanic and Atmospheric Administration*), la Agencia de protección medioambiental (E.P.A.: *U.S. Environmental Protection Agency*) y el Museo Nacional de Historia Natural (N.M.N.H.: *National Museum of Natural History*). Todas las agencias y organismos se unen para ofertar toda la información existente en sus bases de datos de forma unificada.
- **Species 2000**. Proyecto de ámbito internacional que trata de proporcionar una lista global de todas las especies mediante el uso de bases de datos de especies (*Global Species Database, GSD*) federadas. Fundado por la Unión Internacional de Ciencias Biológicas (I.U.B.S.: *International Union of Biological Sciences*) en cooperación con el Comité sobre datos para la ciencia y la tecnología (CO.DATA: *The Committee on Data for Science and Technology*) y la Unión Internacional de la sociedad de microbiología (I.U.M.S.: *International Union of Microbiology Societies*) en Septiembre de 1994.

### Proyectos nomenclaturales

- **I.P.N.I.: International Plant Name Index** (Índice Internacional de nombre de plantas). Es el resultado de la colaboración de tres instituciones: el Real Jardín de Kew (R.B.G.K.: *Royal Botanical Garden of Kew*), el Herbario de la Universidad de Harvard (H.U.H.: *Harvard University Herbaria*), y el Jardín Botánico Nacional

Australiano (A.N.B.G.: *Australian National Botanic Garden*). Este proyecto es importante porque recoge una gran colección de plantas vasculares. Además recibe aportaciones libres de otros usuarios y tiene una editorial que publica toda la información.

- **I.O.N.: Index to Organism Names** (Índice de nombre de organismos). Este proyecto contiene todos los nombres de animales, plantas y virus contenidos en las bases de datos de Thomson BIOSIS. Con más de 1.4. millones de nombres, el proyecto se representa como la mayor recopilación de nomenclatura biológica.

La existencia de proyectos *checklist* nos ayuda a obtener los nombres aceptados y los sinónimos asociados a los mismos. Haciendo uso de proyectos como *ITIS* o *Species 2000* podemos filtrar los datos almacenados en las bases de datos biológicas proporcionando una limpieza de los mismos. Sin embargo, los principales problemas de estos proyectos son:

1. Proyectos internacionales que no recogen todas las casuísticas presentadas en las colecciones.
2. No existe una interfaz de consulta que permita acceder de forma fácil y unificada a la información existente en todos los proyectos.

### **1.3. Gestión de los Datos sobre los especímenes**

Las colecciones biológicas son las principales fuentes de información para estudios de biodiversidad. Éstas son mantenidas por los herbarios y museos de historia natural (Krishtalka and Humphrey, 2000). Como consecuencia de los estudios realizados sobre las colecciones biológicas, los especímenes son revisados continuamente.

Debido a que el catálogo de la vida cambia con frecuencia por la aparición de nuevas técnicas y teorías, la información asociada a las colecciones está siendo

continuamente actualizada y consultada. Por ejemplo, El museo de Zoología de Vertebrados de la Universidad de California (University of California, 2000) ha estimado que un tercio de sus 625.000 registros son actualizados cada año.

Podemos estructurar la información básica para un espécimen en tres categorías:

1. Información intrínseca al espécimen (identificación taxonómica, sexo,...).
2. Información que describen las circunstancias de sus recolección y de su contexto en la naturaleza (fecha de recolección, nombre del colector, hábitat, ...)
3. Otro tipo de información que pueda ser de interés por ejemplo: fotografías, videos,...

Todos estos datos son la base de los estudios taxonómicos, biogeográficos y para el análisis de la biodiversidad. De ellos se extrae la información que un taxónomo necesita cuando desea realizar un estudio para un grupo taxonómico. Un taxónomo, para estudiar un grupo taxonómico, necesita conocer los *taxa* existentes para saber qué material tiene que examinar (Cedric Raguenaud, Jessie Kennedy, 1999). El acceso a la información tiene que permitir al taxónomo procesar los datos, ordenarlos, contarlos y estudiar su distribución.

Por tanto, el taxónomo necesita acceder a toda la información que existe en un formato correcto y bien estructurado. La mayoría de las instituciones no tienen sus colecciones informatizadas, lo cual complica este acceso a la información.

De las áreas de aplicación establecidas por Stanley (Stanley D. Blum, 2000), la principal es la gestión de los especímenes, dada su importancia como base para posteriores tratamientos computacionales.

La correcta informatización de las colecciones permite:

1. Un intercambio estructurado y estandarizado.
2. Un acceso simple y automático.
3. Enviar y recibir información en formato unificado.

Dada la complejidad de las colecciones biológicas y la forma particular de gestión por parte de los centros, cada institución tiende a desarrollar su propio *software*. Esto conlleva la aparición de multitud de esquemas de datos y que sólo son accesibles por los miembros de esa institución, perdiendo las ventajas anteriormente citadas y asociadas a la informatización de colecciones.

Las colecciones biológicas tienen gran cantidad de datos que son fuente de investigaciones en el campo de la supercomputación destinadas al tratamiento masivo de datos para análisis de biodiversidad: el modelado de la distribución de una especie, y la visualización espacial de las muestras existentes, etc.

En general, podemos afirmar que en la Taxonomía y la Sistemática encontramos un frente de trabajo importante, pues la información que se maneja es compleja tanto en lo referente al diseño de una estructura de almacenamiento como al vocabulario con el que se trabaja (muy diverso y variable) (Berendsohn, 2001). Por eso, uno de los objetivos actuales reside en la creación de un sistema de información de bases de datos biológicas que permitan el intercambio de los datos entre distintas instituciones y/o proyectos de investigación. Además, permita la integración de distintas fuentes heterogéneas que facilite la información necesaria para la realización de análisis de biodiversidad.

## **2. El tratamiento de las colecciones.**

La información existente en las colecciones biológicas son de gran utilidad ya que cada espécimen almacenado en ellas es una muestra de lo que hubo en un

lugar y en un momento determinado. El paso previo a todo estudio pasa por la informatización y gestión de las colecciones. Una vez que las colecciones biológicas son informatizadas están disponibles para su tratamiento computacional.

A continuación se describen las principales líneas de investigación descritas por Stockwell (Stockwell, 1997) en el tratamiento de las colecciones sobre biodiversidad (bases de datos, sistemas de información geográfica, predicción, modelado y distribución, etc.). Stockwell distingue entre aplicaciones básicas, aplicaciones avanzadas dependientes de las básicas y las aplicaciones de carácter general (ampliamente investigadas) no específicamente desarrolladas para la biología, para hacer referencia a todas las posibilidades de actuación de la Informática para la Biodiversidad.

Dentro de las aplicaciones básicas podemos encuadrar tres secciones:

- **Bases de datos biológicas:** la informatización y puesta en marcha de bases de datos biológicas que estructuren y recojan todas las muestras existentes es una ardua tarea que no siempre está apoyada por los “conservadores” o gestores de esta información. La informatización supone un gran avance para la extracción e intercambio de información.
- **Bases de datos de variables externas relevantes desde el punto de vista biológico:** Existen determinadas variables que son especialmente relevantes desde el punto de vista biológico pues nos permiten responder preguntas sobre el por qué y dónde viven los organismos (ejemplos de estas variables son el clima, radiación, la geología y los procesos humanos y biológicos). En la actualidad son muy escasas las fuentes de información de este tipo accesibles. El problema se agrava cuando necesitamos una escala concreta. El tamaño de los datos a almacenar se dispara a medida que la granularidad de los mismos aumenta.
- **Modelado de la distribución de especies:** La información biológica disponible sobre la distribución de las especies es casi siempre muy



heterogénea e incompleta. Por ello, resulta de gran utilidad la investigación en modelos no lineales para analizar la respuesta de las especies ante el ambiente y que den respuestas fiables ante la presencia de datos escasos. Para este fin se necesita el apoyo de métodos más complejos y computacionalmente costosos como, por ejemplo, la utilización del sistema GARP (Stockwell and Peters, 1999)(*Genetic Algorithm for Rule Set Production*) para la producción de reglas mediante algoritmos genéticos. GARP realiza la modelización de especies mediante la aplicación de técnicas de Inteligencia Artificial bioinspiradas.

Estas serían las tres aplicaciones básicas descritas por Stockwell. Veamos ahora las aplicaciones avanzadas:

- **Modelado de Biodiversidad:** Los estudios sobre la biodiversidad están íntimamente ligados al estudio de las variaciones en las poblaciones en función de los cambios ambientales. Generalmente se simplifica la representación de la biodiversidad de una región utilizando variables agregadas (por ejemplo, el tipo de vegetación o especies buenas indicadoras de la presencia de otras) que se denominan sustitutos ambientales (*environmental surrogates*). Esta simplificación puede llevar en ocasiones a sistemas algo pobres. Entre las posibles alternativas está el desarrollo de mapas predictivos para cada una de las especies. Aunque esta aproximación es costosa computacionalmente hablando, esta alternativa proveería información novedosa y nos permitiría obtener nuevos patrones para distintos grupos taxonómicos.
- **Modelado de los procesos Ecológicos:** La ecología incorpora un aspecto dinámico muy fundamental para los estudios en biodiversidad (migraciones estacionales, invasiones de organismos exóticos y enfermedades, etc.). La combinación de modelos ecológicos con el componente espacial introduce un mayor grado de exactitud y realismo y, puesto que los estudios en biodiversidad son *per-se* computacionalmente

costosos, la adición de estos aspectos dinámicos repercute en un incremento adicional de la complejidad.

- **sistemas de planificación de áreas protegidas:** El problema de la planificación de las reservas reside en el diseño de áreas naturales protegidas que contengan de forma adecuada y eficiente una muestra extensa del ambiente natural. Incluso los enfoques más sencillos para acometer este problema son computacionalmente costosos, lo que ha limitado el análisis a simplificaciones bastante toscas que sólo garantizan el óptimo bajo condiciones muy estrictas que generalmente no se dan (por ejemplo, un análisis de planificación de aproximación geográfica realizado (G.A.P.: *Geographic Approach to Planning for Biological Diversity*) (Jennings, 1993; Scott, Tear, 1996)).

Estas son las aplicaciones avanzadas que necesitan de las aplicaciones básicas para poder generar modelos y procesar los datos existentes. Veamos ahora las aplicaciones más comunes y que se usan en conjugación con las anteriormente descritas:

- **Estimación de los límites de Predicción:** Tan importante como la precisión de una predicción es la estimación de la confianza o los límites de confianza estadística de la misma. Cuando tratamos con variables estadísticas muy bien definidas podemos calcular los límites de confianza de forma muy sencilla, pero cuando los datos no son adecuados, debemos aplicar otros métodos que requieren generalmente un gran esfuerzo computacional. Los conjuntos de datos biológicos y ambientales nunca cumplen las suposiciones utilizadas por la metodología estadística esto es: las distribuciones no son normales o existen dependencias entre variables. Por esto cobra gran importancia la utilización de métodos de remuestreo (*resampling methods*) para obtener estimaciones más robustas de la confianza en las predicciones.

- **Visualización y generación de Mapas:** La herramienta más utilizada para la visualización en el área de la distribución, la ecología y la conservación es el mapa. El medio ambiente es multi-dimensional y contiene complejas interrelaciones haciéndose necesarios nuevos métodos de visualización.
- **Interactividad:** La utilización de sistemas de Información Geográfica (SIG) ha cobrado un gran interés. Los interfaces para mapas interactivos en la Web proporcionan mejor acceso a las bases de datos sobre biodiversidad y añaden flexibilidad a la visualización de mapas (rotaciones, cambios de escalas, etc.).

Con el estudio realizado por Stockwell se han identificado las áreas de investigación más recientes (ver Tabla 1) estableciendo un marco de trabajo y de avance en la Informática para la Biodiversidad.

Áreas de investigación	Objetivos
Base de datos de especies	Creación de bases de datos internacionales distribuidas de localizaciones de especies. Para el 2001 se esperaban más de un billón de registros georreferenciados.
Variables externas relevantes desde el punto de vista biológico.	Diseño de herramientas que suministren conjuntos de datos relevantes desde el punto de vista biológico a una escala y una región geográfica definida por el usuario.
Modelado de la distribución de especies.	Incrementar la exactitud y disponibilidad de modelos para la distribución de especies. Servicios como el Web permitirán acceder a mapas más exactos de forma más rápida.
Modelado de la biodiversidad	Predicción de la biodiversidad para grupos taxonómicos a partir de la predicción de especies individuales.
Modelado de procesos ecológicos	Incorporación de procesos ecológicos dentro de modelos de biodiversidad de gran resolución espacial.
sistemas para la planificación de áreas protegidas	Delimitación de áreas que contengan de forma adecuada y eficiente una muestra extensa del ambiente natural.
Límites de confianza estadística	Mejorar la robustez del modelado utilizando métodos de remuestreo.
Visualización/Realización de mapas	Métodos mejorados para visualizar datos sobre la Biodiversidad complejos y de múltiples dimensiones.
Interactividad	Provisión de información interactiva vía WWW.

Tabla 1. Áreas de investigación en biodiversidad computacional (Stockwell, 1997)

Muchas disciplinas científicas se han adaptado y aprovechado de los avances tecnológicos. Sin embargo, esto no ha sucedido en el campo de la biodiversidad. Esto es debido a que los biólogos no reciben la formación necesaria en Ciencias de la Computación. Este es el principal motivo de la sinergia necesaria entre biólogos e informáticos para el desarrollo de herramientas que

faciliten el procesamiento de los datos y las necesidades computacionales existentes.

## **2.1. Informatización y gestión de colecciones**

Dada la importancia en la informatización y gestión de las colecciones biológicas, en este apartado se realiza un análisis de los procesos y procedimientos llevados a cabo por los centros gestores de las colecciones. La descripción de los procesos llevados a cabo, nos va a permitir identificar los puntos en los que la informática puede actuar.

Para un espécimen se pueden establecer los siguientes procesos de gestión:

1. Determinación de los especímenes.
2. Informatización de las muestras.
3. Conservación y uso.
4. Préstamos.
5. Revisión y actualización.

Existen procesos internos, tal como la preparación de las muestras para su posterior conservación, tratamientos con insecticidas y procesos de congelación que quedan fuera del objetivo de la informatización.

### **2.1.1. Determinación de los especímenes**

Previo a su conservación, los especímenes han de ser determinados, es decir, identificados y encuadrados en una jerarquía taxonómica.

Una vez que el espécimen es identificado se le asigna un nombre científico aceptado y un número de catálogo que identifica unívocamente la muestra en la colección.

### **2.1.2. Informatización del Material**

Además de la información que asocia el taxónomo a la muestra, es necesario incluir información referente a la recolección (fecha, persona que recolecta, localización, georreferenciación,...).

El diseño de la estructura de almacenamiento para la informatización es complicada puesto que depende, en parte, de la política del Centro. Éste es uno de los principales problemas en la gestión de las colecciones, puesto que cada entidad cree tener unas necesidades diferentes a las demás, por lo que normalmente, cada Centro desarrolla su sistema de información (Bedathur and Haritsa, 2000).

Se puede establecer un conjunto de requerimientos comunes para todos los sistemas de información:

1. Una base de datos centralizada que permita a los usuarios trabajar de forma remota.
2. Una interfaz de consulta amigable y que permita realizar consultas avanzadas sin conocimientos en lenguajes de consulta de bases de datos.
3. Realizar un seguimiento y control del material del Centro.
4. Impresión de etiquetas para que sean incorporadas en las muestras que forman la colección.

Es necesario que los datos informatizados estén centralizados para poder simplificar los procesos de consulta y el acceso a ellos. Los centros gestores

actúan como centros de documentación teniendo que proporcionar servicios de consulta externa. Éstos se realizan a través del contacto directo con las personas encargadas de las colecciones.

### **2.1.3. Conservación de los especímenes y su uso**

Las muestras etiquetadas son almacenadas y conservadas para el uso de cualquier investigador que lo solicite. En este sentido, el centro actúa como centro de documentación que proporciona servicios de consulta a los investigadores.

Estos Centros reciben continuamente peticiones de investigadores que necesitan realizar estudios sobre grupos taxonómicos concretos y requieren conocer todo el material existente para un determinado *taxon* o grupo. Una vez que el personal del Centro realiza las consultas pertinentes, remite los informes oportunos al investigador que realizó la consulta. En algunos casos, el investigador requiere el préstamo de los especímenes para poder observarlos y estudiarlos más detenidamente. Por esta razón el centro tiene que gestionar los materiales prestados.

### **2.1.4. Préstamos de Especímenes.**

Para realizar cualquier estudio biológico es necesario consultar el material conservado en estos Centros. En algunos casos es necesario tener el ejemplar para poder examinarlo.

En el caso de requerir del material cabrían dos posibilidades:

1. El desplazamiento del investigador al Centro para poder manipular y estudiar el espécimen.
2. Solicitar un préstamo al centro. Estos préstamos se realizan a otros centros o investigadores por un periodo limitado de tiempo.

Los préstamos pueden ocupar gran parte del tiempo del conservador del centro. Además, el préstamo del material implica que el centro se pueda quedar temporalmente sin muestras del material solicitado.

Es importante controlar el material prestado para saber la situación actual del Centro y del material ausente. Además, se debe controlar los periodos de devolución del material que tienen las instituciones y solicitarlo en caso de sobrepasar el periodo de validez del préstamo.

La existencia de préstamos conlleva, principalmente, dos problemas al centro:

1. La conservación del material. Debido a la manipulación, la muestra se deteriora.
2. Ausencia total de muestras de un *taxon* en el Centro. Este hecho repercutiría en la consulta de otros investigadores al centro puesto que podrían no manipular la muestra.

El material prestado es estudiado y revisado, completando así la información existente para la muestra. Se incorporan revisiones sobre la determinación del espécimen. Éstas son confirmaciones o nuevas determinaciones del nombre científico asignado al espécimen.

### **2.1.5. Revisiones de los Especímenes.**

Los nombres científicos otorgados al espécimen son revisados por los especialistas, por tanto, tras el estudio del material prestado por el centro, éste es devuelto con anotaciones asociadas que:

1. Confirman el nombre científico con el que el espécimen fue determinado.

2. Modifican el nombre científico del espécimen. El revisor lo indica mediante una etiqueta de revisión.
3. Indican la existencia de más de un espécimen, no referenciado, en la misma muestra. Por error en la preparación del material pueden existir dos o más especímenes dentro de la misma muestra. Este hecho se informa al Centro para que incluya también la etiqueta de adición pertinente. El sistema tiene que contemplar ésta posibilidad.

Todas las observaciones que se realizan por parte de los revisores tienen que ser almacenadas en la base de datos para que conste en la historia del espécimen para el cual se realizó la revisión.

Los préstamos de material son la retroalimentación que permita que la información del Centro esté actualizada y sea de la máxima calidad posible.

### **2.1.6. Resumen del tratamiento del espécimen**

El espécimen sufre una serie de procesos por los cuales se van completando la información que tiene que ser almacenada y estructurada mediante bases de datos. Podemos distinguir tres procesos principales:

- **El etiquetado:** donde se identifican los especímenes e indican las condiciones de su recolección: su procedencia geográfica, fecha de recolección o captura, condiciones ecológicas, etc.
- **La informatización:** donde se vuelca toda esta información en una base de datos.
- **Proceso de revisiones:** información de las revisiones realizadas sobre la muestra.

A continuación, se muestra un esquema del proceso y de la secuencia en la inserción de información en la Base de Datos del centro (ver Figura 2):



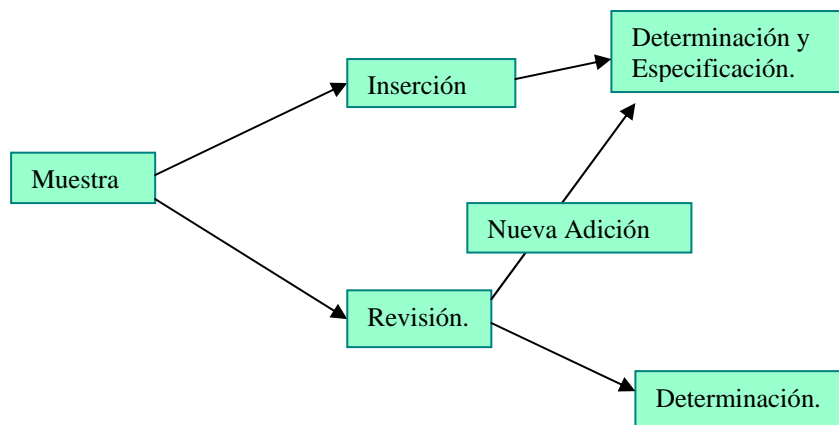


Figura 2. Proceso de informatización y flujo de información.

Actualmente, se han formado grupos de trabajo dentro del grupo internacional de bases de datos taxonómicas para la estandarización de las bases de datos biológicas y establecer un marco común de intercambio entre las instituciones que gestionan colecciones biológicas. A partir de ahí, surgen iniciativas como *GBIF (Global Biodiversity Information Facility)* y *BioCASE (Biological Collection Access Services)* las cuales promueven el intercambio de información. En esta línea, se ha llevado a cabo una intensa investigación en el campo del diseño de modelos de bases de datos para la representación de información taxonómica (Berendsohn, 1997b; Pullan, Watson, 2000).

### 3. Estándares Internacionales

El acceso e integración de la información ofertada por proyectos internacionales, mejoran el trabajo del taxónomo y completan la información existente en la colección.

Como consecuencia, hay que estudiar las distintas formas de comunicarse con los proyectos existentes. Iniciativas como *GBIF (Global Biodiversity Information Facility)* y *BioCASE (Biological Collection Access Service)* promueven el desarrollo de estándares de intercambio que unifican los sistemas de bases de datos heterogéneas existentes.

En este apartado se van a tratar los principales proyectos existentes para el intercambio de datos y estandarización de los mismos.

### **3.1. Iniciativas para la estandarización**

A continuación se detallan dos iniciativas llevadas a cabo para estandarizar y compartir los datos existentes en los centros gestores. Ambas iniciativas han trabajado y trabajaban para unificar el acceso a los datos de los proveedores.

#### **3.1.1. GBIF: Global Biodiversity Information Facility**

GBIF (GBIF, 2001) es una iniciativa internacional a 10 años vista para divulgar a través en Internet, de forma gratuita, toda la información disponible sobre los organismos vivos conocidos a nivel mundial. GBIF se concibe como una red de bases de datos interconectadas que pretende ser una herramienta básica para el desarrollo científico de los países y contribuir significativamente a una mejor protección y uso de la biodiversidad en el planeta.

En la introducción existente en (Unidad de Coordinación de GBIF en España, 2006c) nos muestra una descripción más extensa sobre GBIF. A continuación mostramos un extracto de la misma. Se inicia en 1996 a través de un grupo de trabajo de la Organización para el Desarrollo y Cooperación Económica denominado "*Mega Science Forum Working Group*" cuyo principal interés es lanzar iniciativas científicas de interés fundamental pero que por su escala no eran abordables por país alguno. En este foro surge el concepto de GBIF con la idea de aplicar la informática como mecanismo para facilitar y administrar información proveniente de la naturaleza.

Las prioridades de GBIF se concentran en el nivel de organismos, y de este nivel, la prioridad inicial son las colecciones. Los cuatro programas en curso de GBIF son:

1. *DADI: Data Access and Database Interoperability*: Acceso a los datos e interoperabilidad entre bases de datos.
2. *DIGIT: Digitalisation of Natural History Collections*. Digitalización de colecciones de historia natural.
3. *ECAT: Electronic Catalogue of Names of Know Organism*. Catálogo electrónico de nombres conocidos.
4. *OCB: Outreach and Capacity Building*. Formación y cooperación

GBIF pretende ser una red lo suficientemente flexible para proveer datos y para dar soporte a usuarios. Además, pretende consolidarse como la base principal de consulta de los investigadores en Biodiversidad y usuarios en general.

Para ello se debe de integrar las distintas bases de datos existentes en el mundo. La existencia de distintos estándares permite el flujo de los datos desde las fuentes a GBIF.

Actualmente, existen distintos estándares de representación de información y distintos tipos de rangos o granularidad de la información que se puede intercambiar o almacenar dependiendo del estándar escogido.

Los objetivos que GBIF pretende conseguir son los siguientes:

1. **Información sobre el *Taxon***: proporcionar información a nivel de espécimen y/o observaciones permitiendo tener información de qué es lo que existen en cada lugar del mundo.
  - a. Registros del espécimen.
  - b. Observaciones.
2. **Datos Taxonómicos**:
  - a. Nombres Científicos: la existencia de más de un nombre para determinar a una misma especie genera la existencia de sinónimos.

La identificación de qué nombres son sinónimos es un objetivo a cubrir por GBIF (Proyecto de Nomenclatura):

- i. Nombres científicos.
- ii. Nombres vernáculos o comunes.

**3. Descripciones taxonómicas:** Construir una red mundial de información taxonómica que permita a otros investigadores identificar y determinar a una especie.

**4. Enlaces a otros niveles de información del *taxon*:**

- a. Información sobre la biología del *taxon* y su historia
- b. Interacciones Ecológicas.
- c. Datos genéticos.
- d. Sonidos e imágenes.

GBIF pretende establecerse como una red para el intercambio de datos ofertando dos características fundamentales:

1. Proporcionar la información almacenada en Herbarios y Museos de Historia Natural.
2. Proporcionar un marco común de intercambio en el que los investigadores puedan consultar y publicar sus resultados.

Los usuarios necesitan datos estructurados y bajo un mismo estándar, siendo éste el principal problema. Existen muchas fuentes de información y cada una de ellas está estructurada de forma distinta, respetando o no, algunos de los estándares existentes. Esto hace que GBIF se plantee como primer paso para la consecución de sus objetivos, tener la información en un formato único y común.

Para poder intercambiar los datos entre las distintas fuentes de datos, en un principio se evaluó la posibilidad de imponer un formato común de bases de datos para almacenar la información en todas las fuentes. Este proyecto es inviable, puesto que implica poner de acuerdo a toda la comunidad investigadora y rediseñar el *software* existente.

Por ello surge una segunda solución, donde la información es obtenida de las fuentes de datos con distintos formatos, y es agregada en un formato común y bajo un estándar de intercambio. Es necesario construir sistemas intermedios o mediadores que obtengan los datos de las bases de datos y lo estandaricen a un formato común.

Debido a la heterogeneidad de las fuentes de datos hay que establecer un marco común de intercambio que refleje las necesidades del usuario final. GBIF propone el conjunto mínimo de datos que se deben proporcionar por cada fuente (Proveedor de datos):

1. Nombre del *taxon* al cual ha sido asignado el espécimen.
2. Identificador de la muestra en la colección.
3. Localidad donde el espécimen fue recolectado o la observación fue hecha.
4. Fecha de recolección u observación.
5. Lugar de recolección del espécimen y cómo acceder a más información del espécimen.

Esta información debe de ser integrada por GBIF, en un mismo formato, de entre las distintas fuentes de datos. Para este objetivo existen dos alternativas de estándares:

1. **Darwin Core**: estándar para el intercambio de información de especímenes u observaciones. Promovido por GBIF y el proyecto DiGIR, que posteriormente trataremos.

2. **ABCD Schema:** Estándar que permite mayor nivel de detalle que el establecido por *Darwin Core*. Fue promovido por el proyecto BioCASE y el grupo internacional de trabajo en bases de datos taxonómicas (TDWG), para el intercambio de información de especímenes u observaciones.

Tanto el estándar *Darwin Core* como el estándar *ABCD Schema* intercambian información a nivel de espécimen u observación, consiguiendo uno de los principales objetivos de GBIF. Actualmente, GBIF está trabajando con el grupo internacional de Bases de Datos Taxonómicas (TDWG) y otros, para establecer un estándar que permita el intercambio de datos referentes a nombres y descripciones taxonómicas (proyecto ECAT).

Todos los proveedores de datos facilitan su información en un estándar dependiendo del tipo de dato, y GBIF proporciona la infraestructura necesaria para integrarlo y acceder a ellos.

Podemos denominar como **mediadores** (*wrappers*) a los sistemas intermedios que procesan y estructuran los datos en un formato común (ver Figura 3). GBIF obtiene los datos de los proveedores a través de los mediadores en el formato necesario.

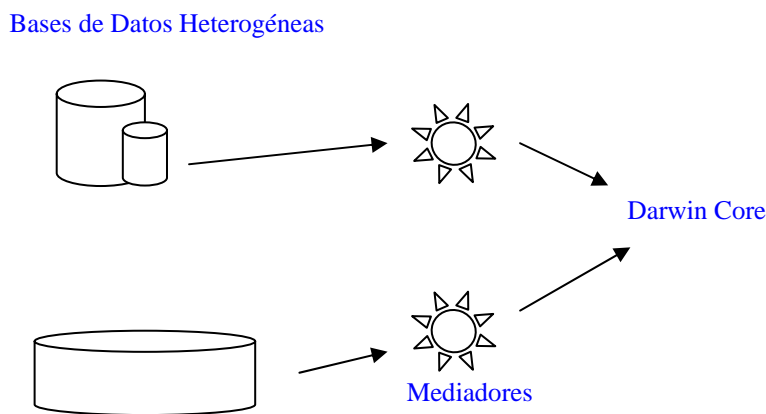


Figura 3. Integración de fuentes heterogéneas

GBIF pretende ser la principal red de consulta y de publicación de los resultados obtenidos por los investigadores a nivel mundial. Mediante la utilización de mediadores se realiza el intercambio de la información contenida en los distintos nodos o proveedores de información de una forma común. Esta forma de actuar hace que todos los proveedores tengan la posibilidad de incorporar sus datos a la red sin necesidad de realizar cambios sustanciales en sus bases de datos.

### 3.1.2. BioCASE: Biological Collection Access Service

El servicio para el acceso a las colecciones biológicas para Europa, BioCASE (Berendsohn, Döring, 2002), es una red europea de las colecciones biológicas de todas las clases. Aunque el proyecto BioCASE finalizó en el 2004, la red se mantiene bajo el proyecto SYNTHESYS (*Synthesis of systematic resources*), el cual tomó el relevo para crear la infraestructura Europea para investigadores en Historia Natural. Éste permite el acceso unificado a los datos distribuidos y heterogéneos de bases de datos observacionales haciendo uso de *software* de libre acceso.

BioCASE estaba basado en los siguientes proyectos:

- CDEFD: *A Common Datastructure for European Floristic Databases* (Berendsohn, 1997a). Estructura de datos Común para Bases de Datos Florísticas Europeas.
- BioCISE: *Biological Collection Information Service in Europe*. Servicio de Información de Colecciones Biológicas en Europa (Berendsohn, 2000).
- ENHSIN: *European Natural History Specimen Information Network* (ENHSIN, 1999). Red Europea de Historia Natural.

La unión de los anteriores proyectos sirvió como base para construir un macro proyecto que une los datos biológicos de los proveedores.

La principal meta a cubrir por este proyecto fue la unificación de las colecciones Europeas. Actualmente, las colecciones están dispersas e inaccesibles. Por esto, se pretendió construir un sistema que permitiera ofertar información de una forma unificada de todos los datos contenidos en los proveedores europeos.

El desarrollo de BioCASE, como un avance en el acceso electrónico, se basó en cuatro áreas de investigación:

1. **Organización de los flujos de Información.** Los nodos nacionales organizan la información y los exportan a la red central. El proyecto proporciona el *software* necesario para la recopilación de la información y para respetar un mismo estándar de salida de datos. Los nodos nacionales proporcionan metadatos sobre las colecciones al sistema central y este recibe información sobre la calidad de los datos enviados.
2. **Desarrollo de Tesoros.** Evaluación de los catálogos de datos existentes, para la creación de un Tesoro (principalmente a la taxonomía y la geocología).
3. **Interfaz de Usuario e Indexación.** Desarrollo de un sitio Web y una interfaz de usuarios para permitir el acceso a las bases de datos existentes en los nodos nacionales
4. **Las necesidades del usuario.** El proyecto cuenta con cientos de investigadores, principalmente en ecología y sistemas de clasificación. Se ha incluido un sistema de retroalimentación que permite al usuario indicar sus opiniones o sugerencias sobre el sistema.

El objetivo del proyecto rebasaba las barreras de las aplicaciones desarrolladas hasta el momento: soluciones aisladas y locales de los diferentes organismos, para poder dar al usuario una interfaz común de consulta de toda la información contenida en bases de datos distribuidas y heterogéneas.



BioCASE estableció la infraestructura necesaria para poder acceder a los datos primarios procedentes de Herbarios y de Museos de Historia Natural, Jardines Botánicos, Zoológicos, etc... Toda esta información presenta una gran base de conocimiento para la biodiversidad. Actualmente, existen más de 30 nodos nacionales proporcionando información a la red.

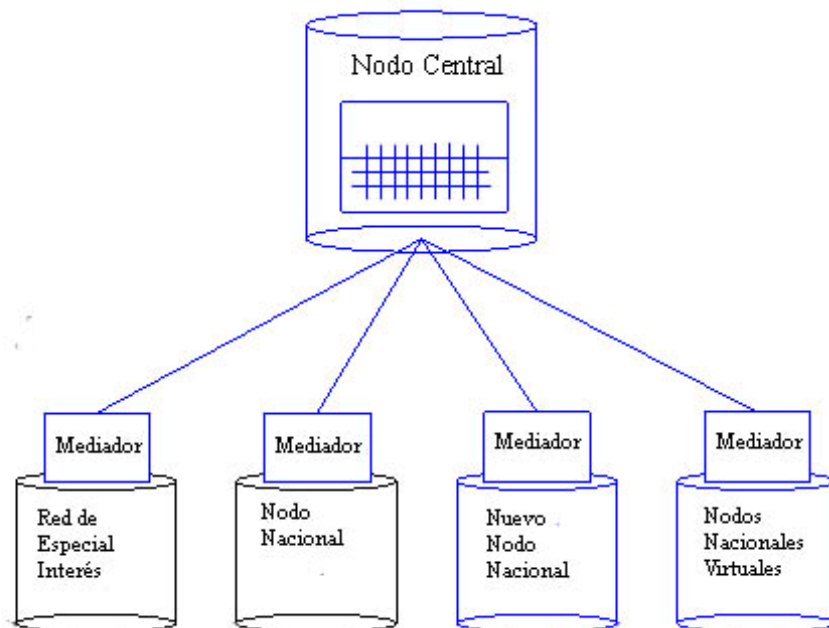


Figura 4. Flujo de Datos en BioCASE.

Como podemos comprobar, al igual que GBIF, BioCASE hace uso de sistemas **mediadores** (*wrappers*) (ver Figura 4) para poder compartir la información existente en las distintas fuentes. Proporciona un formato común al usuario y aplicaciones.

El proyecto BioCASE ha sido el precursor del uso del estándar *ABCD Schema* como esquema común de intercambio de datos de especímenes y observaciones.

Como consecuencia de la creación de proyectos como GBIF y BioCASE, se ha necesitado la realización de estándares que reflejen la información y la

estructura necesaria para intercambiar los datos a nivel de especímenes y observaciones.

### **3.2. Esquemas de intercambio de especímenes u observaciones**

Establecer la estructura de los datos a intercambiar es un paso importante que permite identificar los datos que son útiles compartir. Actualmente, la tecnología más usada para la generación de documentos de intercambio es *XML* (*eXtensible Markup Language*) y el lenguaje *XML Schema* para la definición de las estructuras del documento de intercambio (Balmin and Papakonstantinou, 2004) (ver Apéndice B).

Los esquemas que a continuación se detallan, reflejan la necesidad del intercambio de información entre instituciones y la integración de los datos sobre las colecciones biológicas. Debido a la existencia de distintos proyectos de unificación de información surgen distintos estándares: *Darwin Core* y *ABCD Schema*.

#### **3.2.1. Darwin Core (TDWG, 2005)**

El esquema *Darwin Core* nos indica los campos o atributos que definen el dominio de un espécimen u observación. En el esquema se declaran dos grupos de atributos (obligatorios y opcionales) que formaran el documento sobre el espécimen.

Podemos obtener el esquema del protocolo en la dirección: <http://digir.sourceforge.net/schema/conceptual/darwin/core/2.0/darwincoreWithDiGIRv1.3.xsd>

En la Tabla 2 se muestran la lista de atributos que definen el estándar.

Nombre	Descripción
<i>GlobalUniqueIdentifier*</i>	Valor que identifica de forma unívoca al espécimen dentro de la fuente de datos del proveedor. Tiene la forma: "URN:catalog:[ <i>InstitutionCode</i> ]: [ <i>CollectionCode</i> ]: [ <i>CatalogNumber</i> ]" Ejemplo: "URN:catalog:FMNH:Mammal:145732"
<i>DateLastModified*</i>	La última fecha de modificación del espécimen. Valores en ISO 8601 fecha y hora en UTC(GMT)
<i>BasisOfRecord*</i>	Descripción del registro indicando si existe físicamente o es observación. (P.e. <i>preserved specimen, observation, ...</i> )
<i>InstitutionCode*</i>	Código o acrónimo de la institución que administra la colección.
<i>CollectionCode*</i>	Código de la colección.
<i>CatalogNumber*</i>	Número del espécimen en el catálogo. Debe ser clave primaria.
<i>ScientificName*</i>	Nombre científico del espécimen hasta el nivel más bajo al que el espécimen puede ser descrito.
<i>HigherTaxon</i>	Nombre del <i>taxon</i> más alto.
<i>Kingdom</i>	Reino en el que el espécimen es clasificado.
<i>Phylum</i>	Nombre del filo o división en la que está clasificado.
<i>Class</i>	Nombre de la clase del espécimen.
<i>Order</i>	Nombre del orden.
<i>Family</i>	Nombre de la familia
<i>Genus</i>	Género del espécimen
<i>SpecificEpithet</i>	El epíteto específico del nombre científico.
<i>InfraspecificRank</i>	Rango de la infraespecie (subsp., var., forma.)
<i>InfraspecificEpithet</i>	El infraepíteto específico del nombre científico.
<i>ScientificNameAuthor</i>	El/los nombre/s de lo/s autor/es del nombre científico.
<i>IdentificationQualifier</i>	Indica alguna duda a la hora de determinar el espécimen.
<i>HigherGeography</i>	Rasgos genéricos de la geografía.
<i>Continent</i>	Continente
<i>WaterBody</i>	Nombre de la masa de agua en el que fue localizado el organismo.
<i>IslandGroup</i>	Nombre del grupo de islas.
<i>Island</i>	Nombre de la isla.
<i>Country</i>	País.
<i>StateProvince</i>	Estado.
<i>County</i>	Provincia.
<i>Locality</i>	Descripción de la localidad en la que fue encontrado.
<i>MinimumElevationInMeters</i>	Altitud mínima de la localidad de recogida del organismo.
<i>MaximumElevationInMeters</i>	Altitud máxima.
<i>MinimumDepthInMeters</i>	Profundidad mínima.
<i>MaximumDepthInMeters</i>	Profundidad máxima.
<i>DecimalLatitude</i>	Latitud de la localización del organismo.
<i>DecimalLongitude</i>	Longitud de la longitud del organismo.
<i>GeodeticDatum</i>	Dato del geodésico para la latitud/longitud indicada. Si se desconoce usar la etiqueta "nor recorded".
<i>CoordinateUncertaintyInmeters</i>	Imprecisión en metros de las coordenadas
<i>YearCollected</i>	Año de la recolección, 4 dígitos.
<i>MonthCollected</i>	Mes, 2 dígitos.
<i>DayCollected</i>	Día de la recolección. (1- 31)
<i>TimeCollected</i>	Hora de la recolección. Expresado en decimal donde una unidad es una hora.
<i>JulianDay</i>	El número del día en el año.

<i>Collector</i>	Nombre/s de el/los colector/es
<i>Sex</i>	Sexo del organismo. Valores: masculino, femenino, hermafrodita, andrógino, no grabado, indeterminado, transicional.
<i>LifeStage</i>	La edad del organismo (“ <i>juvenile</i> ”, “ <i>adult</i> ”, “ <i>eff</i> ”, “ <i>nymph</i> ”)
<i>ImageURL</i>	Referencia a una imagen del organismo.
<i>RelatedInformation</i>	Información adicional.

Tabla 2. Atributos de Darwin Core (Unidad de Coordinación de GBIF en España, 2006b). (\*: Denota el carácter de obligatorio)

El conjunto de datos que se pueden intercambiar nos indica los datos por los cuales podremos consultar al sistema. Como podemos observar por la gran cantidad de atributos descritos, disponer de toda la información es bastante complejo. No siempre estarán disponibles todos los datos que se reflejan en este estándar, más aún cuando algunos de ellos son incompatibles entre sí, como por ejemplo *minimumElevationInMeters* y *minimumDepthInMeters*.

### 3.2.2. ABCD Schema (Taxonomic Databases Working Group, 2006)

Se trata de un esquema alternativo, desarrollado en el marco del proyecto europeo BioCASE.

Este esquema alberga más de 2000 conceptos/campos y permite incluir información de los proveedores, metadatos sobre los datos, así como el conjunto de datos a transferir. Está estructurado en módulos, lo que posibilita su ampliación mediante nuevas extensiones.

A continuación se hace un extracto de la información que permite intercambiar.

Nombre	Descripción
<i>UnitGUID</i>	Valor que identifica de forma unívoca al espécimen dentro de la fuente de datos del proveedor. Tiene la forma: “URN:SourceInstitutionCode:[SourceID]: [UnitID]” Ejemplo: “URN:FMNH:Mammal:145732”
<i>SourceInstitutionID*</i>	Código o acrónimo de la institución que administra la colección.
<i>SourceID*</i>	Código de la colección.
<i>UnitID*</i>	Identificador único en la colección.
<i>UnitIDNumeric</i>	Identificador numérico que facilite la ordenación.

<i>LastEditor</i>	Nombre de la persona que modificó por última vez el registro.
<i>DateLastEdited</i>	Fecha de creación o última modificación
<i>Owner</i>	Propietario de la muestra. Concepto independiente del propietario físico de la muestra. Elemento complejo formado por los datos necesarios para el contacto con la persona.
<i>IPRStatements</i>	Especifica las condiciones de uso de la información, licencias, normas para citar la información, ...
<i>UnitContentContacts</i>	Información de contacto para preguntas y aclaraciones.
<i>SourceReference</i>	Referencias que apoyen la información (Título, URI del documento, ...).
<i>UnitReferences</i>	Conjunto de referencias bibliográficas que referencia a la unidad
<i>Identifications</i>	Las identificaciones realizadas a la unidad. Todos los nombres por los que ha pasado la muestra, fechas, referencias utilizadas para la identificación, ...
<i>RecordBasis</i>	Descripción del registro indicando si existe físicamente o es observación. (P.e. <i>preserved specimen, observation, ...</i> )
<i>KindOfUnit</i>	Partes del organismo o clase del material representado en la unidad.
<i>SpecimenUnit /ObservationUnit</i>	Indica si la información es referente a un espécimen o a una observación.
<i>CultureCollectionUnit</i>	Datos específicos para microbiología y disciplinas similares.
<i>MycologicalUnit</i>	Apartado para micología, hongos y similares.
<i>HerbariumUnit</i>	Información específica para especímenes de herbario.
<i>BotanicalGardenUnit</i>	Datos para especímenes hortícola.
<i>PlantGeneticResourcesUnit</i>	Información genética.
<i>ZoologicalUnit</i>	Específico para unidades zoológicas.
<i>PalaeontologicalUnit</i>	Datos sobre paleontología.
<i>MultimediaObjects</i>	Información multimedia asociada a la muestra.
<i>Associations</i>	Relaciones de la unidad con otras unidades, tipos de asociaciones.
<i>Assemblages</i>	Contenedor que agrupa la unidad a otras unidades.
<i>NamedCollectionsOrSurveys</i>	La unidad podría pertenecer a más de una colección.
<i>Gathering</i>	Información sobre la recolección de la unidad. Fecha, nombres de los recolectores u observadores, localidad, detalle del área de recolección, información geoespacial.
<i>CollectorsFieldNumber</i>	Número asignado por el recolector u observador.
<i>MeasurementsOrFacts</i>	Información para cuantificar y categorizar las medidas usadas.
<i>Sex</i>	Sexo de la unidad
<i>Age</i>	Edad del organismo al recolectarlo u observarlo.
<i>Sequences</i>	Contenedor para metadatos referentes a la secuenciación de la muestra.
<i>Notes</i>	Notas fuera de los campos anteriores.
<i>RecordURI</i>	Dirección Web donde ampliar la información de la muestra.
<i>EAnnotations</i>	Anotaciones en formato electrónico tomadas directamente de la muestra.
<i>UnitExtension</i>	Comodín para poder incluir más información que no esté reflejada en el esquema.
<i>Elementos compuestos.</i>	Elementos Obligatorios del esquema.

Tabla 3. Esquema de intercambio de información ABCD Schema

Como se puede observar por la Tabla 3 las posibilidades son mayores a las que oferta el esquema *Darwin Core*. No obstante, esta complejidad ha jugado en contra del esquema puesto que es más complejo adaptar el esquema a las bases de datos de las colecciones.

Se han incluido algunas mejoras muy importantes que permiten intercambiar información multimedia e información genética. Además, se realiza un tratamiento más extenso de la historia de la muestra, permitiendo incluir todos los estados nomenclaturales por los que ha pasado.

Toda la información generada por estos esquemas tiene que ser accesible a proveedores externos permitiendo así su intercambio. Para ello, es necesaria la creación de proveedores de datos que incorporen y gestionen los estándares anteriormente descritos. Es decir, la comunicación entre los proveedores de datos se tiene que realizar con un esquema común de comunicación que permita la interoperabilidad de los sistemas.

### **3.3. Protocolos de Comunicación**

La existencia de distintos proyectos e iniciativas para la estructuración de los datos a intercambiar, hace que existan distintos protocolos de comunicación y por ello distintos proveedores de datos (mediadores). A continuación se detallan las características de los tres principales protocolos existentes.

La iniciativa GBIF incorpora su estándar de comunicación mediante el proyecto *DiGIR*, mientras que *BioCASE* promovió su propio protocolo de comunicación, *BioCASE Access Protocol*. Por tanto, la existencia de distintos protocolos de comunicación llevó a la existencia de distintos tipos de proveedores complicando la integración de las fuentes de datos.

Actualmente, se está trabajando en la segunda generación de los protocolos de comunicación: TAPIR (*TDWG Access Protocol Information Retrieval*) el protocolo de acceso para la recuperación de información del TDWG. Éste pretende unificar los protocolos de comunicación actuales, *DiGIR* y *BioCASE Access Protocol*, promovidos por GBIF y BioCASE, respectivamente.

A continuación se describe el estándar de comunicación DiGIR, como iniciativa de GBIF. Después el protocolo BioCASE y finalmente la última

iniciativa del Grupo Internacional de Bases de Datos Taxonómicas para unificar los protocolos de comunicación existentes, TAPIR.

### 3.3.1. DiGIR

El proyecto DiGIR (DiGIR, 2006) (*Distributed Generic Information Retrieval*) obtiene como resultado un esquema de comunicación denominado DiGIR y un proveedor de datos denominado también DiGIR.

El proveedor de datos DiGIR, creado a partir del esquema de comunicación, permite el intercambio de datos mediante una arquitectura distribuida.

DiGIR intercambia datos estructurados desde múltiples y heterogéneas bases de datos a través de Internet. Es genérico (Döring, Giovanni, 2004) porque implementa un protocolo de comunicación independiente de los datos a intercambiar y de las fuentes de donde se obtienen los datos.

Para el desarrollo de DiGIR se tuvieron en cuenta los siguientes objetivos:

1. Usar protocolos libres y estándar, tales como http y XML.
2. Separar los protocolos de comunicación de los protocolos de intercambio de datos.
3. Construir un proveedor de datos (*data provider*) que cumpliera el estándar de comunicación, que permita el intercambio de datos y que tenga una instalación fácil.
4. Desarrollar un sistema de código abierto bajo la filosofía de GPL (*GNU General Public Licensing*).

Desde el proyecto DiGIR se han generado las herramientas necesarias para hacer que fluyan todos los datos y se pueda consultar la información de los

distintos proveedores. Mediante los proveedores de datos, se intercambian los datos haciendo que estos estén accesibles a los usuarios.

A continuación se muestra la arquitectura propuesta por el proyecto DiGIR y las herramientas generadas que permiten desde intercambiar la información hasta consultarla mediante una interfaz Web común.

### 3.3.1.1. *Arquitectura*

La arquitectura propuesta por DIGIR se compone de tres partes (ver Figura 5):

1. El **Proveedor**: fuente de información.
2. El **Registro de Proveedores**: directorio de proveedores, parecido a las páginas amarillas.
3. El **Portal**: recibe las peticiones de los usuarios.

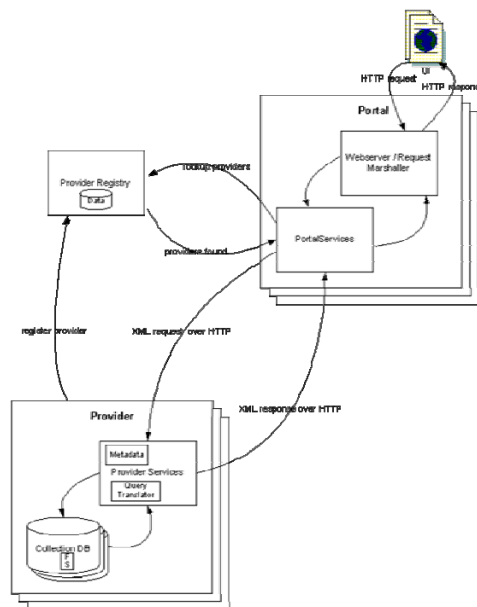


Figura 5. Diagrama de actuación de DIGIR. (DiGIR, 2006)



El principal objetivo del proyecto es la estandarización en el acceso a la información. La composición de los tres elementos hace que los datos puedan fluir entre las distintas fuentes de una forma distribuida.

Veamos cada una de las partes de las que se compone el sistema.

### ***Proveedor de datos***

El proveedor recibe las peticiones provenientes del Portal. El Proveedor devuelve los datos almacenados en las bases de datos que satisfacen la petición realizada por el portal.

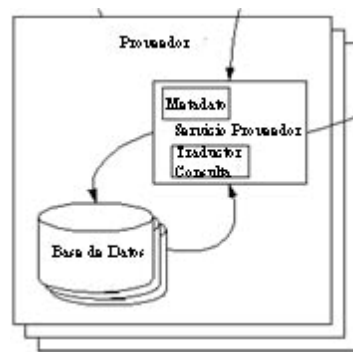


Figura 6. Proveedor de datos DIGIR (DiGIR, 2006)

Los resultados son enviados al solicitante, permitiendo al usuario obtener los metadatos que describen la fuente de datos (forma de acceder al sistema los gestores de la información,...)

### ***Servicio de Registro***

Provee las páginas amarillas de los proveedores existentes. Este componente mantiene la información de cada uno de los proveedores mediante un conjunto de metadatos. Estos registros pueden ser públicos o privados.



Figura 7. Diagrama del registro de proveedores (DiGIR, 2006)

### ***Portal de acceso***

Este es el punto de acceso para los usuarios y/o aplicaciones. El portal remitirá las peticiones a los correspondientes proveedores de forma totalmente transparente al usuario presentando la información en un formato común.

El portal necesita conocer los proveedores existentes. Para ello, consulta al servicio de registro para obtener los proveedores necesarios. Mediante el uso de metadatos el portal determina los proveedores que debe de consultar para resolver una petición realizada por el usuario o aplicación.

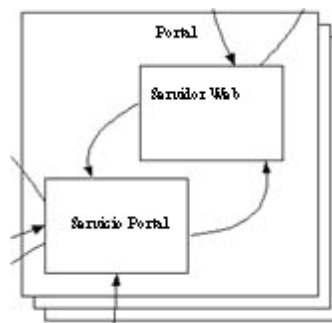


Figura 8. Diagrama del portal de DiGIR (DiGIR, 2006)

El portal envía una petición, en un formato común, a uno o muchos proveedores, generando para el usuario una respuesta unificada.

El conjunto de Proveedores de datos, servicio de registro y el Portal forman las herramientas necesarias para unir las distintas bases de datos de los

proveedores y proporcionar al usuario el acceso unificado a las distintas bases de datos. Para un correcto funcionamiento entre los distintos componentes necesitamos un protocolo de comunicación que permita el entendimiento entre las distintas partes.

### **3.3.1.2. Protocolo de acceso a proveedores DiGIR**

Define los formatos de los mensajes de petición y respuesta para la comunicación entre proveedores, portal, registro y las aplicaciones. El esquema del protocolo está disponible en:

<http://digir.sourceforge.net/schema/protocol/2003/1.0/digir.xsd>

A continuación se muestra un extracto del protocolo. En el Apéndice B se puede consultar los conceptos sobre XML y XML Schema, así como la notación utilizada para representar los XML Schemas de las siguientes figuras.

Para todos los mensajes existe una cabecera común (*header*) que indica la fuente y el destino de los mensajes (ver Figura 9).

Mediante el campo denominado *Type* se indica el tipo de operación que está realizando. Por ejemplo, búsqueda (*search*), metadatos (*metadata*), inventariado (*inventory*), estado (*status*).

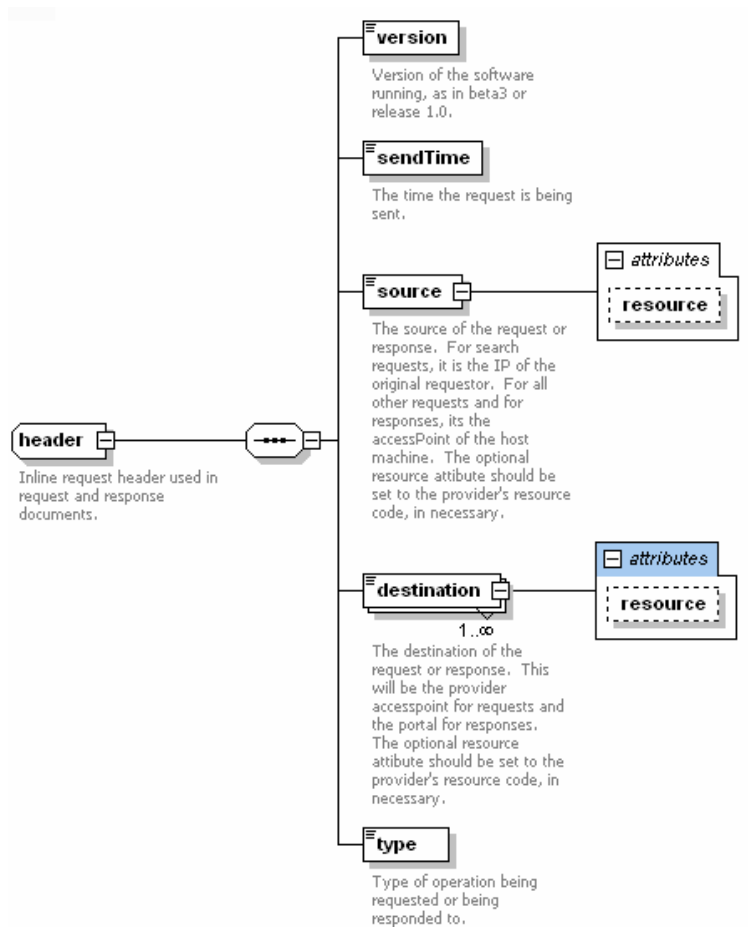


Figura 9. Componentes de una cabecera de un mensaje en el protocolo DiGIR

Veamos las posibilidades de petición de información (ver Figura 10). Independientemente del campo *type*, indicado en la cabecera, se pueden especificar las condiciones necesarias para la búsqueda de información o para realizar una numeración de algún concepto del esquema de intercambio.

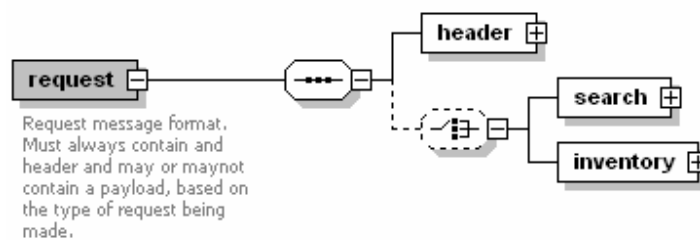


Figura 10. Sentencia Request de DiGIR

Se establecen tres tipos de peticiones posibles: las búsquedas (*search*), el inventario de un concepto dado (*inventory*) y los metadatos del proveedor.

Cuando se requiere realizar una búsqueda en el proveedor incluiremos los campos del esquema de intercambio por los que se van a buscar. El concepto (*searchOperation*) (ver Figura 11) especifica las condiciones de búsqueda (*filter*) así como la estructura de los resultados (*records*) que se deben de mostrar o devolver en la respuesta.

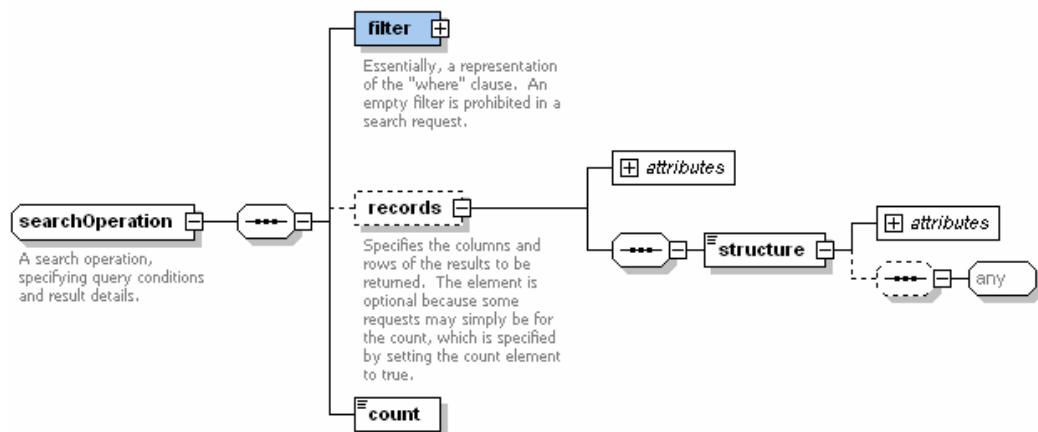


Figura 11. Operación de búsqueda de información. Protocolo DiGIR

En el mensaje de búsqueda de información establecemos: parámetros de filtrado (*Filter*), la estructura de respuesta (*records*), los datos que deseamos recibir tras la consulta y el campo que indica si deseamos obtener el total de ocurrencias que necesitamos.

Mediante la operación de inventariado (*inventory*) permite contar las ocurrencias existentes en el proveedor de consulta. Opcionalmente se pueden establecer condiciones (*filter*) en la consulta (ver Figura 12).

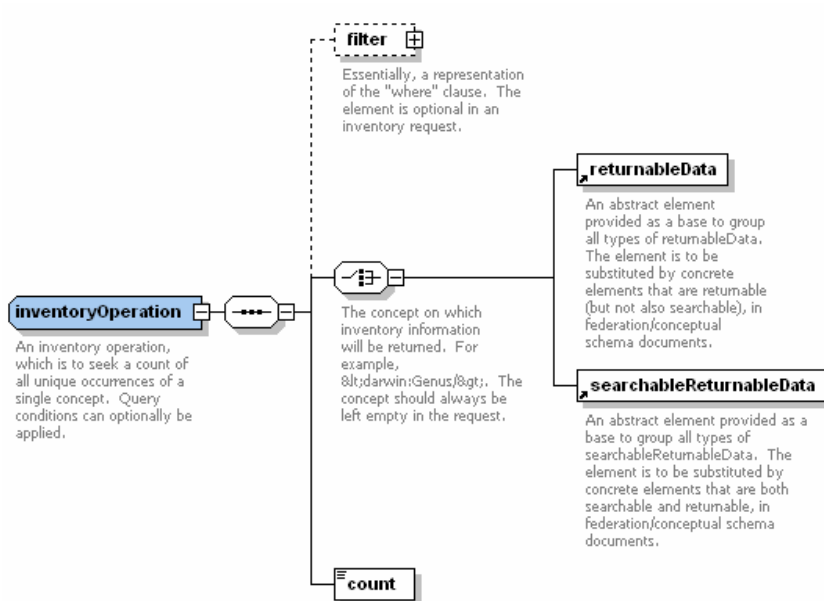


Figura 12. Estructura petición de inventariado. Protocolo DiGIR

Mediante el elemento del tipo *count* podemos obtener, además de la información requerida, el número de ocurrencias existentes en la base del proveedor.

En el proveedor de datos se almacena información que describe la tipología de datos que se intercambian, datos sobre el propio proveedor, datos sobre las personas de contacto y que administran la colección, etc. Esta información se denomina **metadatos**.

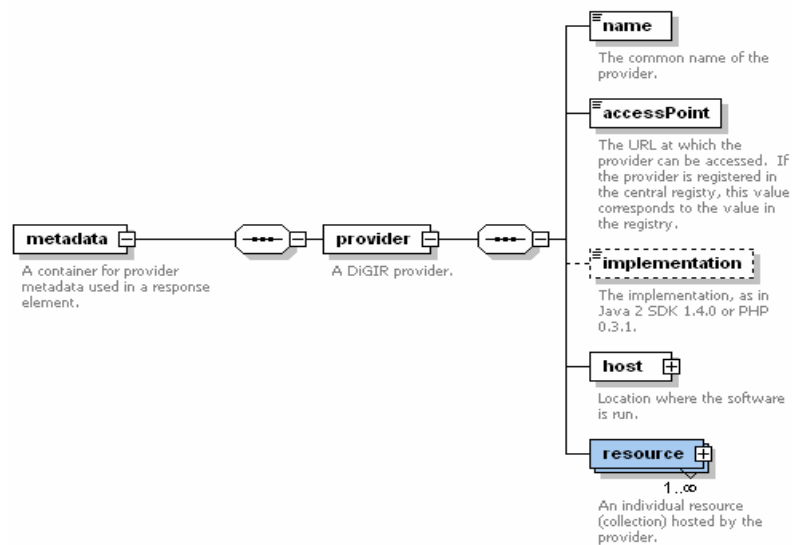


Figura 13. Estructura de los metadatos. Protocolo DiGIR

El tipo complejo *metadata* almacena información referente al proveedor. Para cada proveedor es importante almacenar la dirección de su punto de acceso a través del cual se puede interactuar con él y un conjunto de campos (*resource*) que hacen referencia al conjunto de fuentes disponibles en el proveedor.

Por último, veamos la estructura del mensaje de respuesta a una petición de búsqueda o de inventariado (*response*).

En la respuesta realizada por el proveedor de datos se incluye una cabecera (*header*), el contenido (*content*) y un conjunto de elementos de diagnóstico (*diagnostics*) que nos indican, principalmente, el resultado satisfactorio o no de la operación.

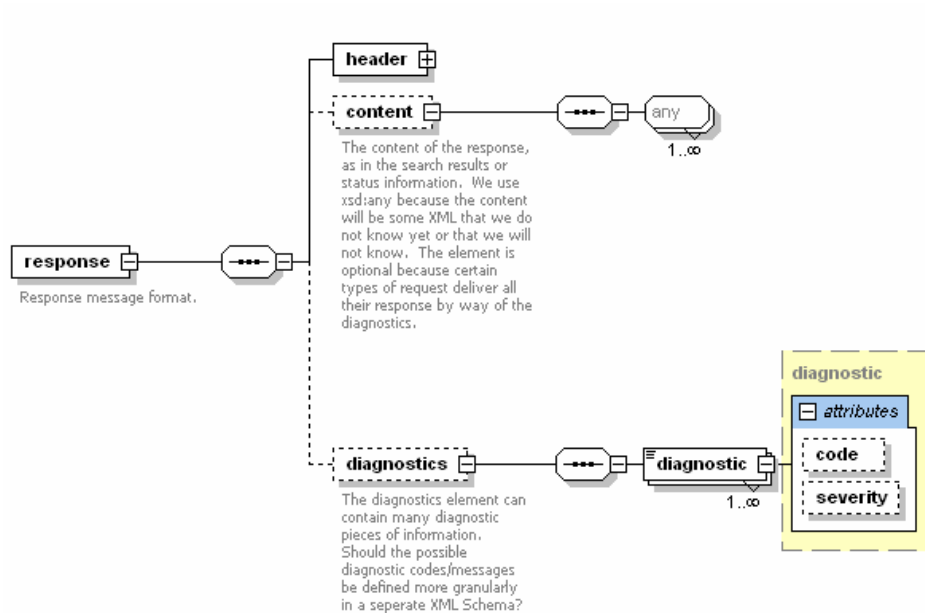


Figura 14. Respuesta a la petición de un mensaje. Protocolo DiGIR

Para cada una de las fuentes consultadas se incluye información de las personas de contacto, descripción de la fuente, palabras claves, cómo citar la información, identificador de los registros, restricciones de uso, etc.

A continuación se muestran dos ejemplos de código *XML* que representan la petición y respuesta de un proveedor de datos, según el protocolo.

### 1. Ejemplo de petición de los metadatos para un proveedor:

```
<?xml version="1.0" encoding="utf-8" ?>
<request xmlns="http://digir.net/schema/protocol/2003/1.0">
  <header>
    <version>1.0.0</version>
    <sendTime>2006-01-20T06:51:45-03:00</sendTime>
    <source>127.0.0.1</source>
    <destination>http://proveedor.org/provider/DiGIR.php</destination>
    <type>metadata</type>
  </header>
</request>
```

Figura 15. Código XML que implementa la petición de Metadatos DiGIR



## 2. Ejemplo de la respuesta del proveedor a la petición realizada:

```

<?xml version="1.0" encoding="utf-8" ?>
<response xmlns="http://digir.net/schema/protocol/2003/1.0">
<header>
  <version>$Revision: 1.14 $</version>
  <sendTime>2006-01-20T06:56:35-03:00</sendTime>
  <source>http://proveedor.org:80/provider/DiGIR.php</source>
  <destination>127.0.0.1</destination>
  <type>metadata</type>
</header>
<content> <metadata>
<provider>
  <name>Nombre Proveedor</name>
<accessPoint>http://proveedor.org:80/provider/DiGIR.php</accessPoint>
<implementation>$Revision: 1.14 $</implementation>
<host>  <name>Nombre Host</name> <code>Codigo</code>
<relatedInformation>http://www.coleccion.org/</relatedInformation>
<contact type="technical">
  <name>Nombre persona</name> <title>Categoria</title>
  <emailAddress>nombre@institucion.org</emailAddress>
  <phone>número teléfono</phone>
</contact> ...
</host>
  <resource><name>Nombre Fuente</name><code>Código Fuente</code>
<relatedInformation>http://fuente.org/</relatedInformation>
<contact type="technical">
  <name>Nombre persona</name>
  <title>Titulo</title>
  <emailAddress>nombre@fuente.com</emailAddress>
  <phone>numero telefono</phone>
</contact>
  <abstract>Descripción fuente</abstract>
  <keywords />
  <citation />
  <useRestrictions />
  <conceptualSchema
schemaLocation="http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd">http://digir.
net/schema/conceptual/darwin/2003/1.0/</conceptualSchema>
  <recordIdentifier>Código_coleccion</recordIdentifier>
  <recordBasis>specimen</recordBasis>
  <numberOfRecords>7887</numberOfRecords>
  <dateLastUpdated>2004-05-06T13:56:00</dateLastUpdated>
  <minQueryTermLength>3</minQueryTermLength>
  <maxSearchResponseRecords>1000000</maxSearchResponseRecords>
  <maxInventoryResponseRecords>1000000</maxInventoryResponseRecords>
  </resource> </provider> </metadata> </content>
<diagnostics>
  <diagnostic code="STATUS_INTERVAL" severity="info">3600</diagnostic>
  <diagnostic code="STATUS_DATA" severity="info">5,0,0</diagnostic>
</diagnostics>
</response>

```

Figura 16. Código XML de respuesta a petición de metadatos en DiGIR

### 3.3.2. Proyecto BioCASE

El proyecto BioCASE, anteriormente comentado, desarrolló su propio protocolo de comunicación entre proveedores de datos para el intercambio de información sobre especímenes, en el estándar *ABCD Schema*. Este estándar presenta modificaciones sobre el protocolo de comunicación *DiGIR*. A continuación se mostrarán las principales diferencias existentes entre ambos.

#### 3.3.2.1. Arquitectura

El proyecto BioCASE, al igual que el proyecto GBIF, creó las herramientas necesarias para que la información pudiera ser compartida por los proveedores y estuviera accesible para los investigadores. Para poder conseguir sus objetivos, BioCASE tuvo que desarrollar el equivalente al proyecto DiGIR.

La Figura 17 describe la arquitectura del proveedor de datos que gestiona las distintas fuentes existentes proporcionando una estructura común (*ABCD Schema*) a la red de información del proyecto BioCASE.

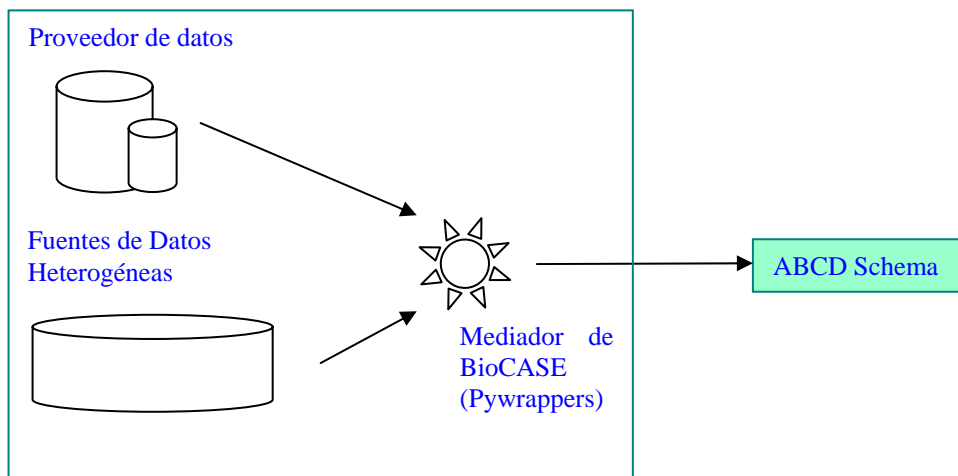


Figura 17. Arquitectura del proveedor de datos

### 3.3.2.2. Protocolo de acceso a proveedores BioCASE (BioCASE Access Protocol)

A continuación mostramos el protocolo de comunicación entre los proveedores de datos. Éste es un protocolo con más posibilidades que el protocolo generado por DiGIR. No obstante, la estructura del protocolo es muy similar.

En la Figura 18 podemos observar como existen conceptos parecidos:

- La cabecera (*header*).
- Un apartado para búsquedas (*search*).
- Una operación de inventariado (*scan*).

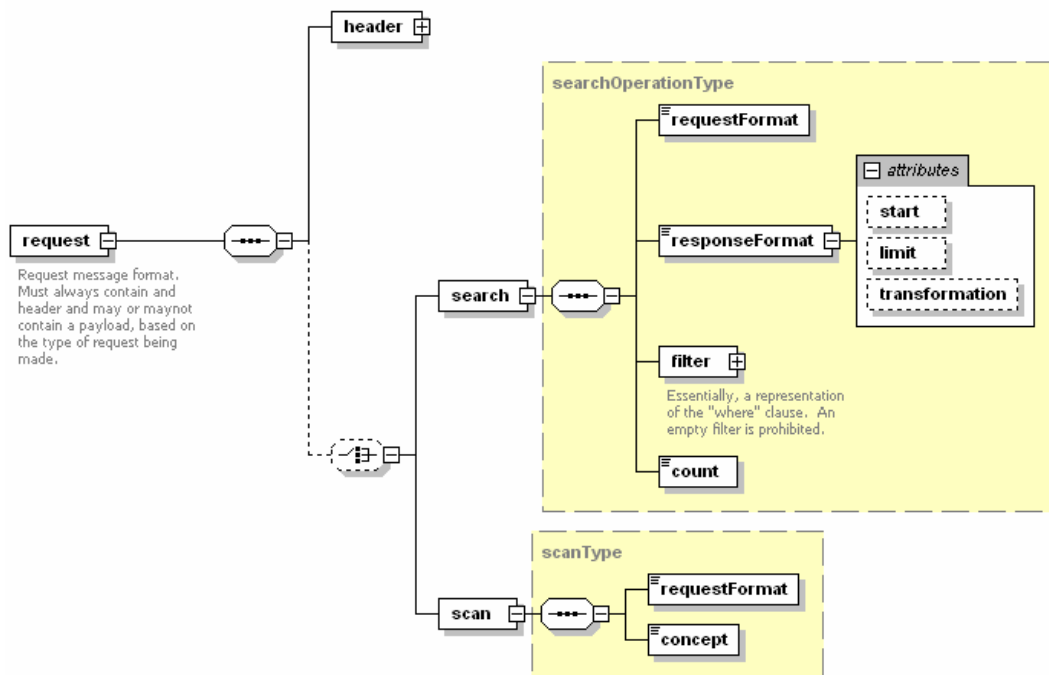


Figura 18. Petición de información a un proveedor BioCASE.

Incluye modificaciones que permiten especificar directamente el formato de respuesta del mensaje, propiedades para limitar la cantidad de resultados devueltos por el proveedor, y generar la posibilidad de paginar los resultados solicitados al proveedor.

El protocolo incluye una nueva operación denominada "capacidades" (*Capabilities*) que permite consultar los esquemas soportados por el proveedor. Un esquema hace relación al tipo de información que es capaz de intercambiar (ver Figura 19).

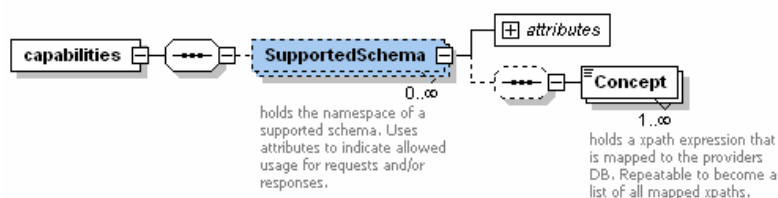


Figura 19. Operación *Capabilities* en el protocolo BioCASE

Aunque los objetivos de los protocolos de comunicación son los mismos, existen diferencias entre ellos que no los hacen compatibles. La existencia de dos estándares distintos para la comunicación con los proveedores y la existencia de distintos tipos de proveedores, complica el acceso unificado a la información.

### 3.3.3. TAPIR: Protocolo de Acceso para la Recuperación de Información

Surge durante la reunión del grupo internacional de bases de datos taxonómicas (TDWG) del 2004 (Taxonomic Databases Working Group, 2004). Se propone para unificar los estándares de comunicación entre proveedores de datos. Un protocolo que permita consultar los datos del proveedor en cualquier estándar de intercambio (por ejemplo, *ABCD Schema* y *Darwin Core*).

El uso de TAPIR aumenta la capacidad de comunicación puesto que permite obtener los datos en distintos formatos (ver Figura 20).

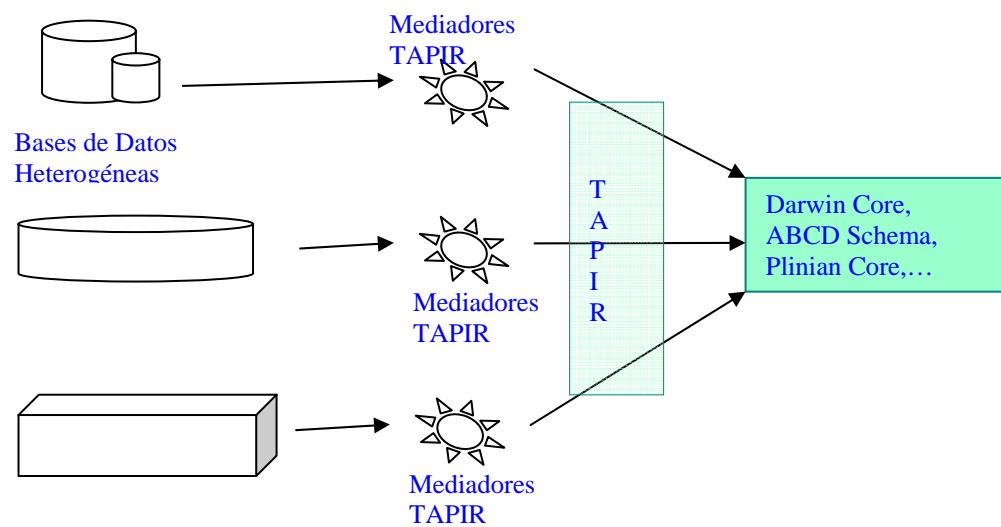


Figura 20. Diagrama del protocolo de comunicación TAPIR

Veamos a continuación las principales características del estándar y cómo resuelve los problemas de sus antecesores (Protocolo *DiGIR* y el Protocolo de Acceso BioCASE).

### 3.3.3.1. Protocolo de acceso a proveedores

El protocolo permite la transmisión de datos en cualquier formato de intercambio ya sea *Darwin Core* o *ABCD Schema*.

Las principales características de TAPIR (Döring, 2005) son:

1. Flexibilidad para crear la estructura de la respuesta.
2. Definición de distintos tipos de vistas (datos en Darwin Core, ABCD Schema, RSS, etc.).
3. Utilización de una capa intermedia entre los conceptos internos y la estructura de respuesta.

Mediante la utilización de vistas permite al cliente realizar preguntas en distintos esquemas, simplemente indicando la dirección de la definición de la vista.

La especificación del protocolo se puede obtener de:  
<http://ww2.BioCASE.org/svn/tapir/trunk/protocol/tapir.xsd>

Veamos a continuación las definiciones más importantes del protocolo:

- Petición de información a un proveedor (*requestType*) :

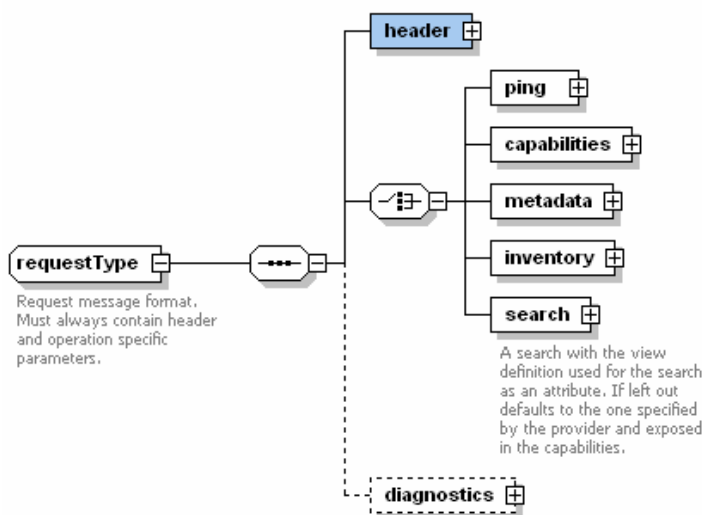


Figura 21. TAPIR: Estructura del mensaje de petición.

En la Figura 21 vemos que un mensaje de petición necesita de dos componentes fijos: una cabecera y una operación (*ping*, *capabilities*, *metadata*, *inventory*, *search*). Opcionalmente se puede incluir un conjunto de mensajes de diagnóstico.

Todos los tipos de peticiones son comunes, a excepción de la petición *capabilities* existente solamente en proveedores BioCASE y la petición *ping* recientemente incorporada. La operación *ping* se usa para comprobar si un proveedor está operativo.

- Petición de capacidades (“*capabilities*”) (Figura 22) el cliente recibe toda la información referente a las operaciones que es capaz de realizar, la información sobre las operaciones que permiten filtrar los resultados, los esquemas conceptuales conocidos, las variables de los proveedores y la

información de configuración de las distintas fuentes de datos existentes en el mismo.

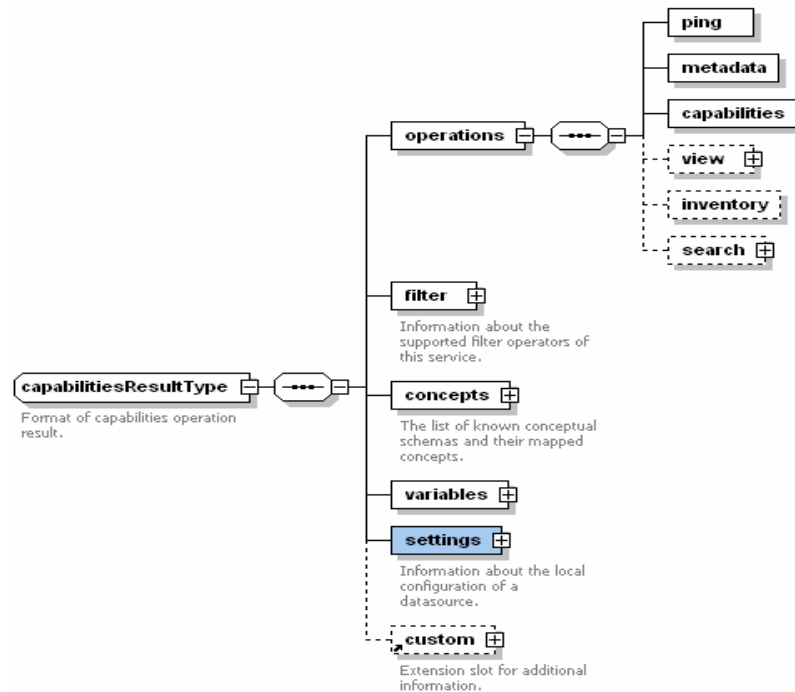


Figura 22. TAPIR: estructura del mensaje *capabilities*

- Petición de los metadatos del proveedor. El cliente obtiene los datos del proveedor tales como: personas de contacto, descripción del proveedor, identificadores del proveedor, restricciones en el uso de los datos, etc. Esta operación ya existía en los proveedores anteriores de *DiGIR* y *BioCASE*, pero se han incorporado más posibilidades.
- La operación de inventariado (*inventorTemplateType*) (ver Figura 23) permite obtener la información existente para un concepto en el proveedor de datos. A su vez se puede especificar un conjunto de condiciones que tienen que cumplir los datos de la respuesta.

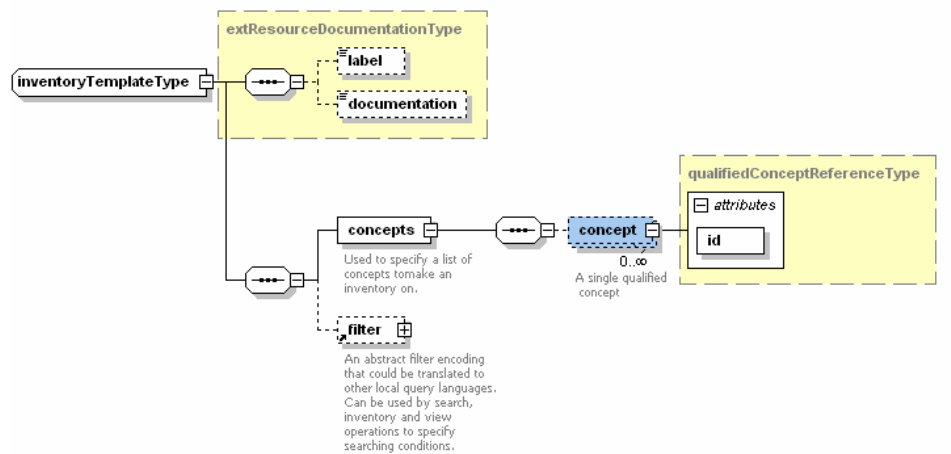


Figura 23. TAPIR: operación de petición de inventariado.

- Operación de búsqueda mediante **vistas** (*outputModel*). Permite cambiar la estructura de respuesta del mensaje. De esta forma, independientemente de cómo se encuentre almacenada la información e independientemente del esquema de intercambio, *TAPIR* es capaz de transformar el mensaje de respuesta.

El uso de vistas permite obtener información en cualquier formato. La vista se define estableciendo un mapeo entre los conceptos de la Base de Datos y la estructura de la vista. Por ejemplo, en formato *Keyhole* (*Google Earth*), RSS, etc.

En la Figura 24 podemos observar el formato que define una vista. Es necesario especificar la estructura de devolución de la vista (*structure*). Definir cuales son los elementos que forma la unidad básica, necesaria para poder establecer un raíz, y por último, indicar una equivalencia para cada uno de los componentes de la estructura con los conceptos que tiene definidos el proveedor.



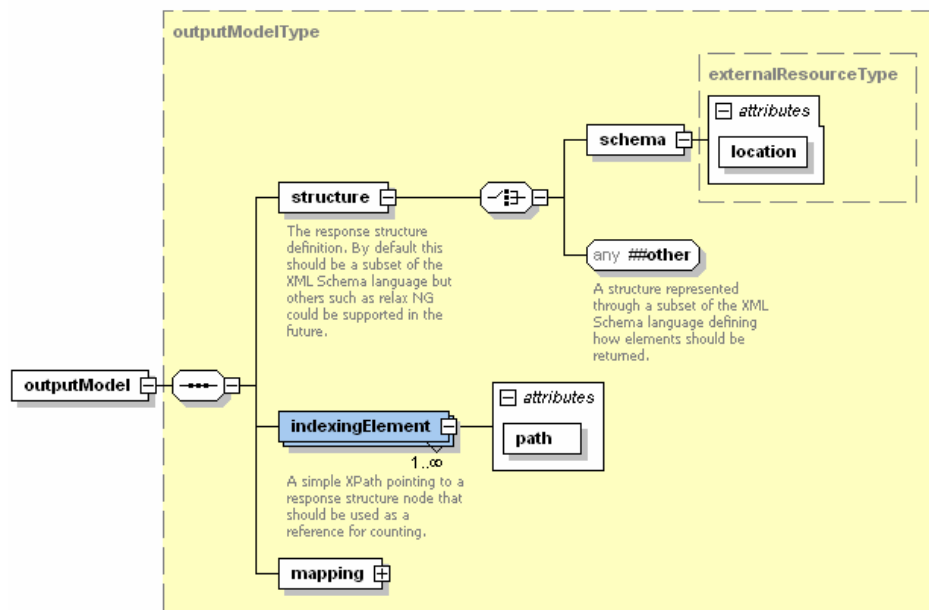


Figura 24. TAPIR: estructura de una vista

En el mensaje de petición de búsqueda tenemos que especificar la vista que deseamos obtener, pudiendo solicitar al proveedor un subconjunto de la vista (*partial*) y las condiciones que tienen que cumplir los datos. A su vez, podemos especificar el conjunto de conceptos por los que deseamos que se ordene.

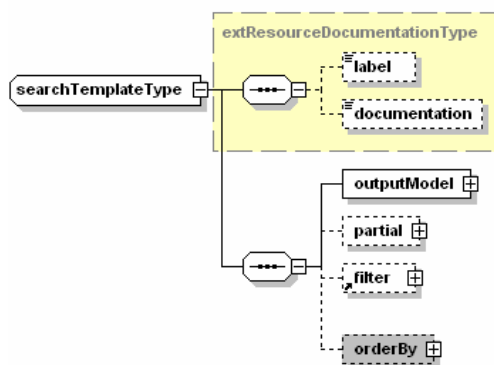


Figura 25. TAPIR: Petición de búsqueda ("search")

Una vez vistas las peticiones hemos de ver que estructura presentan las respuestas. El formato del mensaje de respuesta tiene componentes comunes a las todas peticiones. En él se indica un campo con la cabecera (*header*) y un conjunto

de mensajes de diagnóstico. En la respuesta de búsqueda se incluye información referente a la paginación de los resultados.

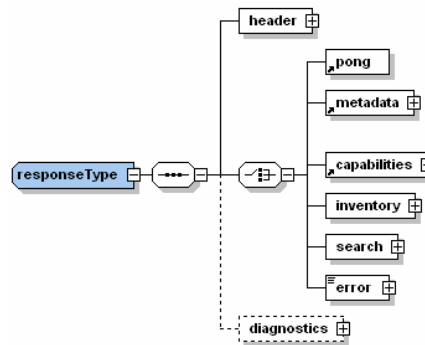


Figura 26. TAPIR: estructura del mensaje de respuesta

*TAPIR* se presenta como el lenguaje de comunicación que permite la unificación a cualquier estándar de intercambio. El uso de vistas y la ampliación de capacidades del lenguaje permiten el intercambio de los estándares actuales. Proponiendo su proveedor de datos como el proveedor capaz de unificar los distintos esquemas existentes (*ABCD Schema*, *Darwin Core*, etc.).

## Capítulo II. Sistemas *Software* existentes

En este capítulo, se realiza un análisis de los sistemas *software* que permiten la informatización de colecciones biológicas. Concretamente, se centrará el estudio en sistemas para colecciones botánicas. La razón principal por la que se han considerado este tipo de colecciones, es que se trata de un campo en el que hemos trabajado prioritariamente. No obstante, todas las aportaciones que se hacen en la presente tesis son extensibles a cualquier colección.

Para poder evaluar los distintos sistemas existentes hay que establecer un marco común de características necesarias para la gestión de las colecciones botánicas. Estas características forman los requerimientos funcionales<sup>1</sup> del *software*.

El tratamiento y gestión de las colecciones es un problema complejo para el cual es necesaria la informatización de los especímenes. Tras el análisis previo de las distintas áreas de aplicación, el tratamiento nomenclatural y los estándares internacionales, se pueden establecer el siguiente conjunto de metas a cubrir por cualquier sistema *software*:

1. **Control de usuarios:** el centro debe de identificar los distintos tipos de usuarios, y establecer las operaciones que cada usuario puede realizar dentro del sistema. Por cada perfil, se establecen un conjunto de privilegios que determinan el conjunto de operaciones permitidas. No todos los usuarios tienen la misma especialización ni capacidad para el acceso y modificación de ciertas partes del sistema.

---

<sup>1</sup> Los requerimientos funcionales nos indican qué metas debe alcanzar el sistema.

2. **Gestión de especímenes:** el sistema debe permitir realizar altas, modificaciones y bajas de nuevos especímenes. Toda la información sobre los especímenes tiene que estar perfectamente enlazada entre sí.
3. **Generación de Etiquetas:** el espécimen es almacenado en un recipiente (pliego, tubo, etc.) para ser conservado en el centro. Por ello, es necesario etiquetar el recipiente para que muestre información sobre el/los espécimen(es) que almacena.
4. **Consultas:** debe permitir el acceso a los especímenes informatizados de una forma rápida y eficiente. El usuario no tiene que conocer los lenguajes de consulta de Bases de Datos, como SQL, para realizar consultas en el sistema.
5. **Préstamos de material:** la conservación del material que forman las colecciones es muy importante para posteriores estudios de biodiversidad. Este hecho nos lleva a tener que generar procedimientos que nos permitan compartir el material. El centro necesita emitir y gestionar los préstamos de material a otras instituciones o investigadores.
6. **Información multimedia:** Este tipo de información es útil para el intercambio *on-line* de información. La información multimedia sirve, en ocasiones, como sustituto del préstamo de material.
7. **Referencias bibliográficas:** en el proceso de asignación del nombre científico al espécimen, generalmente, se hace referencia a la bibliografía consultada para la determinación del espécimen.
8. **Información taxonómica:** el sistema debe mantener un filtro mínimo de la información proporcionada por el usuario. El sistema tiene que mantener un registro de información taxonómica que permitan filtrar gran parte de los errores mecanográficos.

9. **Servicios Externos:** dado el carácter de centro de documentación, para los centros gestores de las colecciones es importante tener vías de comunicación que permitan consultar la información existente en el centro. Las consultas de personal externo para investigaciones son muy importantes y ocupan gran parte del tiempo del personal del centro.
10. **Exportación de información:** el intercambio de la información es necesario para el conocimiento global de la biodiversidad. Por ello, surgen todas las iniciativas anteriores, y se establecen los distintos protocolos y estándares de intercambio.
11. **Distribución de Especímenes:** una parte importante de la información asociada al espécimen es información geoespacial. La utilización de dispositivos de posicionamiento global (GPS) en las campañas de recolección permite, cada vez más, proporcionar las coordenadas espaciales de las muestras. La utilización de herramientas visuales que nos permitan observar la distribución de los especímenes recolectados ayuda al centro, entre otras cosas, a determinar zonas en las que no se ha recolectado material. Este tipo de información es útil para poder orientar las campañas de recolección.

Los requerimientos anteriormente expuestos muestran, sin entrar en detalle sobre la estructura de los datos, las necesidades de los centros de gestión.

Una vez realizado el análisis de los requerimientos funcionales del *software* de gestión de colecciones se evalúan los sistemas *software* existentes. Gran parte de los sistemas evaluados están recomendados por el TDWG para la informatización de colecciones (<http://www.bgbm.fu-berlin.de/TDWG/acc/Software.htm>).

Es importante realizar un análisis de la funcionalidad de los sistemas existentes para poder determinar la necesidad o no de la realización de *software* que permita gestionar las colecciones botánicas y la realización de estudios de biodiversidad.

Para cada uno de los sistemas analizados se describirá la forma de acometer cada uno de los requerimientos funcionales anteriormente descritos.

Una vez que se evalúen las características de los distintos sistemas se mostrará una tabla comparativa con las funciones que incorpora cada uno de los sistemas analizados.

## **1. El gestor de bases de datos de Biodiversidad: Biota 2**

Sistema desarrollado por Robert K. Colwell de la Universidad de Connecticut (Colwell, 2004) durante siete años. Gestiona información sobre especímenes, colecciones de datos para los ecólogos en 40 países y en 47 estados de Estados Unidos. La versión 2 de Biota ha modificado la interfaz e incorporado nuevas características de entrada de datos, mantenimiento y publicación Web.

### **1.1. Características**

Es un *software* comercial para Macintosh que puede ejecutarse también en Windows. Es un sistema de bases de datos relacional que usa el motor de bases de datos de cuarta dimensión (*4th dimension*<sup>2</sup>). Veamos la funcionalidad del sistema teniendo en cuenta los requerimientos funcionales expuestos anteriormente:

1. **Control de usuarios:** Establece cinco grupos de usuarios (Súper usuario, Administrador, Lector, Modificador y Exportador) al cual puede pertenecer un usuario. Con este control, el sistema no

---

<sup>2</sup> *4th dimension*: disponible en <http://www.4d.com>. Producto comercial que incorpora las utilidades necesarias y un marco de trabajo para el desarrollo de bases de datos y componentes Web de forma fácil e integrada.

gestiona las operaciones particulares que un usuario puede realizar dentro del sistema, sino que asigna a cada grupo un conjunto de operaciones por defecto.

2. **Gestión de especímenes:** permite almacenar información referente a distintas colecciones de especímenes, administrar los distintos colectores y las localidades en las que se ha recolectado el material.
3. **Generación de Etiquetas:** la impresión de etiquetas viene gestionada por un módulo separado que permite diseñar su propio formato de etiqueta. No usa código de barras para la codificación de la muestra, siendo éste el principal inconveniente observado en el módulo.
4. **Consultas:** establece un conjunto de consultas fijas y permite mediante un formulario diseñar consultas haciendo uso de despegables con los campos existentes en las distintas tablas de la bases de datos.
5. **Préstamos de material:** control de los préstamos realizados a otras instituciones y del tiempo transcurrido. Permite obtener listados del material prestado y realizar la devolución indicando uno a uno el material devuelto.
6. **Información multimedia:** Sólo permite la asociación de imágenes a especímenes.
7. **Referencias bibliográficas:** mantiene un registro mínimo de información bibliográfica. Las referencias se pueden asociar a una muestra.
8. **Información taxonómica:** mediante un módulo independiente, permite establecer la jerarquía de los distintos nombres científicos. El uso de más de 10 ventanas solicitando información al usuario

sobre el nombre hace que sea muy laborioso dar de alta información en el sistema.

9. **Servicios Externos:** el sistema incorpora la posibilidad de activar un servidor Web propio que permite consultar la información sobre las colecciones existentes en el mismo. El uso del servidor Web no filtra el tipo de información que se muestra a través de la red. No evita que información sensible pueda ser usada mal intencionadamente (por ejemplo, la georreferenciación de un espécimen).
10. **Exportación de información:** el sistema permite exportar información haciendo uso de ficheros planos. No realiza la exportación en ninguno de los estándares establecidos internacionalmente.

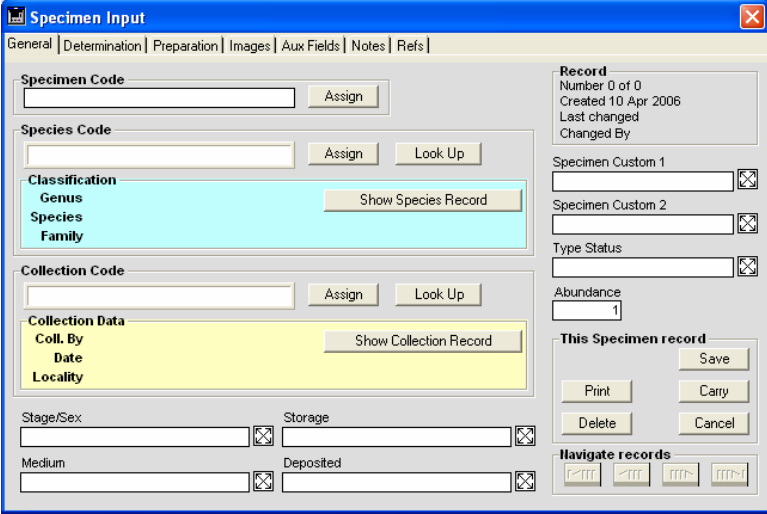


Figura 27. Captura de pantalla de Biota

## 1.2. Análisis de Características

Biota 2 es un sistema muy completo y con grandes posibilidades de gestión. La información está bien estructurada en módulos independiente pero



hace que la entrada de datos sea muy lenta y tediosa. Se pueden destacar las siguientes ventajas e inconvenientes:

- Ventajas:
  - Gran cantidad de módulos para incluir información sobre biodiversidad.
  - Módulo de búsqueda sencillo y accesible desde la Web mediante servidor propio.
  
- Desventajas:
  - Se necesita la adquisición de licencia de uso. Existen distintos tipos de licencias a distintos precios, dependiendo de la funcionalidad proporcionada por el sistema.
  - No es intuitivo y se necesita mucho tiempo para insertar un registro en el sistema.
  - La estructuración en grupos de usuarios no permite la personalización del conjunto de operaciones permitidas a cada usuario.
  - No posee de herramientas de copias de seguridad.
  - Las herramientas de exportación no comparten información en ninguno de los estándares existentes (*DiGIR, ABCD Schema*)

## **2. Specify**

Herramienta *software* especialmente diseñada para el tratamiento de las colecciones biológicas. Specify (Specify, 2005) ha sido financiado por la Fundación Nacional de Ciencia de los Estados Unidos. El objetivo del desarrollo

es proporcionar el soporte *software* necesario para las colecciones. La versión analizada está desarrollada en Microsoft Access y posee un sistema de generación de informes independiente llamado Neurona RAVE.

## 2.1. Características

Veamos cada una de los requerimientos establecidos al inicio del capítulo para analizar el sistema. Sólo se mostrarán aquellos puntos que incorpora el sistema:

1. **Gestión de especímenes:** permite almacenar información referente a distintas colecciones de especímenes (ver Figura 28), sus localidades, colectores e información sobre el evento de recolección.
2. **Consultas:** se establece un conjunto de consultas e informes generados mediante una herramienta gráfica.
3. **Préstamos de material:** control de los préstamos realizados a otras instituciones. Permite obtener listados del material prestado.
4. **Información taxonómica:** mediante el formulario de entrada de especímenes permite indicar información sobre su taxonomía.
5. **Exportación de información:** permite la exportación de información en el estándar de Darwin Core.

Figura 28. Formulario de entrada de datos sobre especímenes en Specify

## 2.2. Análisis de Características

Es un proyecto financiado y que consta de un equipo de trabajo que da soporte y mantenimiento a dicho sistema. Actualmente están desarrollando una nueva versión que trabaja con Microsoft SQL Server.

Como podemos observar este sistema omite gran parte de los requerimientos establecidos por las necesidades de los centros y de las colecciones. No obstante, permite la exportación de su información en un formato estandarizado.

Se pueden destacar las siguientes ventajas e inconvenientes:

- Ventajas:
  - El sistema se puede descargar previo registro en la página Web del proyecto.
  - Herramienta potente de generación de informes.
- Desventajas:

- No controla las operaciones de los usuarios.
- No gestiona información multimedia.
- No es un sistema completo que cumpla con todos los requerimientos establecidos por el análisis del problema.
- No incorpora la posibilidad de consultar de forma externa.

### 3. Biotica

El sistema de Información Biótica (Biótica., 2005), desarrollado por la CONABIO, ha sido diseñado especialmente para el tratamiento de datos nomenclaturales, geográficos, bibliográficos y de parámetros ecológicos. Pretende ayudar en la captura y actualización de los datos, permitiendo una informatización de, prácticamente, cualquier dato. Fue desarrollado en forma modular tanto en la estructura de la base de datos como en su sistema teniendo en cuenta la gran variedad de necesidades de la comunidad biológica

#### 3.1. Características

Biótica fue desarrollado en módulos, permitiendo la relación entre los mismos y posibilitando además la extensión del sistema sin necesidad de rediseñar todo el *software*.



Figura 29. Menú principal de Biótica

Biótica se divide en nueve módulos (ver Figura 29): Base de datos, Directorio, Nomenclatura, Conservación, Ecología, Georreferenciación, Bibliografía, Herramientas y Ayuda.

Permite la gestión de distintas bases de datos y los usuarios que pueden acceder a los datos. El gestor de base de datos utilizado es Microsoft Access, lo que limita las posibilidades del sistema.

Mediante el módulo de directorio es posible insertar la información referente a las instituciones y colecciones (nombre, siglas, dirección, etc.) así como grupos de taxónomos, grupos de colectores y autores de publicaciones.

Veamos su comportamiento con los requerimientos establecidos:

1. **Control de usuarios:** permite realizar un control de usuarios sin especificar las operaciones que puede realizar cada uno. El control de usuarios se establece solamente para reflejar qué usuario accede a los distintos módulos o realiza las modificaciones.
2. **Gestión de especímenes:** permite almacenar información referente a distintas colecciones de especímenes, la georreferenciación y las localidades.
3. **Generación de Etiquetas:** la impresión de etiquetas viene gestionada por el módulo de herramientas. Este módulo permite incorporar un código de barras en la etiqueta.
4. **Consultas:** establece un conjunto de consultas fijas y permite mediante un formulario diseñar cualquier otra consulta mediante el uso de despegables con los campos existentes en las distintas tablas que almacenan la información.
5. **Información multimedia:** permite asociar una imagen a los especímenes. Sólo permite el uso de imágenes.
6. **Referencias bibliográficas:** Dentro del módulo de conservación podemos introducir información relacionada con la recolección, observación e informe bibliográfico del ejemplar que enlazado con

el módulo de nomenclatura permite controlar los nombres científicos.

### **3.2. Análisis de Características**

Como podemos observar es un sistema bastante complejo y orientado sobre todo a la gestión de la información de las colecciones. Podemos enumerar las siguientes ventajas y desventajas.

- Ventajas:
  - sistema estructurado de forma modular.
- Desventajas:
  - El sistema no es de libre distribución. Es posible adquirir el sistema mediante el pago de unas tasas de desarrollo y de distribución.
  - El alta de datos en el sistema es un proceso largo y laborioso.
  - No implementa la exportación de los datos de las colecciones.
  - No permite consultas externas de la información del sistema.

## **4. Biolink**

BioLink (Commonwealth and Organisation, 2005) ha sido diseñado principalmente para trabajar con especímenes. Destinado para el uso de taxónomos, ecólogos, gestores de colecciones y biogeógrafos.

## 4.1. Características

La información referente a los especímenes es almacenada en un conjunto de tablas para su posterior uso. Incluye una serie de herramientas *software* para la importación de datos al sistema BioLink.

Veamos las características de BioLink teniendo en cuenta los requerimientos establecidos:

1. **Gestión de especímenes:** permite almacenar información referente a distintas colecciones de especímenes.
2. **Consultas:** establece un conjunto de consultas sencillas mediante una herramienta independiente.
3. **Préstamos de material:** Gestión de préstamos del material a otros centros. Incluye tratamiento para la devolución parcial de un préstamo.
4. **Información multimedia:** Permite la captura de imágenes, videos y sonidos sobre un espécimen.
5. **Referencias bibliográficas:** mantiene un registro mínimo de la información bibliográfica.
6. **Información taxonómica:** Hace uso de un Explorador que permite buscar nombres científicos existentes en el sistema y examinar las características del mismo.
7. **Distribución de Especímenes:** permite dibujar mapas con la distribución de los taxones elegidos y existentes en el sistema.

## 4.2. Análisis de Características

Este es uno de los sistemas más completos de los encontrados hasta el momento. Biolink ha incorporado algunas utilidades que permiten visualizar la

información espacial mediante mapas. Se pueden enumerar las siguientes ventajas y desventajas del sistema:

- Ventajas:
  - Gran capacidad de gestión de los datos.
  - Incorpora el tratamiento de préstamos.
  - Integración con visualizadores espaciales para la generación de mapas.
- Desventajas:
  - No existe interfaz de consulta externa.
  - No permite la exportación de los datos en algunos de los formatos existentes.

## **5. Herbar**

Herbar es una aplicación para informatizar y gestionar colecciones botánicas (Pando, 1996). Es la aplicación estándar de la Asociación de Herbarios Ibero-macaronésicos (AHIM<sup>3</sup>). Actualmente, el soporte y el desarrollo de Herbar es realizado por el la Unidad de Coordinación del Nodo Español de GBIF (Unidad de Coordinación de GBIF en España, 2006a).

---

<sup>3</sup> AHIM: disponible en: [www.ahim.org](http://www.ahim.org)



## 5.1. Características

Esta aplicación está desarrollada en Microsoft Access y actualmente en su versión 3.5 (ver Figura 30), ha incorporado el tratamiento de información multimedia. Veamos las características del sistema:

1. **Gestión de especímenes:** altas y modificaciones de especímenes. Gestión de las localidades y tratamiento de información para su georreferenciación.
2. **Generación de Etiquetas:** distintos formatos de impresión de etiquetas con códigos de barras.
3. **Consultas:** formulario de consulta simple de los especímenes informatizados. Posibilidad de realizar las consultas accediendo directamente a la herramienta de consulta de Microsoft Access y a las tablas de la base de datos.
4. **Préstamos de material:** formularios de gestión de préstamos de material. Permite listar el material prestado.
5. **Información multimedia:** permite asociar una imagen a una muestra. La imagen es almacenada en una carpeta del sistema.
6. **Información taxonómica:** control de los nombres científicos otorgados a los especímenes permitiendo realizar filtros mecanográficos.
7. **Exportación de información:** permite exportar información en formato Darwin Core.



Figura 30. Pantalla principal de Herbar

## 5.2. Análisis de Características

Uso de *Microsoft Access* como motor de base de datos y gestor de la interfaz del usuario. Reduce un tanto la potencia del sistema así como la posibilidad de posteriores revisiones funcionales del mismo. Se pueden listar las siguientes ventajas e inconvenientes:

- Ventajas:
  - Es un sistema sencillo de usar, lo que es importante para fomentar el uso de sistemas para informatizar las colecciones.
  - Modelo de datos completo.
  - Gestión de localidades.
  - *Software* de libre distribución.
  - Posibilidad de intercambiar información en formato *Darwin Core*
- Desventajas:
  - No oferta servicios externos de consulta.

- No se controlan las operaciones que realizan los usuarios del sistema.

## **6. Sistema de gestión de Herbarios e investigación Botánica: Brahms**

El departamento de Botánica de la Universidad de Oxford ha desarrollado el sistema de Gestión de Herbarios e Investigación Botánica llamado B.R.A.M.H.S (*Botanical Research And Herbarium Management System*) (Filer, 1994; University of Oxford, 1985). El desarrollo comenzó en 1985 y ha pasado ya por más de cinco revisiones importantes del *software*. Es usado por el Herbario de Oxford y por taxónomos de proyectos sobre los géneros *Acacia*, *Agathis*, *Leucaena*, *Lupinus*, etc.

### **6.1. Características**

Proyecto ampliamente soportado por la Universidad de Oxford. Desde su página Web se puede descargar la documentación y ejemplos de distintas colecciones (<http://www.brahmsonline.com/>).

A continuación se muestran las características del sistema teniendo en cuenta el esquema establecido anteriormente:

1. **Control de usuarios:** permite crear usuarios para acceder al sistema con clave. No realiza ningún tipo de control de las operaciones permitidas a los usuarios.
2. **Gestión de especímenes:** altas, modificaciones, bajas de especímenes, información sobre localidades.
3. **Consultas:** módulo de consulta simple y pre-establecido que no permite al usuario realizar consultas avanzadas.

4. **Información multimedia:** gestiona imágenes asociadas a especímenes mediante un módulo independiente.
5. **Referencias bibliográficas:** formulario simple de gestión de referencias bibliográficas.
6. **Distribución de Especímenes:** al gestionar las localidades de un espécimen facilita el tratamiento de las referencias espaciales de un espécimen y permite mostrar la distribución de los especímenes.

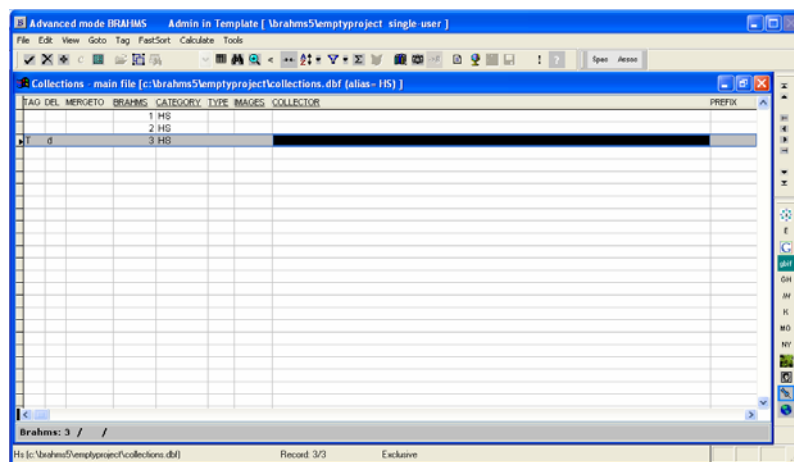


Figura 31. Captura pantalla principal BRAHMS

## 6.2. Análisis de Características

Este sistema presenta una gran madurez debido al amplio periodo de desarrollo que ha sufrido desde sus inicios en 1985. No obstante, también presenta deficiencias que hacen del sistema un *software* que dicta mucho de ser intuitivo. La interfaz de usuario no facilita el trabajo de los usuarios como se puede observar en la Figura 31.

Tras evaluar el *software* se pueden mostrar las siguientes ventajas e inconvenientes:

- Ventajas

- Gran capacidad de gestión de los datos de las colecciones biológicas.
- sistema desarrollado en módulos que clarifican las partes del mismo.
- Desventajas
  - El sistema no incorpora tareas propias de gestión de un centro como pueden ser los préstamos de material o consultas externas.
  - No existe un control de las distintas operaciones que puede realizar un usuario dentro del sistema.
  - No realiza exportaciones de especímenes en alguno de los estándares comentados anteriormente.
  - No incorpora ningún mecanismo de consulta externa.

## **7. Sistema para la Gestión de Herbario mediante Access (Herbarium Access Management System)**

Sistema desarrollado por la Universidad de California para la gestión de Herbarios (University of California Davis, 2004). Esta versión está disponible para *Microsoft Access 2000*. Es un sistema simple que permite almacenar información de especímenes.

### **7.1. Características**

El sistema desarrollado mediante Microsoft Access presenta una interfaz amigable. No obstante, es un sistema con poca funcionalidad y que no cubre las necesidades para la gestión de las colecciones biológicas. Mediante este sistema podemos:

1. **Gestión de especímenes:** Almacena información sobre su determinación nomenclatural y localización (ver Figura 32).
2. **Consultas:** permite realizar consultas por un conjunto de campos establecidos por el sistema sin poder enlazarlos entre sí. Presenta un sistema muy amigable de consulta. Mediante la utilización de Microsoft Access se pueden realizar consultas y modificaciones accediendo directamente a las tablas que forman la base de datos.
3. **Préstamos de material:** formulario que registra el material prestado. Mediante un campo de texto se pueden concatenar los identificadores del material prestado.
4. **Referencias bibliográficas:** permite gestionar el material bibliográfico del centro. No está asociado a la muestra, solamente gestiona el material bibliográfico del centro.
5. **Servicios Externos:** permite acceder al sistema mediante un conjunto de páginas en lenguaje ASP (*Active Script Page*). Es necesario tener instalado el Servidor Web propietario de Microsoft, llamado IIS (*Internet Information Server*).

Figura 32. Entrada de datos de especímenes en Herbarium Access

## 7.2. Análisis de Características

La sencillez es la principal característica de este sistema pero no presenta muchas de las características importantes necesarias para el control de los datos de las colecciones como el control de usuarios, módulos de exportación de datos, etc.

A continuación se muestra el conjunto de ventajas e inconvenientes del uso del sistema.

- Ventajas:
  - Tiene una interfaz amigable.
  - Permite almacenar información sobre especímenes.
  - Es un sistema de libre distribución.
  - Exportación en formato Darwin Core.
- Desventajas:
  - No permite el control de usuarios.
  - No tiene un módulo de búsqueda de información.
  - Las consultas externas al sistema están limitadas para el uso del Servidor Web IIS, caracterizado por sus problemas de seguridad.
  - No permite la exportación de información.

## 8. Un sistema para la Gestión de Herbarios: TRACY

El *software* Tracy (Minnigerode, 1997) es un sistema para la gestión de colecciones de herbarios. No es un sistema de grandes capacidades ya que fue diseñado principalmente para facilitar solamente la entrada rápida de datos sobre

especímenes. El sistema fue desarrollado por M. David Minnigerode (miembro de un grupo de trabajo de Bioinformática en Texas).

## 8.1. Características

Tracy ha sido desarrollado principalmente para:

1. **Gestión de especímenes:** formulario de entrada rápida que permite registrar información sobre el espécimen.
2. **Consultas:** establece distintas consultas predefinidas y permite combinar condiciones para realizar una búsqueda.

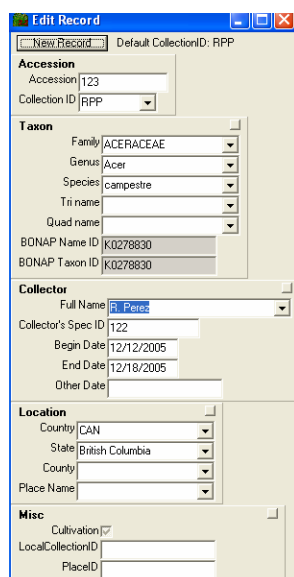


Figura 33. Plantilla de datos sobre especímenes

El sistema presenta una interfaz amigable (ver Figura 33) pero no incorpora la funcionalidad necesaria para el centro, tal como la gestión de préstamos, usuarios, imágenes, etc.

## 8.2. Análisis de Características

Uno de los sistemas más simples de entre los analizados. Su sencillez le ha llevado a incorporar muy pocas funcionalidades. Este sistema sólo permite



almacenar y consultar información sobre especímenes. No permite la exportación, control de usuarios, gestión nomenclatural, etc.

## 9. PLabel: Herbarium Label Program

El sistema PLabel (University of Florida Herbarium, 2000) ha sido diseñado para ayudar a la preparación de las etiquetas de herbario y para almacenar la información que éstas necesitan. El sistema ha sido desarrollado por el Herbario de la Universidad de Florida.

### 9.1. Características

PLabel permite almacenar toda la información referente a especímenes mediante una plantilla generada en Clipper<sup>4</sup>.

The screenshot shows the 'PLabel for Windows - Main Form' window. The form is divided into several sections for data entry:

- Header:** Labels File, Browse Page, Labels File, Edit Page, Add Label, Taxa, Familia, Counties, Print.
- Identification:** Fam./C. (128), Genus (Medicago), Species (lupulina), Authority (L.).
- Ranking:** Rank1 (Intraspecific 1), Authority, Rank2 (Intraspecific 2), Authority, Gn. Modif., Sp. Modif.
- Collector:** Det. (K. D. P.), Det. Date, Common.
- Geography:** Country (United States), State (Florida), County (Alachua), Locality Name (don't repeat in locality).
- UTM Zone:** Easting, Northing, Latitude, Hg., Long., Elev., Datum, Sec. TR, Elev.
- Locality:** NW edge Gainesville on E side of US 441 at NE corner of jct. with NW 53rd. Ave.
- Habitat:** Edge of ditch adjacent to wet pinewoods.
- Plant Description:** Flowers yellow.
- Collector:** Collector First (Kent D.), Collector Last (Perkins), Coll. # (2000), Coll. With (Ralph Nader).
- Date Coll.:** 21 June 2004.
- Other Fields:** # Labels (1), Cult? (checkbox), Espro. St., Field Priv., Photo, DNA, Spitt, Wood, Living, Prosted.
- Notes:** A major retailer's proposal to develop this site has been turned down by the city.
- Footer:** Search, Add Label, Quick Search, AS ENTERED, Record #: 1, Close.

Figura 34. Entrada de datos de especímenes de PLabel

<sup>4</sup> Lenguaje de programación para MS-DOS de Microsoft y que utiliza DBASE III como gestor de bases de datos. Creado originalmente como compilador de DBASE III en 1985.

El sistema PLabel tiene las siguientes características:

1. **Gestión de especímenes:** conjunto de formularios (ver Figura 34) que permiten almacenar información sobre localidad, hábitat, colector,...
2. **Generación de Etiquetas:** gestión simple de impresión de etiquetas con información y formato predeterminado.
3. **Consultas:** permite la navegación entre los registros existentes mediante una barra inferior. Además permite la búsqueda simple de especímenes por alguno de los campos que recoge la plantilla de datos (colector, familia, localidad, etc.).
4. **Información taxonómica:** formulario independiente para almacenar información taxonómica que será referenciada por el espécimen.

Pese a ser un sistema pobre en funcionalidad, permite gestionar la información sobre colecciones. No obstante, no incorpora operaciones como pueden ser los préstamos de material, consultas avanzadas, gestión de imágenes, etc.

## 9.2. Análisis de Características

PLabel fue desarrollado para satisfacer una sola necesidad, la generación de las etiquetas que rotulan las muestras. Por ello, la información que el sistema permite almacenar no es muy extensa. Actualmente PLabel está en su versión beta y no hay información acerca de la versión final del sistema.

- Ventajas:
  - Es un sistema de libre distribución.
  - Tiene una interfaz amigable.

- Gestión de información sobre especímenes.
- Desventajas:
  - Sistema en versión beta.
  - No existen usuarios y control de las operaciones que estos puedan realizar sobre la información del sistema.
  - No existe la posibilidad de exportar información en ninguno de los estándares vistos con anterioridad.
  - No gestiona los préstamos de material.
  - La impresión de etiquetas no incorpora código de barras para facilitar la lectura automatizada de la información del mismo.

## **10. Herbario Virtual Express: VHE (Virtual Herbarium Express)**

El sistema Herbario Virtual Express (V.H.E.: *Virtual Herbarium Express*) ha sido desarrollado por el Jardín Botánico de Nueva York (John D. and Catherine T. MacArthur Foundation, 1996). El sistema está diseñado para ser utilizado por botánicos que gestionan la información sobre especímenes (ver Figura 35). El sistema hace uso de Microsoft Access como motor de bases de datos y para la interfaz de usuario.

### **10.1. Características**

El sistema está desarrollado en *Microsoft Access* lo que lo hace bastante limitado. A continuación se muestran las características del sistema:

1. **Gestión de especímenes:** altas y modificaciones de especímenes. Almacenamiento de información sobre su identificación y sobre su recolección.
2. **Generación de Etiquetas:** no incluye la incorporación de códigos de barra. La impresión se realiza en un formato fijo.
3. **Consultas:** sistema básico de consulta por número de registro o número de colector.
4. **Referencias bibliográficas:** gestiona información sobre los autores y referencias bibliográficas.



Figura 35. Pantalla principal de VHE

## 10.2. Análisis de Características

VHE es un sistema gratuito desarrollado por el Jardín Botánico de Nueva York para la gestión de información sobre especímenes. Para una mejor gestión de los especímenes se ha incorporado un módulo de gestión de autores. Tras evaluar el sistema no se ha encontrado ningún mecanismo para el intercambio de la información existente en las bases de datos. A continuación se detallan las principales ventajas y desventajas del sistema:

- Ventajas:
  - Sistema de libre distribución.

- Formularios amigables.
- Gestión de autores, publicaciones y nombres taxonómicos.
- Desventajas:
  - No permite intercambiar información en los estándares conocidos.
  - No realiza labores de gestión del centro como consultas externas y préstamos de material.
  - No incorpora mecanismos de consulta de los datos existentes en la colección.

## 11. Resumen comparativo

Tras examinar y analizar los distintos sistemas existentes, se puede apreciar la disparidad en las características que estos incorporan, y en la forma de abordar las distintas funcionalidades establecidas como requerimientos del *software*.

A continuación mostramos un cuadro resumen (ver Tabla 4) de las principales características encontradas tras el análisis del *software*.

sistemas\ Características	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
Biota 2											
Specify											
Biotica											
Biolink											
Herbar											
Brahms											
Herbarium Access Management System											
Tracy											
PLabel											
VHE											

Tabla 4. Tabla comparativa de los sistemas *software* existentes (ver leyenda en siguiente tabla)

Identificador	Significado
1.	Control de usuarios.
2.	Gestión de especímenes.
3.	Generación de Etiquetas.
4.	Consultas.
5.	Préstamos de material.
6.	Información multimedia.
7.	Referencias bibliográficas.
8.	Información taxonómica.
9.	Servicios Externos.
10.	Exportación de información.
11.	Distribución de Especímenes.

Tabla 5. Leyendas de tabla comparativa

### 11.1. Análisis de la situación actual

Tras examinar y establecer los requerimientos del sistema que satisfacen las necesidades de gestión de las colecciones, se han evaluado los diez sistemas más representativos.

Como se muestra en la Tabla 4 no existe ningún sistema que cumpla con todos los requerimientos establecidos y que, por tanto, satisfaga completamente las restricciones del problema.

Veamos a continuación, cada uno de los requerimientos y la forma de acometerlos por los distintos sistemas evaluados.

El primer inconveniente que se encuentra en los distintos sistemas es la carencia de un **control de usuarios** y de las operaciones que estos pueden realizar dentro del sistema. Tan solo tres de los diez sistemas evaluados realizan un control de acceso al sistema mediante usuario y clave.

La **gestión de especímenes** es incorporada por todos los sistemas evaluados, con mayor o menor grado de dificultad en la entrada de los datos por parte del usuario.

La **generación de Etiquetas** es un paso importante para la catalogación y almacenamiento de la muestra. Tan solo la mitad de los sistemas incorpora la realización de etiquetas. La incorporación de códigos de barras acelera el proceso

de gestión, por ejemplo, cuando necesitamos realizar un préstamo. La generación de etiquetas con código de barras sólo está presente en dos de los cinco sistemas que generan etiquetas.

Una vez que el sistema sirve de soporte para la informatización de las colecciones, su principal tarea será la de permitir obtener información a partir de los datos existentes. Todos los sistemas analizados incorporan una parte de **consulta** no siendo iguales en todos los sistemas. Generalmente, se plantean sistemas de consulta muy simples en los que el usuario elige los campos de búsqueda y establece sus condiciones.

La principal tarea del centro que gestiona y mantiene las colecciones es actuar como centro de documentación que proporciona un mínimo de servicios a otros centros e investigadores. El **préstamo de materiales** se ha incluido en cinco de los sistemas analizados, lo cual se presenta como un inconveniente puesto que no se podrá saber qué material está en el centro o cuál está prestado.

Se puede asociar **información multimedia** referente a la muestra para que el proceso de informatización sea más sencillo. Además, mediante la información multimedia (por ejemplo, las imágenes) en algunos casos, se puede evitar manipular la muestra evitando el consiguiente deterioro. De los sistemas analizados, sólo la mitad de ellos permite almacenar información multimedia sobre la muestra.

Las **referencias bibliográficas** permiten asociar las citas utilizadas para la asignación del nombre científico al espécimen. La mayor parte de los sistemas incorporan el tratamiento de la bibliografía (seis de los diez sistemas analizados).

La **información taxonómica** permitiría filtrar y depurar los datos almacenados en las bases de datos. Solamente cinco de los sistemas evaluados realiza un tratamiento independiente de la taxonomía. No obstante, ninguno de los sistemas incorpora algún mecanismo de comunicación con los distintos proyectos nomenclaturales existentes y comentados en el Capítulo I.

La consulta del material existente en un centro de estas características es un procedimiento usual que tienen que llevar a cabo los investigadores. La incorporación de mecanismos de **consulta externa** permite al investigador realizar sus consultas sin necesidad de ponerse en contacto con el centro. El acceso *on-line* a un subconjunto de los datos almacenados es un paso importante para los centros gestores. Tan solo uno de los sistemas evaluados incorpora un servidor Web que permite acceder a los datos del sistema. Este sistema permite el acceso indiscriminado a los datos del sistema creando un efecto contrario al deseado. Es necesario mostrar la información de forma *on-line*, pero no cualquier información ya que cualquier persona podría hacer un mal uso de la misma (por ejemplo, localidades exactas de especies amenazadas).

Actualmente, existe la necesidad de compartir los datos de las colecciones. Por ello, surgen iniciativas mundiales como GBIF y BioCASE que proponen estándares de intercambio, proveen del *software* necesario y dan soporte a los proveedores que desean exportar sus datos. A pesar de la importancia de la adaptación de los sistemas, sólo dos de los diez sistemas evaluados incorporan de forma directa los mecanismos para **exportar sus datos** en algunos de los protocolos analizados en el Capítulo I.

La georreferenciación almacenada sirve para mostrar la **distribución de Especímenes**. Solamente dos de los sistemas analizados incorporan la posibilidad de dibujar un mapa en el que sitúan el punto donde se recogió el espécimen.

Como se puede observar, no existen sistemas que cumplan con todas las necesidades básicas de gestión de las colecciones complejas.

## **11.2. El uso de los datos.**

La informatización de las colecciones es el primer y principal paso para la realización de estudios de biodiversidad. La realización de aplicaciones avanzadas (Stockwell, 1997), basadas en las aplicaciones generales de informatización son la fuente para el avance y la generación de conocimiento.



Hasta ahora, todos los sistemas estudiados no han incorporado ninguna funcionalidad que permita al investigador del centro analizar los datos existentes en su sistema. Esto limita las posibilidades del *software* y hace que el investigador tenga que trabajar con datos primarios.

Una vez examinadas las alternativas existentes no encontramos un *software* que cubra todas las necesidades.

Por ello se planteó el desarrollo de un sistema que permita cubrir las perspectivas desde dos puntos de vista:

1. Desde el punto de vista de **centro de documentación**: permitiendo almacenar y gestionar toda la información necesaria.
2. Desde el punto de vista de **centro de investigación**: aplicación de técnicas de inteligencia artificial para la extracción de información. Facilitar las operaciones necesarias al investigador para la realización de estudios de biodiversidad.



### **Capítulo III. BioMEN: Un Sistema de Información para el tratamiento de colecciones complejas**

Las colecciones complejas requieren de un sistema que permita gestionar tanto la información existente, como los procesos asociados a las mismas. Es importante que el sistema intercambie información de forma estructurada y estandarizada con los proyectos existentes.

Tras analizar los sistemas y comprobar la inexistencia de un sistema que reúna los requerimientos funcionales establecidos realizamos el diseño y desarrollo de un sistema de información basado en un Sistema Multiagente capaz de gestionar las colecciones de forma inteligente, denominado Bio.M.E.N. (*Biological Management Executed over Network*).

La utilización de técnicas de inteligencia artificial permiten que el sistema no sea solamente un repositorio de datos, proporcionando al usuario la extracción de conocimiento tras un tratamiento inteligente de los datos.

La comunicación con fuentes externas heterogéneas y la integración de la información procedentes de las mismas con los datos existentes en el sistema, aumentan la calidad de los datos y de la información que el sistema puede proporcionar al investigador.

Previo a cualquier estudio sobre los datos, tenemos que dotar al sistema de las características necesarias para gestionarlos. Por ello, BioMEN se ha realizado con el objetivo de cubrir tanto las necesidades de gestión, como las de investigación.

A continuación se describen los pasos seguidos para el desarrollo e implementación de BioMEN. Posteriormente se detallarán sus características desde el punto de vista de la gestión de colecciones. En el siguiente capítulo se

tratarán las características de BioMEN que lo hacen un sistema inteligente para el tratamiento de las colecciones.

## **1. Desarrollo e implementación**

En este epígrafe se tratan el modelo de ciclo de vida utilizado para el desarrollo, los requerimientos funcionales y no funcionales del sistema y la arquitectura del sistema.

### **1.1. Modelo de ciclo de vida**

De entre todos los distintos modelos de ciclo de vida existentes (Modelo en Cascada, Construcción de Prototipos, Desarrollo Rápido de Aplicaciones, Espiral de Böhem y Modelo Evolutivo (Pressman, 1997)) se ha optado por el **Modelo Evolutivo**.

El **Modelo Evolutivo** (Pressman, 1997) combina elementos del Modelo en Cascada con la filosofía interactiva de Construcción de Prototipos. Se basa en la construcción sucesiva de prototipos, que cada vez se aproximan más a la aplicación final. Cada secuencia lineal produce un incremento en la funcionalidad del *software*.

Un prototipo no es una simulación, sino un paso hacia la realidad que se quiere construir. Facilita al analista la creación de un modelo del *software* que se va a construir. Además, no tiene que ser necesariamente un sistema completo, puede ser una descripción en papel que refleje la interacción persona – máquina y facilite al usuario la comprensión de ésta; también puede ser un sistema que implemente un subconjunto de las funciones requeridas o, incluso, un sistema ya existente que implemente las funciones deseadas pero que necesita ser mejorado.

Una diferencia con respecto al modelo de Construcción de Prototipos, es que el modelo evolutivo se centra en la entrega de un prototipo operacional con cada incremento.

Hay varias formas de prototipado:

- **El prototipado de las especificaciones**, cuyo objetivo es indicar los requisitos del *software* y extraer los requerimientos. Genera prototipos desechables.
- **El prototipado cíclico**, en el que el prototipo va evolucionando hasta convertirse en el definitivo. Los prototipos que genera son “prototipos incrementales”.

La estructura básica del proceso viene dada por la Figura 36:

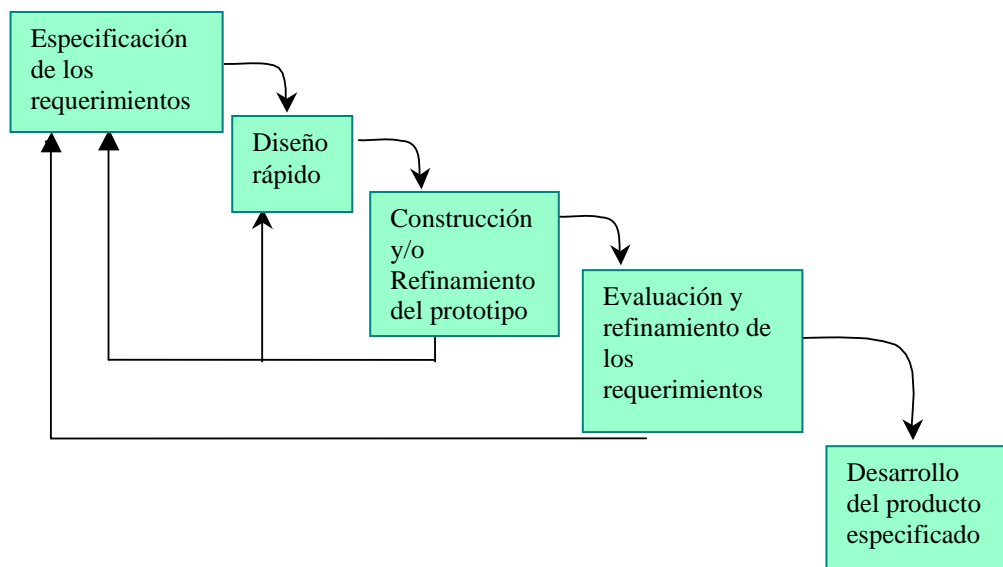


Figura 36. Modelo de ciclo de vida evolutivo

Las ventajas de este modelo de ciclo de vida son:

- Mayor fiabilidad de la especificación de los requerimientos y del proyecto.
- Se obtienen pronto versiones preliminares y operativas del proyecto, evitando la impaciencia del usuario.

- Documentación completa y de calidad (gracias al prototipado de las especificaciones).
- Facilidad para realizar pruebas con el prototipo.
- Comodidad y sencillez para realizar modificaciones.
- Al tratarse de un proceso incremental se reutilizan partes implementadas.

Aunque también hay inconvenientes tales como:

- Incorporación de características no deseables de forma inadvertida (defectos que en versiones preliminares pueden ser permisibles, pero que bajo ningún concepto lo son en versiones finales, detalles descuidados, etc.).
- Puede suceder que el usuario se conforme con las versiones preliminares sin favorecer o limitando las posibilidades de llegar al producto definitivo.

Este es el modelo de ciclo de vida utilizado para el desarrollo de BioMEN. En particular se han desarrollado dos prototipos. El primero de ellos se centró en satisfacer los requerimientos funcionales y no funcionales expuestos hasta el momento. El segundo prototipo se ha desarrollado para satisfacer las necesidades investigadoras basadas sobre los datos ya gestionados por el primer prototipo.

## **1.2. Requerimientos del sistema**

A lo largo del presente documento se han tratado los requerimientos funcionales del sistema (control de usuarios, gestión de especímenes, generación de etiquetas, etc.) establecidos en el Capítulo II. Además, para el desarrollo de

BioMEN se han tenido en cuenta un conjunto de requerimientos no funcionales<sup>5</sup> que se deben de satisfacer y que condicionan el desarrollo del mismo.

### **1.2.1. Requerimientos no funcionales**

Los requerimientos no funcionales establecidos para el desarrollo de BioMEN fueron:

- El sistema tiene que ser accesible desde cualquier plataforma, independientemente del tipo de sistema operativo.
- El sistema tiene que ser de libre distribución y desarrollado bajo la filosofía GPL.
- Debe de ser accesible desde Internet.
- La interfaz de usuario debe ser adaptable a las necesidades de cada centro.
- El tiempo de respuesta del sistema ha de ser aceptable.
- Debe de incluir herramientas de Administración para poder ser mantenido con facilidad.
- El acceso al sistema debe de ser seguro. Las comunicaciones tienen que estar cifradas entre el cliente y el servidor.
- Se establecerán permisos para los distintos usuarios del sistema, impidiendo así la utilización de ciertas partes del sistema a usuarios no autorizados.
- El sistema debe permitir realizar copias de seguridad con facilidad.

---

<sup>5</sup> Los requerimientos no funcionales son restricciones al conjunto de las posibles soluciones.

El sistema desarrollado tiene que cumplir con todas los requerimientos expuestos.

### 1.3. Arquitecturas y *Software* de desarrollo.

Debido a los requerimientos no funcionales debemos de realizar un sistema que sea accesible desde cualquier parte.

BioMEN presenta una arquitectura Cliente /Servidor donde los navegadores Web actúan como clientes y el servidor Web actúa como Servidor del sistema. El uso de la Web hace que BioMEN esté accesible desde cualquier parte del mundo conectada a la red.

La arquitectura Cliente/Servidor permite además:

- Centralizar los recursos necesarios para el servidor de la aplicación.
- Actualizar del *software* sin afectar a los clientes.
- Acceder de forma unificada a la información del centro.

BioMEN presenta una estructura Cliente/Servidor (ver Figura 37) que haciendo uso de servidores Web permite acceder al sistema de forma totalmente *on-line*.

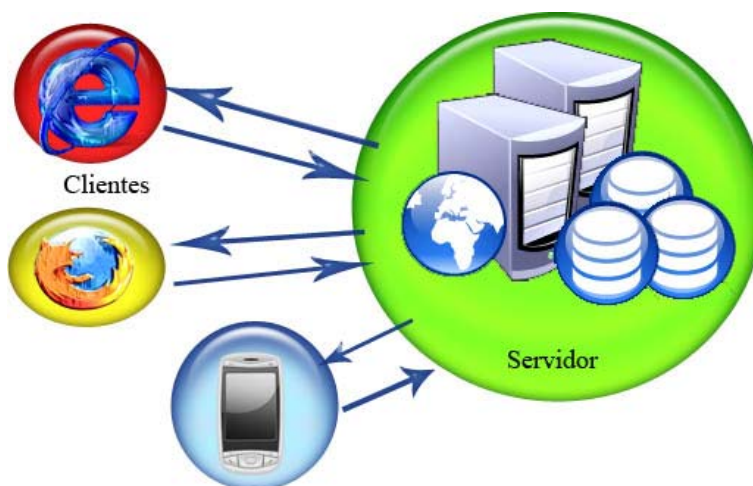


Figura 37. Estructura Cliente/Servidor de BioMEN



La realización de un sistema Web permite adaptarse, prácticamente, a cualquier plataforma debido a la existencia de tecnologías de desarrollo que permiten ejecutarse en distintos sistemas operativos.

### 1.3.1. *Software* necesario para la solución del problema.

En este apartado vamos a tratar de comentar tanto el *software* destinado tanto para el Servidor, como el *Software* destinado para el Cliente, teniendo en cuenta los siguientes apartados:

- **Servidor Web.** La elección del servidor Web dependerá del lenguaje de desarrollo elegido y de su posibilidad para ejecutarse distintas plataformas.
- **Lenguaje de Desarrollo.** El lenguaje de desarrollo nos marcará las posibilidades de portabilidad, fiabilidad y seguridad en el desarrollo del sistema.
- **Gestor de Base de Datos.** El tratamiento de las colecciones complejas está marcado por la gran cantidad de datos a gestionar. Por ello, se debe de elegir un sistema gestor de bases de datos estable y seguro que gestione este volumen elevado de datos con suficiente rapidez y fiabilidad.
- ***Software* de navegación.** El navegador actúa de cliente en el sistema. La forma de interpretar las etiquetas HTML (*hypertext markup language*) no es la misma en distintos navegadores (por ejemplo: Internet Explorer, Firefox, opera, etc.).

El núcleo principal del sistema lo establece la terna: Lenguaje de Programación, Servidor Web y Gestor de Bases de Datos. Hoy en día, existen alternativas conjuntas que permite su ejecución en distintas plataformas. Debido a la utilización de la Web como puerta para el acceso a BioMEN, el sistema operativo del cliente es independiente del sistema operativo utilizado para ejecutar el servidor.

## ***Servidor Web***

Un **Servidor Web** es un sistema que implementa el protocolo HTTP (*hypertext transfer protocol*). Este protocolo está diseñado para transferir lo que llamamos hipertextos, páginas Web o páginas HTML. El servidor Web se mantiene a la espera de peticiones HTTP llevadas a cabo por un cliente HTTP, llamado **navegador**.

Sobre el servicio Web se pueden disponer aplicaciones Web. Estas son fragmentos de código que se ejecutan en el servidor y que devuelven el código html correspondiente, al cliente que lanzó la petición.

Algunos de los servidores Web más importantes a destacar son:

- **Apache (the Apache Software Foundation, 2006)**: es un servidor de gran escalabilidad soportado principalmente por módulos dinámicos. Esto hace que sea adaptable a la interacción con *software* muy diverso. Es un proyecto de código abierto para plataformas *Unix*, *Windows* y otras. El servidor Apache se desarrolla dentro del proyecto *HTTP Server* de la fundación Apache (*Apache Software Foundation*).
- **Internet Information Server (IIS)**: Es un conjunto de servicios propietarios de *Microsoft* que permiten proporcionar páginas Web tanto de forma local como remota. Sólo puede ser ejecutado bajo sistemas Operativos *Windows*. El servidor está basado en módulos que dan capacidad al servidor para ejecutar distintos tipos de lenguajes (*Active Server Pages (ASP)*, *PHP*, *Perl*, ... )
- **Cherokee**: (<http://www.0x50.org>) Es un proyecto de *Software Libre* y publicado bajo licencia GPL con el objetivo de que sea accesible por el mayor número de personas y exista una comunidad de usuarios y desarrolladores que lo soporten y mejoren. Actualmente sólo está disponible para plataformas *Unix*.

De entre los servidores Web existentes nos hemos decantado por la utilización de Apache como servidor Web debido principalmente a dos razones:

1. Servidor que permite ejecutarse en múltiples plataformas.
2. Está soportado por la fundación Apache *Software* por lo que aporta fiabilidad y seguridad al proyecto.

El último informe proporcionado por la empresa NetCraft (WWS, 2005) nos muestra que cerca del 69% de los servidores Web son servidores Apache.

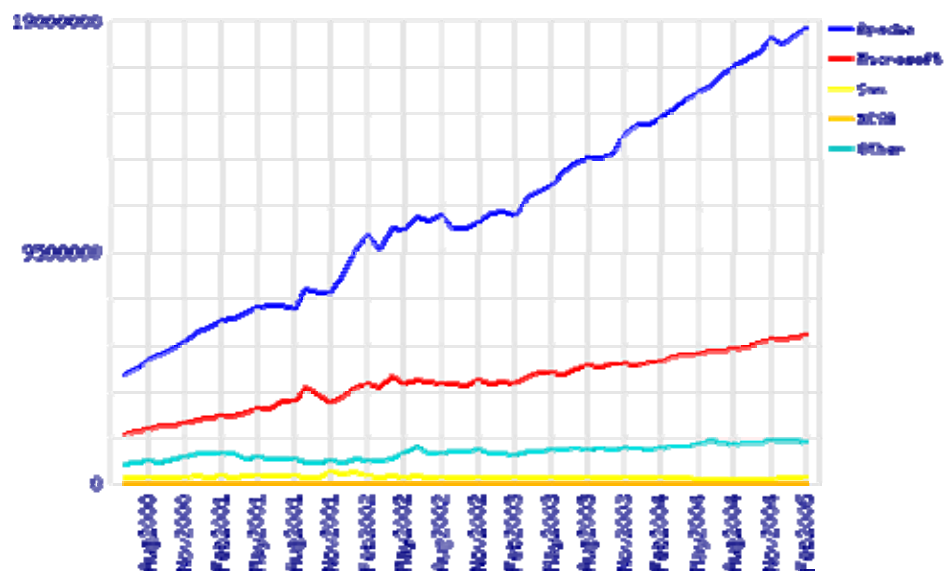


Figura 38. Comparativa de uso de los distintos servidores Web (WWS, 2005)

## ***Lenguaje de Desarrollo***

El lenguaje de desarrollo que se use para la construcción del sistema Web no tiene porqué ser único, aunque se ha de tener en cuenta que debe soportarlo el servidor Web, así como el cliente de navegación (Netscape, Internet Explorer, Firefox, ...).

### **HTML**

El *HyperText Markup Language* (HTML) es el lenguaje estándar que se usa para la creación de páginas Web y, por tanto, la base para poder construir el

sistema. La aplicación Web se ejecuta en el servidor Web construyendo el código HTML necesario que será interpretado por el navegador del cliente.

### JavaScript

Es un **lenguaje interpretado**<sup>6</sup> utilizado en páginas Web. Éste es interpretado por el navegador del cliente, haciendo que gran parte de las verificaciones y restricciones de la información que proporciona el cliente al servidor se validen mediante JavaScript. Por ejemplo, comprobar la existencia de un campo obligatorio y el formato de una fecha.

### Servlet's de Java

Un **Servlet** es una aplicación escrita en Java que se ejecuta en el marco de un servicio de red, (un servidor HTTP, por ejemplo), y que recibe y responde a las peticiones de uno o más clientes.

Las principales diferencias entre las tecnologías **CGI**<sup>7</sup> y Servlet son las siguientes:

- La tecnología Servlet proporciona las mismas ventajas del lenguaje Java en cuanto a **portabilidad** ("escribe una vez y ejecuta en cualquier lugar") y **seguridad**, ya que un servlet es una clase de Java.
- Los Servlet's proporcionan mayor **rendimiento**. Cuando son llamados por primera vez quedan activos en la memoria del servidor hasta que el

---

<sup>6</sup> Lenguaje interpretado es aquel que se ejecuta sin ser compilado. En cada ejecución del código, un interprete traduce él mismo a código máquina para que pueda ser ejecutado.

<sup>7</sup> CGI es el acrónimo de *Common Gateway Interface* (Pasarela de Interfaz Común). Es una tecnología importante que permite al cliente solicitar datos a una aplicación Web. Puede estar escrito en cualquier lenguaje, principalmente en *Perl*, *C* y *C++*.

sistema que controla el servidor los desactiva. De esta manera se minimiza en gran medida el tiempo de respuesta.

La utilización de Servlet's de Java y la posibilidad de cooperación entre los Servlet's y las distintas clases Java dotan al sistema de grandes proyecciones de desarrollo.

### ***Gestor de la Base de Datos.***

El gestor de base de datos sirve de acceso entre las aplicaciones y las bases de datos. Se compone de un lenguaje de definición de datos, lenguaje de manipulación de datos y de un lenguaje de consulta.

Podemos distinguir dos grupos de gestores de bases de datos:

- Uso libre y gratuito como por ejemplo *PostgreSQL*, *MySQL* y *SQLite*.
- Uso Comercial como por ejemplo *Oracle*, *dBase* y *Microsoft Access*.

Debido a que deseamos que el proyecto realizado sea de libre distribución y no necesite la adquisición de licencias de uso de ninguno de sus componentes se descartaron las alternativas comerciales.

De entre las alternativas de libre distribución *PostgreSQL* y *MySQL* son las alternativas a considerar para el sistema. Aunque ambos servidores tienen características muy similares nos hemos decantado por la utilización de **PostgreSQL** como gestor de bases de datos.

El principal motivo para no usar *MySQL* es que en su versión 4.1 estable no incorpora vistas, disparadores y procedimientos que facilitan el desarrollo y control de la integridad de los datos. No obstante, se podría usar cualquier gestor de bases de datos para el que exista un conector **JDBC** (*Java DataBase Connectivity*). La JDBC es una capa intermedia que aísla las particularidades de programación de cualquier sistema gestor de bases de datos. La JDBC permite la

ejecución de operaciones sobre Base de Datos desde el lenguaje Java independiente del sistema Gestor de Base de Datos.

### 1.3.2. Arquitectura de tres capas.

La construcción de una arquitectura por capas es un estilo de desarrollo en el que el principal objetivo es separar la capa de interfaz de usuario de la capa de los datos. La principal ventaja de esta metodología es que el desarrollo se puede realizar a varios niveles. Además, existe una mayor abstracción en el proceso de desarrollo y hace las partes del mismo más independientes. El principal inconveniente es la necesidad de establecer los cauces de comunicación entre las distintas capas que forman la arquitectura.

Las tres etapas que componen esta arquitectura son:

- **Capa de presentación:** es la parte de interfaz de usuario. Esta capa se encarga de recoger y mostrar las peticiones del usuario. Realiza un pre-procesamiento de los datos que se envían al sistema. Sólo se comunica con la capa lógica.
- **Capa lógica o de negocio:** es la capa que contiene las tareas y reglas que rigen el proceso. Se denomina capa de negocio (o lógica de aplicación) pues es aquí donde se establecen todas las reglas que deben cumplirse. Esta capa se comunica con la capa de presentación, para recibir las solicitudes y devolver los resultados, y con la capa de datos, para solicitar, almacenar o recuperar los datos.
- **Capa de datos:** es donde residen los datos. Puede estar formado por uno o más gestores de datos, que realizan todo el proceso de almacenamiento, gestión y recuperación de datos.

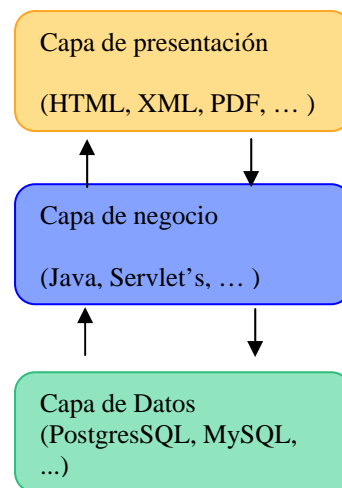


Figura 39. Arquitectura de tres capas

Como podemos observar en la Figura 39, modificar la interfaz a demanda del usuario final no implica la modificación del código del sistema, puesto que las capas están separadas.

Para el desarrollo de sistemas Web es importante mantener de forma independiente la interfaz del usuario de la lógica del sistema. La independencia entre las capas hace posible la personalización de las interfaces. De esta forma podemos incorporar tantos interfaces como usuarios (adaptación para discapacitados, multi-idioma, etc.).

Aunque el desarrollo en una arquitectura de tres capas es más complejo, la recompensa es mayor. La utilización de esta metodología de desarrollo ha permitido que BioMEN tenga las siguientes ventajas:

- Adaptación de la interfaz al usuario.
- Generación de distintos formatos de salida del sistema (XML, HTML, Excel, PDF, etc.).
- Intercambio de la capa de datos por otros sistemas gestores de datos.

### *Capa de presentación*

La evolución de los lenguajes de etiquetas ha facilitado la separación entre la capa de presentación y la capa de negocio. La combinación de XML (*Extensible Markup Language*) con XSLT (*Extensible StyleSheet Language Transformations*) nos ha permitido generar la interfaz de usuario a partir de los datos comunicados por la capa de negocio (en formato XML).

La función de XSLT es la de transformar de un formato en otros (ver Figura 40), tomando como formato fuente, datos generados en formato XML. El formato de salida puede ser muy diverso:

- XML: puede actuar como entrada para otro sistema.
- HTML: distintas interfaces dependiendo de las preferencias de cada usuario.
- PDF: obtener la salida en un formato portable de documentos (PDF).
- XLS: La salida de consultas se puede formatear en hojas de estilo.
- KML: representación de toda la información geoespacial en *Keyhole* para que pueda ser interpretado por *Google Earth*.

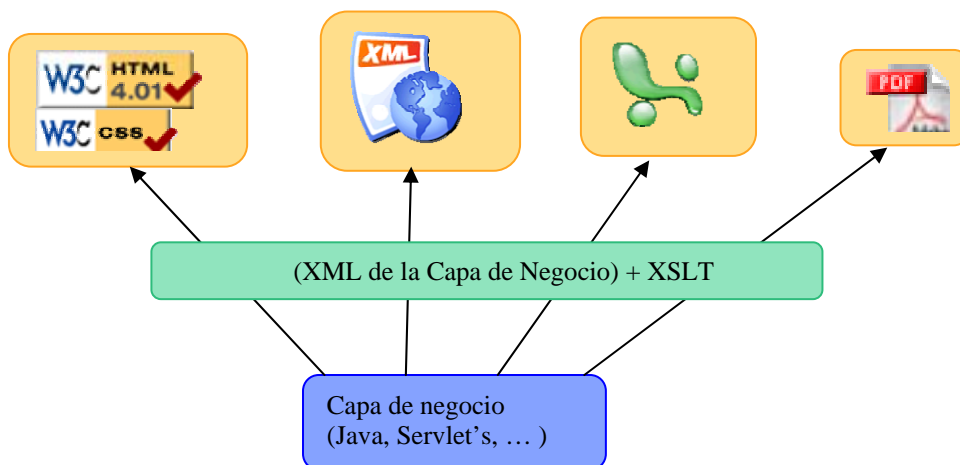


Figura 40. Generación dinámica mediante XML y XSLT



La Figura 40 muestra las distintas posibilidades existentes en BiOMEN para obtener los datos que se muestran al usuario. Un ejemplo de la salida en Excel se detalla a continuación:

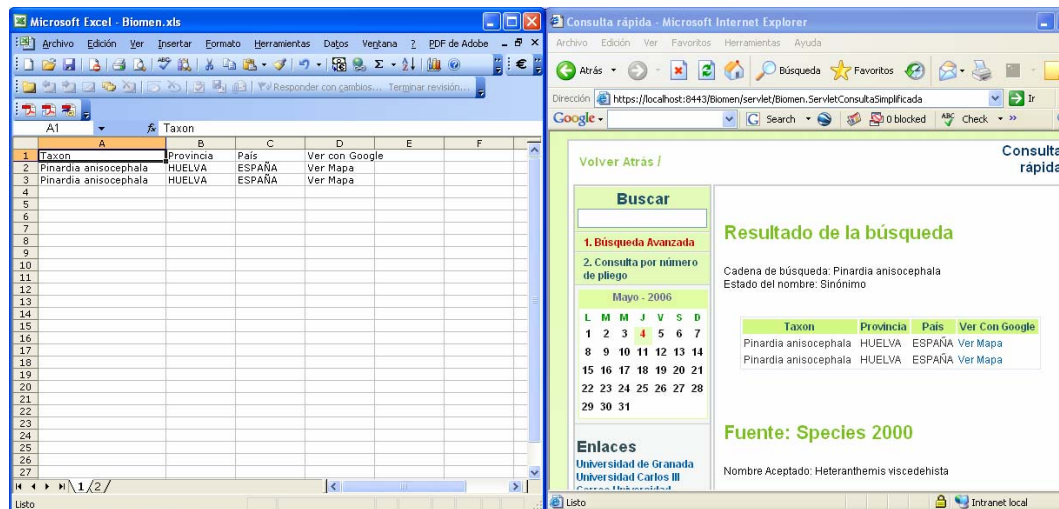


Figura 41. Salida de la interfaz en Hoja de cálculo Excel

El uso de transformaciones XSLT hace más versátil el sistema. Toda la presentación del usuario se realiza de una forma totalmente independiente de la capa de negocio. La capa de negocio se comunica con la capa de presentación usando XML.

### *Capa de Negocio*

La capa de negocio utiliza una ontología para expresar la información que transfiere a la capa de presentación. La información intercambiada se expresa en XML. El XML generado por la capa de negocio tiene que verificar una serie de características del dominio que están reflejadas por una ontología.

La ontología está formulada usando el lenguaje XML Schema, que aunque no sea muy expresivo, permite cubrir las necesidades del sistema.

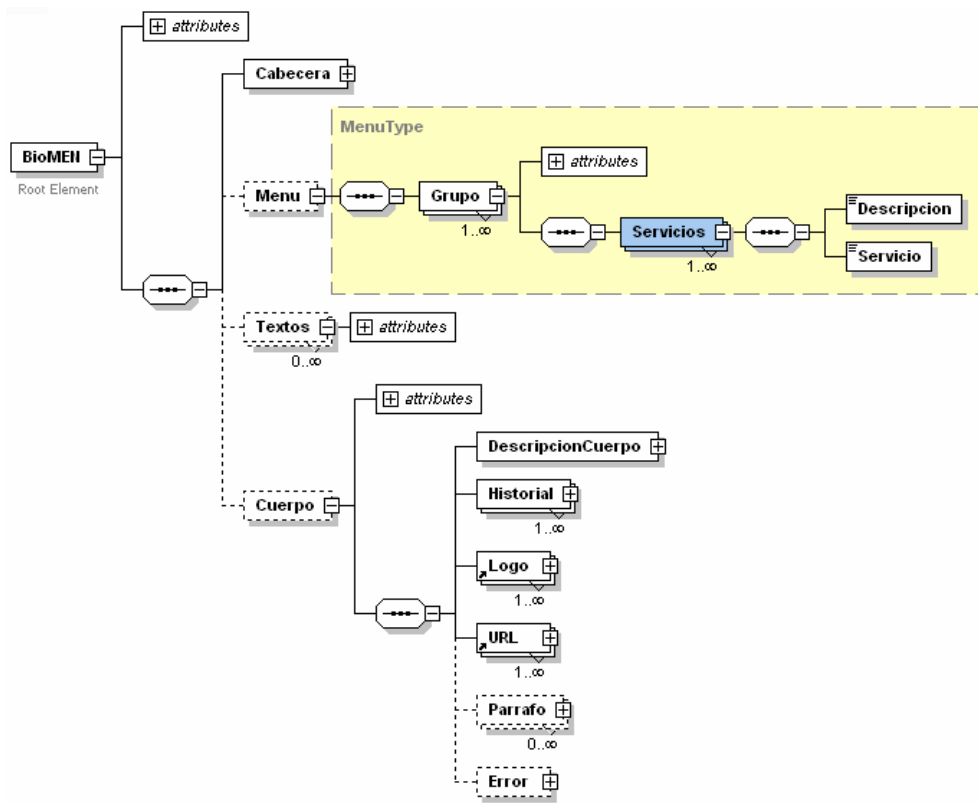


Figura 42. Fragmento de la ontología que define el dominio de la interfaz de usuario.

### ***Capa de datos.***

Obviamente es donde residen los datos. Está formada por uno o más gestores de bases de datos que realiza todo el almacenamiento de datos y recibe las solicitudes de almacenamiento o de recuperación de información desde la capa de negocio.

La capa lógica se comunica con la capa de datos a través de clases que implementan los objetos del sistema, bajo un paradigma de orientación a objetos.

## **2. Características**

La realización de un sistema gestionado a través de Internet hace que el sistema sea accesible desde cualquier parte del mundo. Este tipo de sistemas

adquieren más complejidad en el desarrollo que los sistemas de escritorio. Todas las características que se describen a continuación se realizan a través de Internet, mediante el acceso controlado al sistema y bajo las pertinentes medidas de seguridad.

A continuación se van a mostrar las características de BioMEN teniendo en cuenta los requerimientos funcionales marcados en el Capítulo II.

## 2.1. Control de usuarios

La existencia de distintos tipos de usuarios en el centro crea la necesidad de realizar distintos perfiles de usuarios en el sistema. Como se ha comentado con anterioridad, el control de los usuarios y las operaciones que estos pueden realizar en el sistema es importante para controlar la calidad de los datos.

Teniendo en cuenta estas necesidades, en BioMEN se ha establecido un conjunto de identificadores de operación. Cada usuario tiene un conjunto de identificadores que indican las operaciones permitidas en el sistema. Cuando el usuario accede al sistema mediante su nombre de usuario y clave, el sistema le presenta solamente las operaciones que puede ejecutar.

Veamos a continuación, un conjunto de las principales operaciones permitidas en BioMEN (ver Figura 43):

1. **Introducir nuevas muestras:** mediante este permiso se facilita al usuario introducir nuevos especímenes en el sistema.
2. **Introducir Revisiones:** el investigador que estudia el material prestado realiza revisiones nomenclaturales sobre el/los espécimen(es). Estas revisiones tienen que incorporarse en el sistema ya que forman parte importante de la información del espécimen.
3. **Introducir datos sobre Nomenclatura, Provincias o Países:** permisos que establecen valores por los que se filtra la información

introducida por el usuario. Mediante estos filtros el sistema se asegura que los datos tienen una calidad mínima.

4. **Modificación de los datos de un espécimen:** En BioMEN se distingue entre las operaciones de alta y de modificación. Los operadores de datos pueden introducir nueva información pero no modificar datos de muestras ya incorporados en el sistema.
5. **Gestión Multimedia:** permiso para permitir a un usuario gestionar las imágenes, videos y sonidos asociados a una muestra.
6. **Gestión de etiquetas:** operación que permite acceder al módulo de impresión de etiquetas.
7. **Consulta de datos sobre Nomenclatura, Provincias o Países:** si deseamos dar la posibilidad de consultar los datos por los que se puede filtrar la información.
8. **Consulta individual de un espécimen:** operación que permite acceder a los datos almacenados para una muestra.
9. **Búsqueda Avanzada:** operación que permite acceder al módulo de consulta avanzada desde el cual se pueden generar consultas de diversos tipos que posteriormente se comentarán.
10. **Gestión de Préstamos:** operaciones para permitir la creación, consulta, reclamación y devolución de los préstamos.
11. **Administración de usuarios:** establece la posibilidad de acceder y gestionar los permisos establecidos para los distintos usuarios.
12. **Control de Sistema Multiagentes:** permiso para habilitar o deshabilitar la actuación de los agentes en el sistema.

**13. Acceso a la información producida por el Sistema Multiagente:**  
permiso para poder consultar los resultados e informes que obtiene el sistema multi- agente.

Permiso	Activo/Inactivo
Introducir Revisiones	Si
Introducir Generos, Provincias o Países	Si
Gestión de Etiquetas	Si
Gestión de Especificaciones Asignadas	Si
Consulta de Generos	Si
Consulta de Provincias	Si
Busqueda Avanzada	Si
Mostrar Mapas	Si
Solicitar Préstamo	Si
Creación de Préstamos	Si
Consultar Préstamos pendientes	Si
Reclamación de Préstamos	Si
Devolución de Préstamos	Si
Generar Exportaciones	Si
Administración de usuarios	Si
Migración de Base de Datos	Si
Ver Datos Personales	Si
Modificar Noticias	Si
Control de Agentes	Si
Consultar Resultados de los Agentes	Si
Introducir Muestra o Adicion	Si
Modificación de Muestras	Si
Gestión de Imágenes de Muestras	Si
Consulta Individual de Muestra	Si
Gestión Bibliografica	Si
Consultar Bibliografía	Si

Figura 43. Lista de permisos en BioMEN

Como se puede observar existen muchos tipos de permisos, lo que hace que el sistema se pueda adaptar a las necesidades de los distintos tipos de usuarios existentes en el centro. La construcción modular del sistema permite dar de alta nuevos permisos y asociarlos con las páginas Web que gestionan dicha operación.

Cuando un usuario accede al sistema mediante la utilización de un nombre de usuario (login) y una clave (*password*), la interfaz del usuario se adapta a las posibilidades que tiene éste en el sistema.

BioMEN es un sistema gestionado a través de la Web, lo que impone la necesidad de incorporar medidas de seguridad para no permitir el acceso al sistema de usuarios no identificados. Toda la información intercambiada entre el usuario y el sistema es cifrada para que no pueda ser filtrada.

Una vez que el usuario proporciona su nombre de usuario y clave el sistema realiza los siguientes pasos:

1. Comprobar si existe el usuario en el sistema y si la clave proporcionada es correcta.
2. Obtener las operaciones permitidas para el usuario.
3. Construir la interfaz a partir de las operaciones permitidas y las preferencias del usuario.

Los servidores Web no permiten almacenar el estado de las peticiones del usuario, pero existen dos tecnologías independientes de desarrollo de aplicaciones que lo permiten: el uso de galletas (*cookies*) y el uso de sesiones.

- El sistema usa un conjunto de datos almacenados en las *cookies* del cliente para identificar al usuario, almacenar el estado de las peticiones, etc. Las *cookies* usadas por el sistema son almacenadas en la parte del cliente, lo que hace que un usuario mal intencionado pueda acceder a esta información.
- El *uso de sesiones* permite realizar las mismas operaciones que con el uso de *cookies*, pero mejoran la seguridad del sistema. Las sesiones son transferidas entre el sistema y el navegador del usuario sin necesidad de estar almacenadas físicamente en la parte del cliente.

BioMEN hace uso de sesiones para almacenar el estado del usuario mientras que éste accede y navega por el sistema. La navegación y el intercambio de información entre el usuario y BioMEN, se realiza de forma segura. Para ello, BioMEN incorpora un sistema de cifrado **SSL (Secure Socket Layer) de 128 bits**. La información es cifrada antes de ser transmitida, impidiendo que se pueda obtener el nombre de usuario y clave mediante una escucha de la información transmitida por la red.

## 2.2. Gestión de especímenes

Mediante este módulo BioMEN permite almacenar todos los datos que serán útiles para posteriores estudios y que son importantes para los investigadores.

Existen distintos niveles o grupos de datos que se pueden almacenar dentro de este módulo:

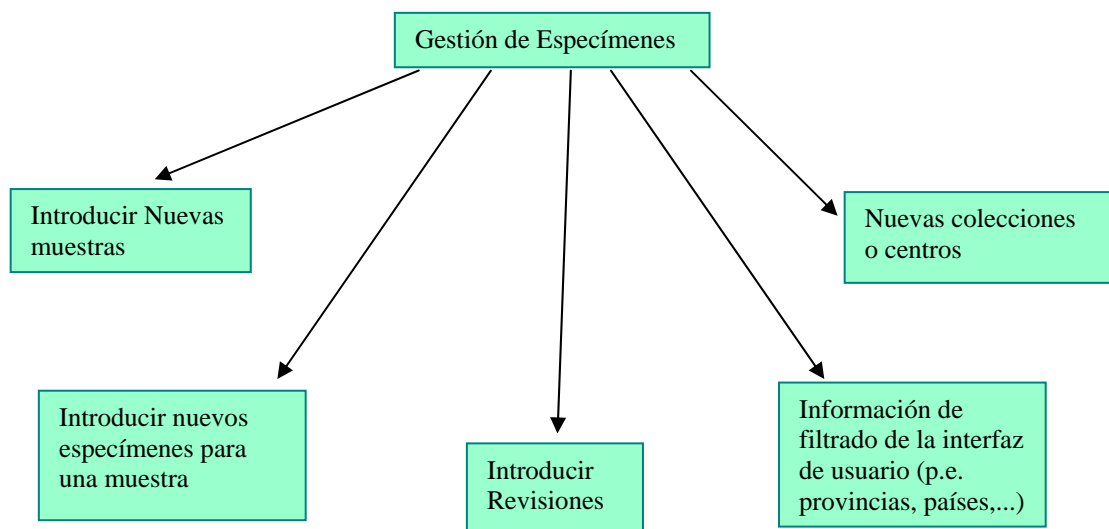


Figura 44. Tipos de información en el módulo de especímenes.

En la Figura 44 se pueden observar la descomposición del módulo de gestión de especímenes realizada en BioMEN:

1. **Introducir nuevas muestras:** desde este módulo se podrá introducir la información necesaria para registrar un nuevo espécimen en el sistema. En el proceso de alta de los datos existen filtros de los datos proporcionados por el usuario para mejorar la calidad de los datos y evitar errores mecanográficos (ver Figura 45).

### Tipo de Operación

Ha solicitado introducir una nueva Adición N° 2 para el Pliego N°: 1  
 - El Género introducido no existe en la BD.  
 - La provincia introducida y el país no están correctamente asociados.

Campos	Valor
Genero	Abia
País	MARRUECOS
Provincia	-

Figura 45. Filtrado de datos en BioMEN

- Introducir Adiciones:** En ocasiones se detecta más de un espécimen en la muestra: por ser similar al primer espécimen determinado o por que se traten de distintas fases de la misma especie. La conjunción del identificador de la muestra más el identificador de la adición forman un identificador único del espécimen dentro del sistema.

Campo	Valor
Nombre del Herbario	RPP
Colección	Fanerogamia
N° Pliego	1
Adición	1
D E T E R M I N A C I Ó N	
Familia	Pinaceae
Género	Abies
Especie	marroccana
Aut_Espec	Trabut
Infrank	
Infra	
Aut_Infra	
Idqual	
Determ	Font Quer
Camisa	S
Mesaj	f
Tipo	
Observación	

País	MARRUECOS
Provincia	-
Localidad	Monte Tisuka, supra Xauen.
Utm	
Latitud	
Longitud	
Altura	1400
Habitat	
Habitat	21-V-1928
Colector	Font Quer
N° Colector	
Num. Pliegos	1
Preparación	
Exsicc	Her Maroccanum
N° Exsicc	16
Observación	

Figura 46. Extracto de información para un espécimen

- Introducir Revisiones:** desde este módulo se pueden introducir las revisiones realizadas por los investigadores que han estudiado el espécimen y han realizado algún tipo de modificación, corrección o confirmación taxonómica.
- Información de filtrado:** datos que utiliza el sistema para filtrar la información proporcionada por el usuario.



5. **Nuevas Colecciones o Centros:** el sistema permite gestionar de forma independiente distintas colecciones y a su vez distintos centros (ver Figura 47).

**Nuevo Centro**

Introduzca los datos para el nuevo Centro

Campos	Valores
Código de la Institución	<input type="text"/>
Nombre Completo	<input type="text"/>
Dirección	<input type="text"/>
Población	<input type="text"/>
Provincia	Sin especificar
País	Sin especificar
Código Postal	<input type="text"/>
Teléfono	<input type="text"/>
Email	<input type="text"/>
Activo	<input type="checkbox"/>

Figura 47. Datos para un nuevo centro

Las tareas asociadas a la introducción de datos en el sistema se encuentran separadas en distintas operaciones para que el administrador del sistema pueda otorgar permisos, dependiendo del perfil de usuario.

Una vez que los datos del espécimen son almacenados en el sistema, se procederá a su etiquetado final y conservación para posteriores estudios.

### 2.3. Generación de Etiquetas

Una vez que la información está almacenada en el sistema, BioMEN permite generar las etiquetas entre dos identificadores dados. El sistema emite un informe en formato PDF (*Portable Document Format*) listo para ser impreso y colocado en las muestras.

Las etiquetas emitidas incorporan un código de barras para la identificación de la muestra de una forma unívoca y automatizada, mediante el uso de lectores de códigos de barra.

El código de barras es especialmente útil para la realización de préstamos. Permite dar de alta un préstamo con el lector de códigos de barra indicando así las muestras que forman parte del mismo.

El **código de barras** es la representación de una determinada información mediante un conjunto de líneas paralelas verticales de diferente grosor y espaciado. La correspondencia o mapeo entre los mensajes que representan y el código de barras se denomina *simbología*. Las simbologías pueden ser clasificadas en dos grupos, atendiendo a dos criterios diferentes:

1. **Continuo frente a discreto** (ver Tabla 6): los caracteres en las simbologías continuas, comienzan con un espacio y el siguiente comienza con una barra, o viceversa. Los caracteres en las simbologías discretas comienzan y terminan con barras; el espacio entre caracteres es ignorado, para no ser interpretado como el final del código.
2. **Bidimensional frente a multidimensional** (ver Tabla 7): las barras en las simbologías bidimensionales son anchas o estrechas; cuánto sean de anchas, no importa y pueden variar de un carácter al siguiente. Las barras, en las simbologías multidimensionales, son múltiplos de una anchura llamada X; generalmente, se emplean barras con anchura X, 2X, 3X, y 4X.

Simbología	Cont/Disc	Usos
Plessey	Continuo	Catálogos, estanterías, inventarios
EAN		Uso para productos alimentarios
Codabar	Discreto	Bibliotecas, bancos de sangre, billetes de avión
Código 39	Discreto	Variado
Código 93	Continuo	Variado
Código 128	Continuo	Variado

Tabla 6. Códigos de barras de una dimensión más representativos.

Simbología	Notas
3-DI	Desarrollado por Lynn Ltd.
ArrayTag	De ArrayTech Systems.
Ojo de buey( <i>bullseye</i> )	Este fue el código de barras probado un almacén Kroger en Cincinnati, EE.UU. Utiliza barras concéntricas.
Código 1	Dominio público.
Código 16K	Basado en Código 128.
Código 49	Códigos de barras 1D apilados de Intermec Corp.
Código CP	De CP Tron, Inc.
Data Matrix	De RSVI Acuity CiMatrix.

Datastrip	De Datastrip, Inc.
PDF417	El código 2D más común. Dominio público.
SmartCode	De InfoImaging Technologies.
Snowflake	De Marconi Data Systems, Inc.
SuperCódigo	Dominio público.
UltraCódigo	Versiones en blanco y negro y en color. Dominio público.

Tabla 7. Códigos de barras multidimensionales

La existencia de dos tipos de códigos de barra se explica por la necesidad de representar y almacenar caracteres de distinta longitud. Mientras que los códigos de barras unidimensionales tienen una capacidad que oscila entorno a los 15 caracteres, según sea el ancho del lector de códigos de barras, el código bidimensional permite representar hasta los 2800 caracteres, en el caso del *PDF417*.

Hasta el momento, en todos los sistemas analizados, sólo se han utilizado códigos de barras de una dimensión, puesto que las necesidades para representar el identificador de la muestra son mínimas. No obstante, la utilización de códigos de barras multidimensionales permite almacenar toda la historia de la muestra en una sola etiqueta. El código podrá ser leído en cualquier momento recuperando así toda la información asociada a la muestra.

Por ello, BioMEN, en su módulo de etiquetado permite:

- Utilización de **códigos de barras de una dimensión** para la generación de etiquetas inteligentes.
- Generación de etiquetas de **barras bidimensionales** que almacenan toda la información de la muestra.

De entre todas las alternativas, para el código de barras unidimensional, se ha elegido el **CÓDIGO 39**, por ser el más utilizado y soportado por, prácticamente, todos los lectores de códigos de barras. Además, permite la codificación de caracteres alfanuméricos (ver Figura 48).

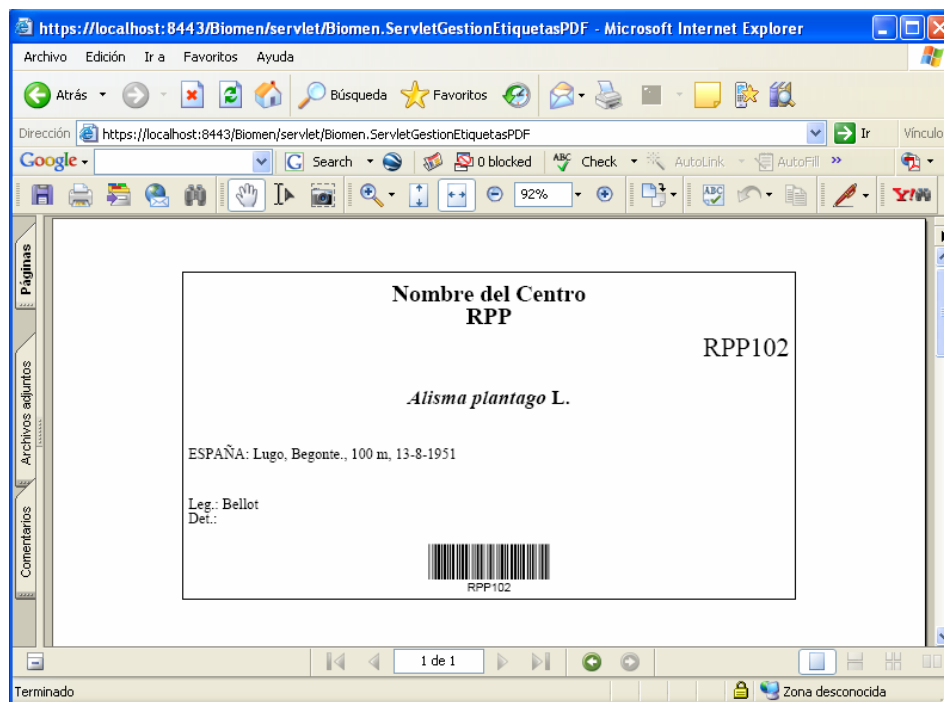


Figura 48. Etiqueta con código de barras en BioMEN

Por otro lado, se ha elegido el código PDF417, como código bidimensional, por ser un código de dominio público, estar estandarizado y por ser el código que más caracteres permite representar.

BioMEN permite representar en un código bidimensional la información almacenada para una muestra en un formato conocido. Toda la información almacenada en el código de barras debe de estar estructurada para que se puedan distinguir los distintos campos que la forman.

Por ello, BioMEN estructura la información en el estándar Darwin Core, promocionado por GBIF, y genera el código de barras PDF417 equivalente (ver Figura 49). Esto permite:

1. Representar la información mediante un estándar consensuado por toda la comunidad científica.
2. Puede ser tratado por cualquier *software* que permita la importación en el estándar Darwin Core.

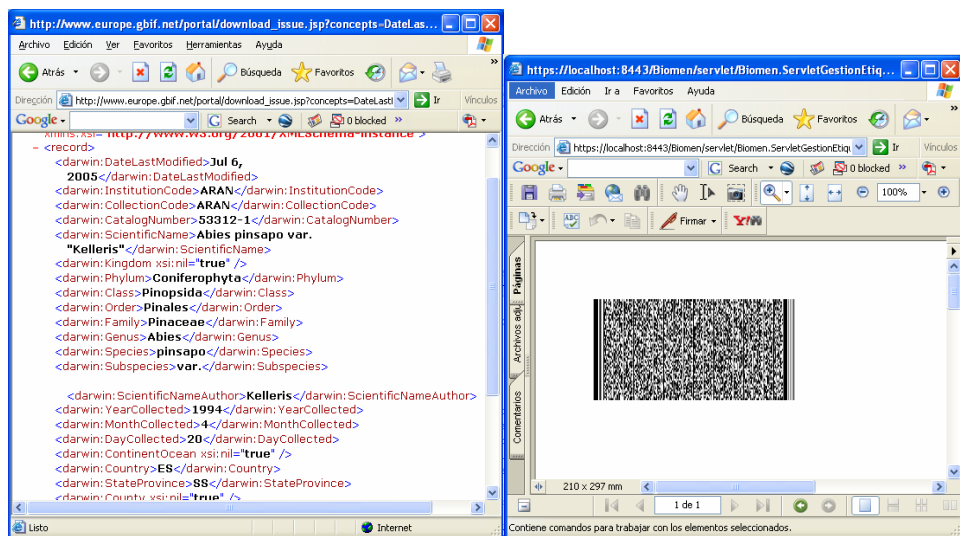


Figura 49. Equivalencia entre Darwin Core y código bidimensional PDF417 en BioMEN

Esta es una característica única y que permite recuperar la información existente para una muestra de forma automática, rápida y efectiva.

## 2.4. Consultas

El acceso a los datos almacenados en el sistema es un apartado importante en cualquier desarrollo. Los sistemas hasta el momento evaluados han permitido, en mayor o menor grado, la consulta de los datos almacenados.

En BioMEN se ha dividido el módulo de consulta en tres grupos (ver Figura 50):

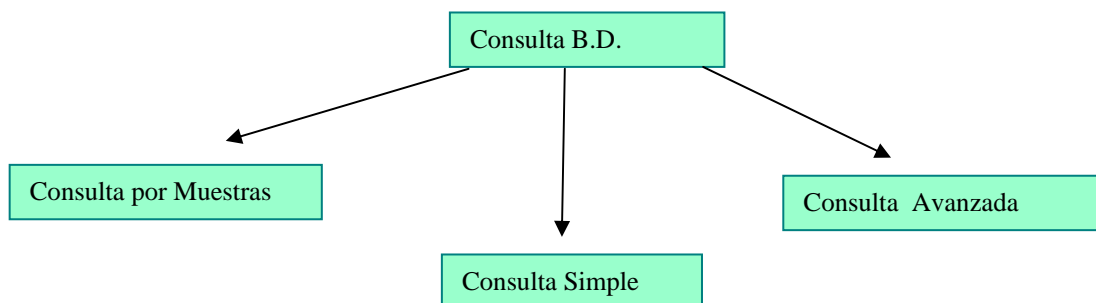


Figura 50. Tres submódulos de consulta en BioMEN.

- **Consulta por muestra:** permite acceder a toda la información asociada a una muestra dada.

- **Consulta Simple:** el sistema consulta para un nombre científico dado.
- **Consulta Avanzada:** permite al usuario especificar cualquier condición y seleccionar los campos deseados para el resultado de la consulta. Todo el proceso de consulta se realiza por la Web y es posible obtener el resultado de la consulta en diversos formatos (PDF, Hoja de Cálculo Excel, ...).

La **consulta por muestra** permite acceder a toda la historia de la muestra almacenado en el sistema (información taxonómica, determinación, localización, ...) (ver Figura 51).

**Adición1**  
Datos de la Determinación

Campos	Revisión Número 1
Familia	Alismataceae
Género	Alisma
Especie	plantago
Aut_Espec	L.
Infrank	
Infra	
Aut_Infra	
Idqual	
Determ	
Camisa	S
Mesan	/
Tipo	
Observación	

**Consultas**  
Consulta de identificadores

Campos	Valores
Número de identificador	<input type="text"/>
Nombre centro	GDA

Restablecer Enviar

Campos	Etiqueta Original
País	ESP
Provincia	LU
Localidad	Lugo, Begonte.
Utm	

Figura 51. Fragmento de consulta por muestra

Por medio de la consulta por muestra se accede a la información multimedia, estado de la muestra (prestada o no) y la posibilidad de mostrar de forma gráfica y sobre un mapa el lugar de recolección. Posteriormente mostraremos los detalles de representación mediante servicios externos de Sistemas de Información geográfica.

La **consulta simple** permite consultar los especímenes existentes en el sistema bajo un nombre científico (ver Figura 52). Es una consulta simple y sencilla que permite al usuario obtener un listado de las muestras que existen en el centro de una forma rápida y ágil.



Figura 52. Consulta simple en BioMEN

El último de los submódulos desarrollados para BioMEN es el de la **consulta avanzada**. Desde este módulo el usuario puede consultar cualquier dato almacenado en el sistema. Además, puede configurar el aspecto de salida de la consulta.

### Búsqueda Avanzada

Establezca las condiciones de la búsqueda.

Enlace	Campo	Condición	Valor
	País	Igual	ESP
Y	Provincia	Igual	GR
Y	Número de Revisión	CONDICION	
Y	Número de Revisión	CONDICION	

Mostrar Resultados

Restablecer    Enviar

Figura 53. Paso 1: Estableciendo condiciones de consulta

### Búsqueda Avanzada

Parámetros para la búsqueda avanzada.  
Establezca las columnas para el resultado de la búsqueda.

Columnas			
Familia	Género	Especie	Localidad
Columna Resultado	Columna Resultado	Columna Resultado	Columna Resultado
Columna Resultado	Columna Resultado	Columna Resultado	Columna Resultado

Recibir el número de táxones.  
 Agrupar los resultados  
 Ver con Google Earth  
 No desea introducir más columnas para el resultado de la consulta.

Restablecer    Enviar

Figura 54. Paso 2: selección de columnas de resultado.

### Resultado de la búsqueda avanzada

Resultado de la búsqueda  
Total resultados: 64











FAMILIA	GENERO	ESPECIE	LOCALIDAD	Ver En Mapa
Pinaceae	Cedrus	deodora	Granada, Lanjarón, Tello.	
Pinaceae	Cedrus	deodora	Granada, Lanjarón, Tello.	
Aphylliphorales	Clavaria	stricta	Granada, Lanjarón, Tello.	
Pinaceae	Cedrus	libani	Granada, Lanjarón, Tello.	
Pinaceae	Cedrus	libani	Granada, Lanjarón, Tello.	
Aphylliphorales	Clavaria	stricta	Granada, Lanjarón, Tello.	
Pinaceae	Pinus	amiltroni	Granada, Lanjarón, Tello.	
Aphylliphorales	Hymenochaete	cinnamomea	Granada, Lanjarón, Tello.	
Pinaceae	Pinus	baciano	Granada, Lanjarón, Tello.	
Rhytismatales	Hypoderma	nevisequium	Granada, Lanjarón, Tello.	
Pinaceae	Pinus	canariensis	Granada, Lanjarón, Tello.	

Figura 55. Resultado de consulta avanzada

Una vez establecidas las condiciones de búsqueda (Figura 53), el usuario selecciona las columnas de resultado (Figura 54) y el tipo de resultado que requiere. Como se puede observar, por las figuras anteriores, las consultas se pueden mostrar en mapas mediante la herramienta *Google Earth* (Figura 55).

El usuario puede solicitar en la consulta:

- El número de taxa que cumplen las condiciones establecidas.
- Las últimas revisiones nomenclaturales de la consulta realizada.
- Enlazar con un visualizador de información geográfica.

Una vez que la consulta es obtenida por el sistema, BioMEN permite:

- Refinar la búsqueda (ver Figura 56): realizar una consulta sobre los resultados ya obtenidos.
- Obtener la consulta en distintos formatos: PDF, Hoja de cálculo Excel, ...
- Realizar un préstamo a partir de los resultados obtenidos por la consulta (sólo para aquellos usuarios con permisos para ello).



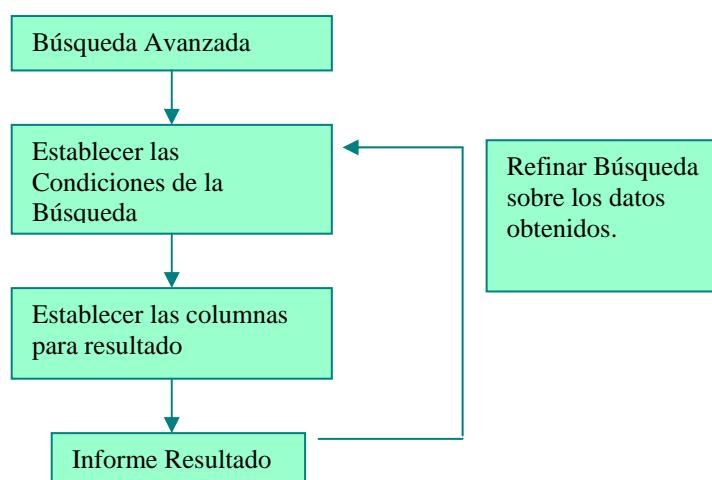


Figura 56. Estructura de búsqueda y refinamiento de la consulta

## 2.5. Préstamos de material

El material del centro es prestado a otros investigadores y centros. Por ello, BioMEN incorpora las herramientas necesarias para llevar a cabo el control de los préstamos a otras instituciones.

En este módulo, existen tres submódulos que permiten gestionar los préstamos: alta, consultar préstamos y devolución. Veamos cada uno de ellos con más detalle.

### *Alta de Préstamos*

El alta de un préstamo mediante BioMEN se puede realizar por tres vías:

1. Mediante el uso de lector de código de barras. Permite al gestor del centro registrar el material que forma el préstamo.
2. A través de una consulta avanzada en el sistema. A partir del listado de material solicitado por un investigado, el personal autorizado puede realizar una consulta. Los resultados de la misma, son el punto de partida del proceso de préstamo.

3. El usuario del sistema puede solicitar el préstamo desde el apartado de búsqueda avanzada. El administrador del sistema da permisos a los usuarios para solicitar préstamos y éste selecciona el material que desea. El sistema informa a los gestores de préstamos de forma totalmente automática y emitiendo alertas a través del Email para que se realice el préstamo solicitado por el usuario.

Desde este módulo se introduce la información necesaria para recoger adecuadamente los préstamos. El diagrama del módulo es el siguiente:

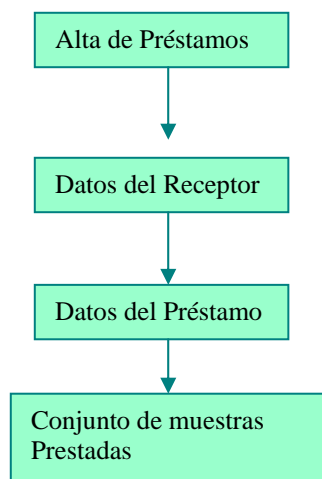


Figura 57. Proceso de alta de un préstamo.

Los datos del Receptor recogen la información necesaria para saber a quién se envía el préstamo. Los datos del Préstamo reflejan las características del préstamo, tales como; observaciones, fecha del préstamo, periodo de validez del mismo, etc. Por último, el conjunto de muestras prestadas.

### ***Consulta de Préstamos***

La consulta de los préstamos realizados en el centro permite controlar el estado de los mismos. Se pueden distinguir de dos tipos de consultas de préstamos:

- **Listado global de Préstamos:** Este listado muestra de una forma enumerada el conjunto de préstamos que el centro tiene pendientes de

devolución. Indica el número total de ejemplares prestados en cada préstamo, así como los datos de la entidad receptora.

- **Listado individual de Préstamo:** opción que detalla la información de un préstamo. Se obtiene el conjunto de muestras que han sido enviadas en dicho préstamo, pudiendo observar la naturaleza del préstamo (grupo de géneros enviados).

### ***Devolución de Préstamos***

Los préstamos que se envían a las correspondientes entidades o instituciones deben ser devueltos tras cumplir su periodo de validez.

Para realizar la devolución del préstamo se introduce el identificador del mismo y el sistema proporciona al usuario autorizado el conjunto de material que forma parte del préstamo.

El usuario tiene que comprobar la devolución de cada una de las muestras que forman parte del préstamo, ya que no siempre se devuelve completamente. Por ello, el sistema tiene que incorporar los mecanismos necesarios para realizar devoluciones parciales.

El préstamo de material es una de las principales tareas que permiten intercambiar información, pero implica un deterioro del material. Por ello, sería interesante realizar cualquier avance que reduzca el número de materiales prestados. En este aspecto, es fundamental la transferencia e intercambio de información multimedia.

## **2.6. Información multimedia**

Debido al carácter del centro como servicio de documentación es de gran utilidad proporcionar información multimedia del espécimen. La información multimedia puede disminuir el número de préstamos, por ejemplo al permitir que el investigador acceda de forma visual al espécimen.

Mediante esta propuesta se solucionan algunos de los problemas que son importantes considerar y evaluar:

- Deterioro del Material.
- El investigador no tiene que desplazarse al centro para poder observar el espécimen.

Por estos motivos, BioMEN asocia cualquier tipo de información multimedia a una muestra (Figura 58, Figura 59). Esta información multimedia puede ser consultada por el módulo de consulta y por los servicios externos, que posteriormente se comentarán.

### Asociación de multimedia a un identificador

Rellene el formulario

Campo	Origen Del Multimedia
Número de identificador	<input type="text"/>
Nombre del centro	RPP <input type="button" value="v"/>
Dirección de la Archivo	<input type="text"/> <input type="button" value="Examinar..."/>
Operación	Añadir Multimedia <input type="button" value="v"/>

Figura 58. Asociación de información multimedia a una muestra

### Consulta de muestras

Identificador	Multimedia	Eliminar
Número de identificador: 102 del centro RPP Total Adiciones: 1		

Figura 59. Acceso a la información multimedia desde el módulo de consulta

## 2.7. Referencias bibliográficas

BioMEN permite gestionar la información bibliográfica utilizada para la asignación de un nombre científico.

Para cada espécimen es posible asociar más de una referencia bibliográfica (ver Figura 60), permitiendo a su vez asociar una misma referencia a varios especímenes.

### Asignación Bibliografía

Asociar la bibliografía utilizada para la determinación de un espécimen.

Campos		Valores
Identificador del Especimen		1-1

Autor	Título
<input type="checkbox"/> R. Perez-Perez	A MultiAgent System to perform

Bibliografía	
Cadena de búsqueda	<input type="text"/>

Figura 60. Asignación de bibliografía a una muestra

Permite además gestionar la Bibliografía del sistema de forma independiente, esté o no, asociada o no a un espécimen.

Este módulo se gestiona de forma independiente y BioMEN permite establecer los permisos correspondientes para limitar el acceso de los usuarios al mismo.

Una vez que las referencias son asociadas a una muestra, BioMEN permite consultar la bibliografía cuando se consulta una muestra (ver Figura 61).

### Consultas

Consulta de Contenedores

Campos	Valores
Número de Contenedor	RPP19912
Nombre Centro	RPP

### Bibliografía asociada

Identificador	Título	Detalles
123	Veronica fruticulosa L. en los Pirineos orientales	Ver
231	Contribuciones al conocimiento de la flora Cantábrica, II	Ver
234	Contribuciones al conocimiento de la flora de Asturias, II	Ver

Figura 61. Captura del módulo de bibliografía de BioMEN

## 2.8. Información taxonómica

Para mejorar la calidad de los datos almacenados en el sistema, BioMEN gestiona información taxonómica (ver Figura 62). Ésta es utilizada para filtrar los datos proporcionados por los usuarios evitando así errores mecanográficos.

### Nuevo Género

Introduzca los datos para el nuevo género

Campos	Valores
Grupo	<input type="text"/>
Familia	<input type="text"/>
Género	<input type="text"/>

Figura 62. Plantilla para un nuevo género en BioMEN

Desde este módulo, BioMEN permite gestionar nuevos taxones que actuarán de filtro para los datos proporcionados por los usuarios.

La información taxonómica asociada a este módulo está gestionada por un conjunto de operaciones que no todos los usuarios pueden ejecutar.

Este módulo es tratado con más detalle en el capítulo siguiente cuando tratemos la integración de fuentes heterogéneas y la comunicación con proyectos externos.

El uso de proyectos taxonómicos para el soporte y filtrado de la información existente en las bases de datos del sistema es importante, puesto que puede mantener actualizada la información existente en el mismo.

## 2.9. Servicios Externos

Con BioMEN se ha acentuado, sobre todo, el tratamiento virtual del centro y el acceso deslocalizado al mismo.

Toda la información contenida en el sistema es de utilidad para la comunidad científica. No obstante, existe información que puede ser perjudicial para la conservación de las especies. Es por ello que cierta información, como pueden ser, los datos espaciales, no son mostrados a todos los usuarios externos del centro. La información no puede ser facilitada en su totalidad, ya que podría ser utilizada mal intencionadamente.

Desde el módulo de servicios externos se han dado distintas soluciones para satisfacer todas las necesidades existentes según los distintos tipos de usuarios e investigadores.

Los mecanismos de consulta externa habilitados por BioMEN son los siguientes:

- Consulta de las especies existentes en el sistema.
- Consulta de los especímenes existentes para un *taxon* dado.
- Un conjunto de Servicios Web que permiten enlazar, de forma automatizada y por aplicaciones, las consultas al sistema.

La consulta al sistema se puede realizar mediante dos alternativas:

1. Mediante la **petición http** a un punto de acceso del sistema. El paso de parámetros permite filtrar el resultado de la consulta de los usuarios.

2. Mediante el uso de **Servicios Web**. Se ha incorporado a BioMEN la tecnología necesaria para proporcionar servicios Web. El uso de este tipo de servicios permite enlazar BioMEN de forma automática con otros sistemas.

Veamos cada una de estas alternativas.

En primer lugar examinaremos las alternativas desarrolladas para las peticiones *http*. Éstas se pueden realizar a dos puntos de acceso incorporados en BioMEN:

1. **Consulta de especies existentes:** El primer punto de acceso externo al sistema permite consultar las distintas especies existentes teniendo en cuenta los siguientes parámetros:
  - i. *Taxon*: subcadena que filtra el resultado por el nombre científico a buscar.
  - ii. País: cadena que especifica el nombre del país para el cual se desean filtrar los resultados.
  - iii. Provincia: cadena que filtra los resultados por la provincia en la que se ha recolectado el espécimen.
2. **Consulta de los especímenes:** informe de los especímenes asociados a un *taxon* dado. Permite adaptar la respuesta a un formato e interfaz especificada:
  - i. *Taxon*: subcadena que filtra el resultado por el nombre científico asociado al espécimen.
  - ii. XSLT (*Extensible Stylesheet Language Transformation*): mediante este parámetro el usuario puede transformar el resultado de la consulta en su propio formato. Esto permite integrar a BioMEN al aspecto de cualquier otro sistema.



- iii. Formato: parámetro que especifica el formato de salida de la consulta. Este parámetro puede tomar valores como HTML, XML, Excel o PDF.



Figura 63. Esquema de petición http a través del navegador Web

Como podemos ver en el diagrama la consulta externa permite obtener información de suficiente interés y permite una búsqueda adecuada. Mediante esta estructura de consulta permite responder a una serie de preguntas, tales como: ¿qué especímenes se encuentran en una determinada provincia? ¿En qué lugar se encuentra un *taxon* específico? ¿Existe información multimedia? etc.

La información proporcionada al usuario está formada por el siguiente esquema de intercambio:

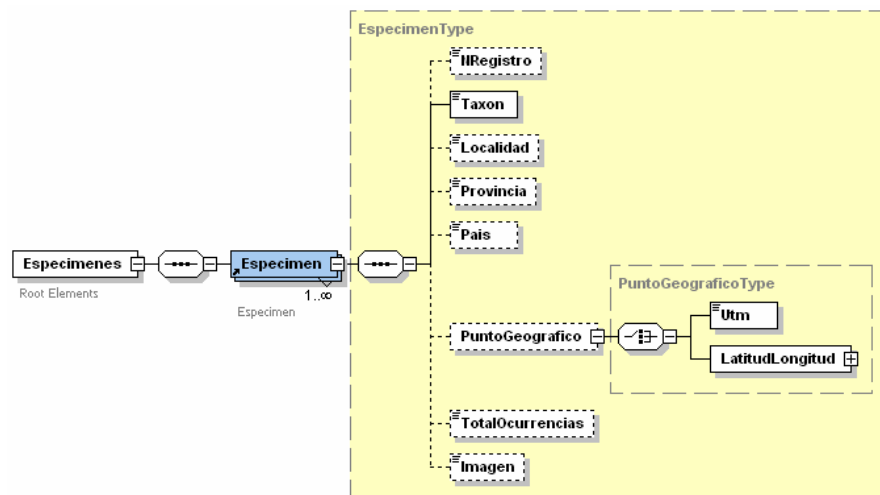


Figura 64. Esquema propio para las consultas externas

Para el resultado de las consultas externas hemos generado un esquema que estructura la respuesta del sistema. Se ha definido un *XML Schema* que define la estructura de la respuesta de la consulta en uno o más ítems del tipo *Espécimen*.

La información posible para un ítem del tipo espécimen es la siguiente (ver Figura 64):

- **Nregistro:** número de registro de la muestra en el sistema. Este identifica de forma unívoca la muestra dentro del sistema. De esta forma un usuario externo puede solicitar el préstamo de una muestra concreta al centro consultado.
- **Taxon:** nombre científico con el que fue determinado el espécimen.
- **Localidad:** descripción de la localidad donde se recogió el espécimen.
- **Provincia:** especifica el nombre de la provincia donde se recolectó la muestra.
- **País:** especifica el nombre del país donde se recolectó la muestra.
- **PuntoGeográfico:** especifica las coordenadas latitud/longitud de recolección del espécimen. En las consultas externas se reduce la precisión para no proporcionar la posición exacta del espécimen.
- **TotalOurrencias:** nos indica el número de ocurrencias que existen en la base de datos y que verifican las condiciones establecidas por el usuario.
- **Imagen:** es la URL que accede a la información multimedia asociada a la muestra.

Todos estos campos permiten al usuario obtener los datos necesarios para tener información de qué es lo que existe en el centro y si existe información multimedia asociada al mismo puede también consultarla.

Se ha desarrollado un estándar propio puesto que la información necesaria para la respuesta a las consultas es un fragmento muy pequeño de las posibilidades que ofrecen los estándares como Darwin Core o ABCD Schema.

Dado que BioMEN es gestionado a través de la Web, si el investigador necesitara más información de la que se proporciona por las consultas externas, se puede crear un usuario y dar los permisos necesarios para la consulta.

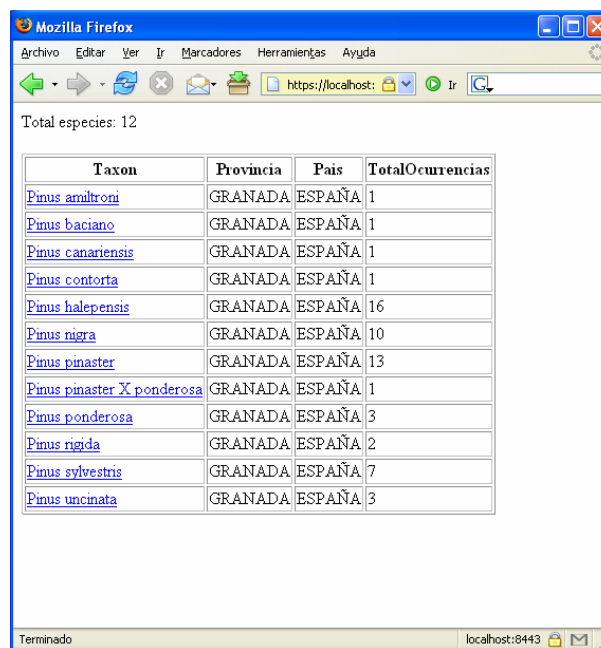
Veamos dos ejemplos de consulta externa mediante petición http:

1. Listado de especies: listado de especies en la provincia de Granada y cuya cadena de *taxon* comience por "*Pinus*".

a. Petición:

<https://localhost:8443/Biomen/servlet/Biomen.ServletConsultaExterna?taxon=Pinus&provincia=GRANADA>

- b. Respuesta (los datos de prueba que se muestran son parte de las bases de datos del Herbario de la Universidad de Granada (Herbario de la Universidad de Granada, 2006)):



Taxon	Provincia	Pais	TotalOcurrencias
<a href="#">Pinus amiltoni</a>	GRANADA	ESPAÑA	1
<a href="#">Pinus baciano</a>	GRANADA	ESPAÑA	1
<a href="#">Pinus canariensis</a>	GRANADA	ESPAÑA	1
<a href="#">Pinus contorta</a>	GRANADA	ESPAÑA	1
<a href="#">Pinus halepensis</a>	GRANADA	ESPAÑA	16
<a href="#">Pinus nigra</a>	GRANADA	ESPAÑA	10
<a href="#">Pinus pinaster</a>	GRANADA	ESPAÑA	13
<a href="#">Pinus pinaster X ponderosa</a>	GRANADA	ESPAÑA	1
<a href="#">Pinus ponderosa</a>	GRANADA	ESPAÑA	3
<a href="#">Pinus rigida</a>	GRANADA	ESPAÑA	2
<a href="#">Pinus sylvestris</a>	GRANADA	ESPAÑA	7
<a href="#">Pinus uncinata</a>	GRANADA	ESPAÑA	3

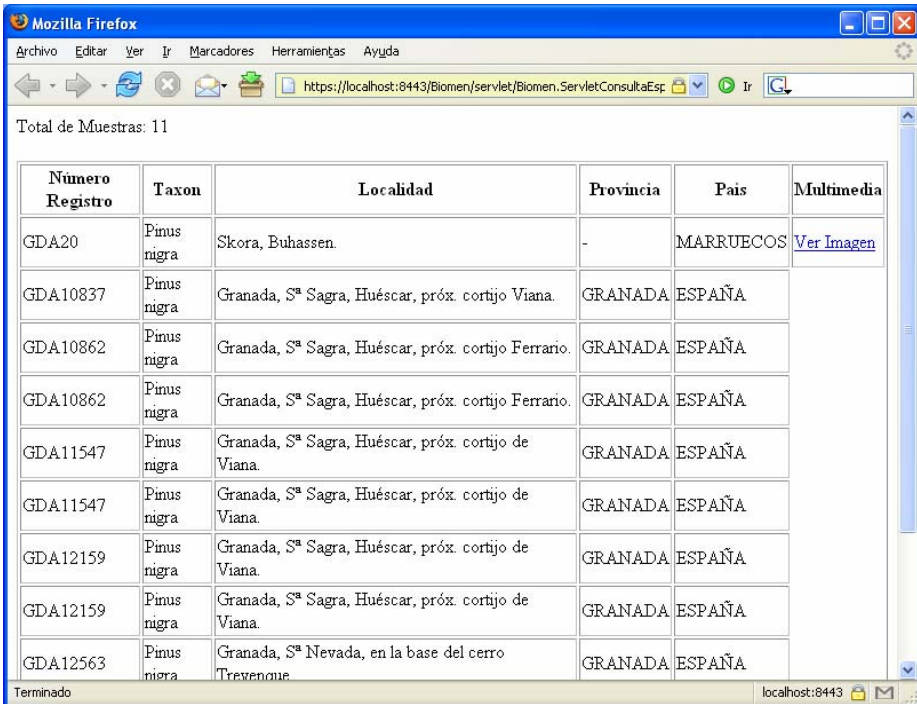
Figura 65. Respuesta a consulta externa para listar especies

2. Listado de los especímenes para la especie *Pinus nigra*:

## a. Petición:

<https://localhost:8443/Biomen/servlet/Biomen.ServletConsultaEspecie?taxon=Pinus%20nigra>

## b. Resultado (Herbario de la Universidad de Granada, 2006):



Total de Muestras: 11

Número Registro	Taxon	Localidad	Provincia	Pais	Multimedia
GDA20	Pinus nigra	Skora, Buhassen.	-	MARRUECOS	<a href="#">Ver Imagen</a>
GDA10837	Pinus nigra	Granada, Sª Sagra, Huéscar, próx. cortijo Viana.	GRANADA	ESPAÑA	
GDA10862	Pinus nigra	Granada, Sª Sagra, Huéscar, próx. cortijo Ferrario.	GRANADA	ESPAÑA	
GDA10862	Pinus nigra	Granada, Sª Sagra, Huéscar, próx. cortijo Ferrario.	GRANADA	ESPAÑA	
GDA11547	Pinus nigra	Granada, Sª Sagra, Huéscar, próx. cortijo de Viana.	GRANADA	ESPAÑA	
GDA11547	Pinus nigra	Granada, Sª Sagra, Huéscar, próx. cortijo de Viana.	GRANADA	ESPAÑA	
GDA12159	Pinus nigra	Granada, Sª Sagra, Huéscar, próx. cortijo de Viana.	GRANADA	ESPAÑA	
GDA12159	Pinus nigra	Granada, Sª Sagra, Huéscar, próx. cortijo de Viana.	GRANADA	ESPAÑA	
GDA12563	Pinus nigra	Granada, Sª Nevada, en la base del cerro Trevenque.	GRANADA	ESPAÑA	

Como podemos observar en esta última consulta, aparece un enlace para aquellas muestras en las que existe información multimedia. El usuario externo puede acceder a ésta y examinar la muestra sin necesidad de solicitar el préstamo.

Por último, en este módulo se han incorporado todos los mecanismos y tecnologías necesarias para habilitar dos **servicios Web** que hacen referencia a cada una de las consultas anteriores pero accesibles con otro canal de información.

### *Qué es un servicio Web*

Un servicio Web es una colección de protocolos y estándares que sirve para intercambiar datos entre aplicaciones. Distintas aplicaciones de *software*

desarrolladas en lenguajes de programación diferentes y ejecutadas sobre cualquier plataforma pueden utilizar los servicios Web para intercambiar datos. La interoperabilidad se consigue mediante la adopción de estándares públicos. Las organizaciones OASIS<sup>8</sup> y W3C<sup>9</sup> son los comités responsables de la arquitectura y reglamentación de los servicios Web. Para mejorar la interoperabilidad entre distintas implementaciones de servicios Web se ha creado el organismo WS-I, encargado de desarrollar diversos perfiles para definir de manera más exhaustiva estos estándares.



Figura 66. Diagrama de flujo de datos en un servicio Web

Para el funcionamiento de los Servicios Web se han empleado los siguientes estándares:

1. Pila de Protocolos de los Servicios Web (*Web Services Protocol Stack*): Así se denomina al conjunto de servicios y protocolos de los servicios Web.

---

<sup>8</sup> OASIS es un acrónimo de *Organization for the Advancement of Structured Information Standards*. Se trata de un consorcio internacional sin ánimo de lucro que orienta, desarrolla y soporta el desarrollo de estándares para empresas.

<sup>9</sup> W3C es un acrónimo de *World Wide Web Consortium*. W3C es una organización internacional que produce estándares para la WWW.

2. XML (*eXtensible Markup Language*): Es el formato estándar para los datos intercambiados.
3. SOAP<sup>10</sup> (*Simple Object Access Protocol*): Protocolo sobre los que se establece el intercambio.
4. WSDL (*Web Services Description Language*): describe la interfaz pública a los servicios Web. Está basado en XML y describe la forma de comunicación con los Servicios Web.
5. UDDI (*Universal Description, Discovery, and Integration*): Protocolo para publicar la información de los servicios Web. Permite a las aplicaciones comprobar qué servicios Web están disponibles.
6. WS-Security (*Web Service Security*): Protocolo de seguridad aceptado como estándar por OASIS. Garantiza la autenticación de los actores y la confidencialidad de los mensajes enviados.

El funcionamiento de los servicios Web hace que distintas aplicaciones se pueden comunicar de una forma eficiente y sin tener en cuenta las necesidades gráficas de los usuarios.

El proveedor de un servicio Web registra su servicio (punto de acceso) en un servidor UDDI. Éste servidor es el encargado de hacer público el servicio y de actuar de mecanismo centralizador para el descubrimiento de nuevos servicios.

Un consumidor de servicios solicita al servidor UDDI los servicios que cubren sus objetivos o necesidades y este le informa del punto de acceso al servicio y de la forma de interactuar con él (mediante mensajes en WSDL).

---

<sup>10</sup> SOAP es un protocolo estándar creado por Microsoft, IBM y otros, está actualmente bajo el control de la W3C. SOAP define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos en XML

Una vez que el consumidor conoce de la existencia del proveedor se realizan intercambios directos de mensajes (en SOAP) para satisfacer el servicio proporcionado.

### ***Ventajas de los servicios Web***

Se pueden enumerar una serie de ventajas de la utilización de servicios Web por parte de los sistemas que desean ofertar un servicio:

- Aportan interoperabilidad entre aplicaciones de *software* independientemente de sus propiedades o de las plataformas sobre las que se instalen.
- Los servicios Web fomentan los estándares y protocolos basados en texto, que hacen más fácil acceder a su contenido y entender su funcionamiento.
- Al apoyarse en HTTP, los servicios Web pueden aprovecharse de los sistemas de seguridad (*firewall*) sin necesidad de cambiar las reglas de filtrado.
- Permiten que servicios y *software* de diferentes compañías ubicadas en diferentes lugares geográficos puedan ser combinados fácilmente para proveer servicios integrados.

### ***Servicios Web en BioMEN***

En BioMEN se ha incorporado el uso de Apache AXIS para incorporar en el sistema las tecnologías necesarias para proporcionar servicios Web.

El acceso a las descripciones de los servicios Web (WSDL) se realiza a través de dos puntos de acceso, uno para cada uno de los servicios Web proporcionados:

- <https://localhost:8443/Biomen/ConsultaEspecie.jws?wsdl> : para la consulta de las especies existentes para un subcadena que represente el *taxon*, una provincia o un país.
- <https://localhost:8443/Biomen/ConsultaEspecimenes.jws?wsdl> : servicio Web que permite obtener un listado de los especímenes almacenados para un *taxon* dado.

Veamos por ejemplo un fragmento del código WSDL y los parámetros necesarios para realizar la petición del Servicio Web de consulta de especímenes:

```

- <wsdl:definitions targetNamespace="https://localhost:8443/Biomen/ConsultaEspecimenes.jws">
- <!--
  WSDL created by Apache Axis version: 1.3
  Built on Oct 05, 2005 (05:23:37 EDT)
-->
- <wsdl:message name="consultaRequest">
  <wsdl:part name="taxon" type="xsd:string"/>
  <wsdl:part name="provincia" type="xsd:string"/>
  <wsdl:part name="pais" type="xsd:string"/>
</wsdl:message>
- <wsdl:message name="consultaResponse">
  <wsdl:part name="consultaReturn" type="xsd:string"/>
</wsdl:message>

```

Figura 67. Mensaje WSDL para el Servicio Web Consulta de Especímenes.

En la figura se puede observar los tres parámetros que recibe el servicio Web (*taxon*, provincia y país) de tipo cadena (*string*) y el resultado del servicio será una cadena (*consultaResponse*).

La existencia de Servicios Web hace que se pueden generar flujos de información entre distintos sistemas que permitan al investigador realizar sus estudios de biodiversidad. Por ejemplo, mediante los servicios Web de BioMEN se obtienen los datos sobre los especímenes existentes para una determinada especie. Esta información es usada a su vez por un sistema de información geográfica que nos permite delimitar las áreas de interés, ...

Ejemplo de petición y respuesta al servicio Web:



```

public class Consulta {
    public static void main(String [] args) throws Exception {
        String endpoint = "http://localhost:8080/Biomen/ConsultaEspecímenes.jws";
        // Make the call
        Service service = new Service();
        Call call = (Call) service.createCall();
        call.setTargetEndpointAddress(new java.net.URL(endpoint));
        call.setOperationName("consulta");
        call.removeAllParameters();
        call.setReturnType(XMLType.XSD_STRING);
        call.addParameter("op1", XMLType.XSD_STRING, ParameterMode.IN);
        call.addParameter("op2", XMLType.XSD_STRING, ParameterMode.IN);
        call.addParameter("op3", XMLType.XSD_STRING, ParameterMode.IN);
        String res = (String) call.invoke( new Object [] { new String("Abies pinsapo"), new
String("MALAGA"),new String("") });
        System.out.println("Especímenes: " + res);
    }
}

```

Figura 68. Comunicación desde Java con los servicios Web de BioMEN

```

Especímenes: <?xml version="1.0"?>
<Especímenes>
  <Especímen><NRegistro><Taxon>Abies pinsapo</Taxon> <Localidad/>
<Provincia>MALAGA</Provincia> <Pais>ESPAÑA</Pais>
<PuntoGeografico><LatitudLongitud><Latitud/><Longitud/></LatitudLongitud></P
untoGeografico> <TotalOcurrencias>12</TotalOcurrencias> <Imagen/>
</Especímen> </Especímenes>

Especies : <?xml version="1.0"?>
<Especímenes>
<Especímen><NRegistro>GDA8</NRegistro><Taxon>Abies pinsapo</Taxon>
<Localidad>Talasentant.</Localidad> <Provincia>-</Provincia>
<Pais>MARRUECOS</Pais><PuntoGeografico><LatitudLongitud><Latitud/><Lon
gitud/></LatitudLongitud></PuntoGeografico><Imagen/></Especímen>

<Especímen><NRegistro>GDA3</NRegistro><Taxon>Abies
pinsapo</Taxon><Localidad>Málaga, Sª de Ronda.</Localidad>
<Provincia>MALAGA</Provincia><Pais>ESPAÑA</Pais>
<PuntoGeografico><LatitudLongitud><Latitud/><Longitud/></LatitudLongitud></P
untoGeografico><Imagen>https://localhost:8443/Biomen/identificado/Imagenes/plieg
os/GDA3.jpg</Imagen></Especímen>

...

```

Figura 69. Resultado de las llamadas a los servicios Web de BioMEN

## 2.10. Exportación de información

Como ya se ha comentado a lo largo del presente documento, la incorporación de mecanismos que permitan el intercambio de información es muy importante para la comunidad científica.

Un sistema que trate cualquier tipo de colección debe de permitir intercambiar sus datos con los proyectos existentes y en los esquemas estandarizados propuestos por la comunidad científica.

BioMEN ha incorporado un módulo en el sistema que permite definir distintos formatos de exportación de forma dinámica. Mediante la definición de vistas de exportación, BioMEN se adapta a las distintas necesidades de los proyectos de intercambio.

### Listado de exportaciones

Seleccione cualquiera de las exportaciones para ver detalles.

Descripción	Nombre De La Vista	Total De Identificadores
<a href="#">ABCD Schema</a>	ABCDSchema	45552
<a href="#">Darwin Core v2</a>	dwc2	997
<a href="#">Generar una nueva exportación</a>		

Figura 70. Listado de vistas de exportaciones en BioMEN

Como podemos observar en la Figura 70, se pueden generar tantas vistas de exportación de datos como sea necesario. En la Figura 70 se puede observar como se han definido dos vistas de exportación, una para exportar datos en formato *ABCD Schema* y otra en formato *Darwin Core*.

La generación de una vista de exportación implica:

- Definir los campos de salida de la exportación.
- Definir las condiciones de los datos que van a ser exportados.

- Se crea una vista de datos en el sistema de gestión de bases de datos que permite acceder a los datos de forma totalmente actualizada y que sólo tiene permisos de lectura.

Al generar una vista de exportación se realizan los cambios necesarios en el gestor de Bases de Datos de BioMEN, de forma que los datos proporcionados por la vista de exportación están totalmente actualizados (ver Figura 71).

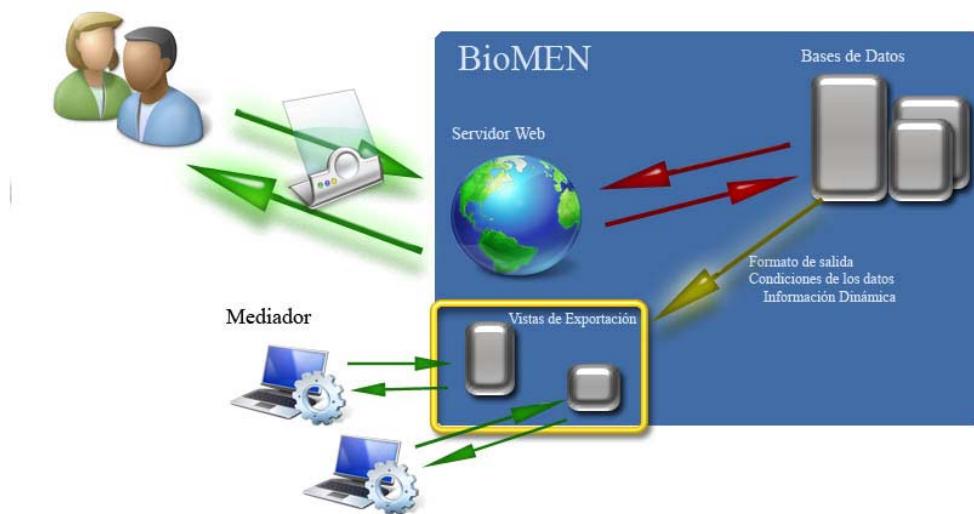


Figura 71. Datos dinámicos para las vistas de exportaciones en BioMEN

La generación de una vista de exportación se realiza en tres pasos:

1. Datos sobre la exportación: datos que indican el nombre de la vista, dirección del esquema que define la exportación, si existe.
2. Definición de los campos de salida: indicando el nombre del campo que tiene el campo exportado. El campo de salida puede ser:
  - Un literal de texto fijo.
  - Un campo de los existentes en el sistema.
  - Una función aplicada sobre un campo del sistema. Por ejemplo, una función que devuelve el campo día de una fecha o una función que reduce la precisión de una coordenada geográfica.

### Campos a exportar

Nombre Del Campo Exportado	Función Sobre Los Datos	Campo A Exportar	Texto Fijo	Eliminar Campo
InstitutionCode	Ninguna Función	Ninguno	RPP	
Family	Ninguna Función	Familia		
DayRecolection	Devolver el Día	Fecha		
	Ninguna Función	Ninguno		
	Ninguna Función	Ninguno		

Figura 72. Extracto de la definición de los campos a exportar en BioMEN

- Definición de las condiciones de exportación: establece las condiciones que deben de satisfacer los datos exportados.

### Condiciones de la exportación

Campo De Biomen	Condición	Texto	Eliminar Condición
Familia	Igual que	Pinaceae	
Ninguno	CONDICION		

Figura 73. Extracto de la definición de condiciones para la exportación en BioMEN

Una vez que todos los campos son definidos, se genera una vista en el sistema de base de datos que permite a otros sistemas externos conectarse a la vista, sólo con permisos de lectura, e intercambiar los datos mediante la utilización de sistemas mediadores (*wrappers*).

## 2.11. Distribución de Especímenes

La visualización de la distribución de los especímenes es una opción importante para la realización de estudios de biodiversidad. La observación de la distribución de los especímenes existentes para una especie determinada puede servir de base para establecer su área de distribución.

Una vez que desde el sistema se permite la incorporación y tratamiento de la información espacial, no todos los sistemas analizados permiten la generación de mapas con este tipo de información.

Esto está principalmente ligado a la dificultad de incorporar al sistema *software* un sistema de información geográfica. No obstante, existen otras alternativas, menos potentes pero que sirven para nuestro propósito.

Desde BioMEN se ha realizado la integración con servicios externos que permiten mostrar de forma visual la situación de un espécimen.

Por ello, BioMEN incorpora la integración con los siguientes servicios de visualización espacial:

1. **MapQuest** (disponible en <http://www.mapquest.com/>): sistema de acceso libre para uso no comercial. El sistema permite representar un punto expresado en latitud/longitud de forma visual.
2. **Google Earth** (<http://earth.google.com/>): servicio ofertado por Google con distintos tipos de licencias de uso (gratuita, plus y profesional) dependiendo de la funcionalidad del *software*. El sistema permite representar un conjunto de puntos.

Cada uno de los sistemas integrados necesita diferentes formas de comunicación. Por ello, se va a mostrar los pasos a seguir para cada uno:

- **MaqQuest:** este servicio sólo permite pintar un punto a la vez. Por ello, este servicio se ha usado en la consulta individual de muestras.
  - Petición:  
<http://www.mapquest.com/maps/map.adp?searchtype=address&formtype=address&latlongtype=decimal&latitude=<latitud>&longituede=<longitud>>
  - Ejemplo:
    - i. Petición:  
<http://www.mapquest.com/maps/map.adp?searchtype=address&formtype=address&latlongtype=decimal&latitude=36.9166667&longitude=-3.4833333>

## ii. Respuesta:

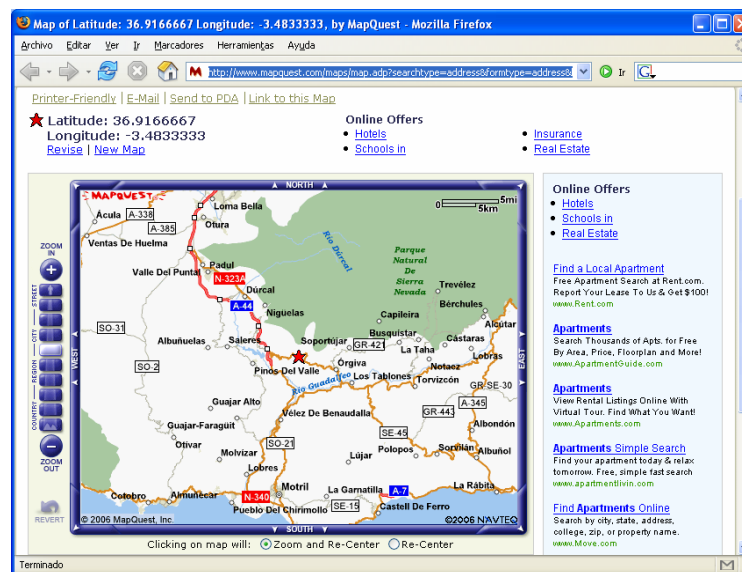


Figura 74. Respuesta de solicitud de visualización de mapa en MapQuest

- **Google Earth:** Es una aplicación de escritorio que combina una base cartográfica de imágenes aéreas y de satélite de alta resolución. *Google Earth* permite guardar o compartir su información mediante ficheros **KML** (*Keyhole Markup Language*), en los cuales se pueden definir lugares de interés, capas de datos, rutas y polígonos.
  - Petición: desde distintas partes de BioMEN se permite la visualización de información espacial con *Google Earth* (consulta avanzada, consulta simple y consulta por muestra). Una vez que el sistema recibe la petición, BioMEN genera un fichero en formato KML que contiene toda la información necesaria para representar los especímenes relacionados con la consulta. A continuación se muestra el código generado por BioMEN:

```

<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://earth.google.com/kml/2.0">
  <Folder>
    <name>BioMEN</name>
    <open>1</open>
    <description><![CDATA[<p><strong>BioMEN Data: INSTITUCION DE PRUEBA.</p>
      <p>Data retrieved on: 2006-01-17T12:04:23Z</p>]]></description>
    <Folder>
      <name>Occurrence Records</name>
      <open>1</open>
      <Placemark><name>Cedrus deodora</name>
        <description> ... (Puede ser una página Web) </description>
        <LookAt><longitude>-
          3.4833333</longitude><latitude>36.9166667</latitude><range>305.8880792294568</range>
          <tilt>46.72425699662645</tilt><heading>49.06133439171233</heading>
        </LookAt>
        <Style><IconStyle><Icon><href>root://icons/palette-
          2.png</href><x>128</x><y>224</y><w>32</w><h>32</h><w>32</w><h>32</h></Icon>
        </IconStyle></Style>
        <Point><extrude>1</extrude><altitudeMode>relativeToGround</altitudeMode><coordinates>
          >-3.4833333,36.9166667,125</coordinates>
        </Point>
      </Placemark>

      <Placemark><name>Cedrus deodora</name>
        <description> ... </description>
        <LookAt><longitude>-
          3.4833333</longitude><latitude>36.9166667</latitude><range>305.8880792294568</range>
          <tilt>46.72425699662645</tilt><heading>49.06133439171233</heading>
        </LookAt>
        <Style><IconStyle><Icon><href>root://icons/palette-
          2.png</href><x>128</x><y>224</y><w>32</w><h>32</h><w>32</w><h>32</h></Icon>
        </IconStyle></Style>
        <Point><extrude>1</extrude><altitudeMode>relativeToGround</altitudeMode><coordinates>
          >-3.4833333,36.9166667,125</coordinates>
        </Point>
      </Placemark>
    </Folder>
  </Folder>
</kml>

```

Figura 75. Fichero KML generado por BioMEN para Google Earth

- Respuesta: Para visualizar la respuesta es necesario tener instalado Google Earth (ver Figura 76).

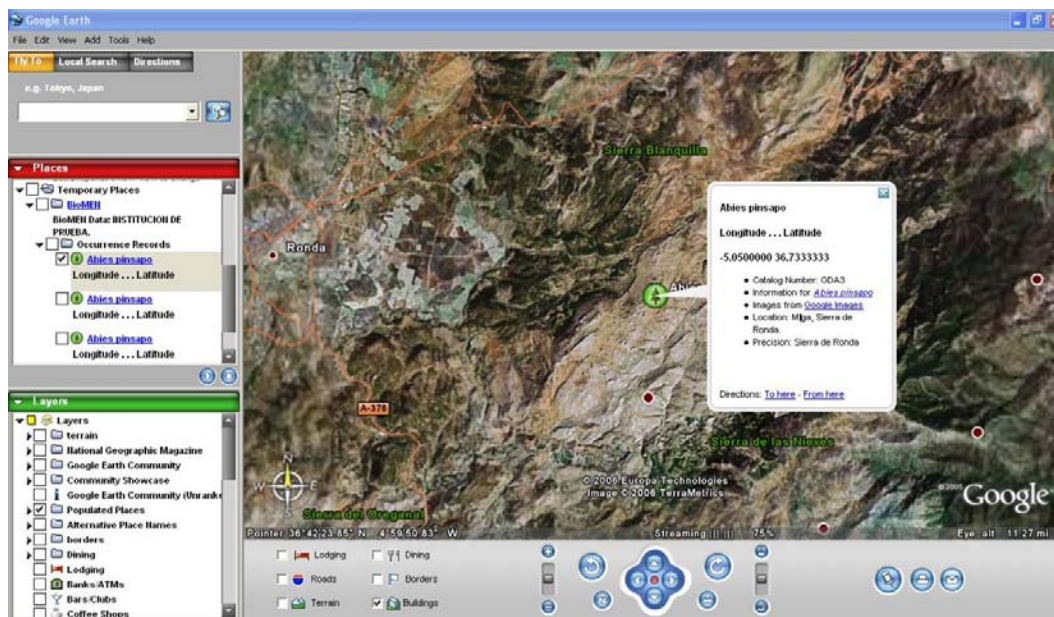


Figura 76. Visualización con Google Earth de datos geospaciales de BioMEN

Además de la posibilidad de visualizar los datos de los especímenes, BioMEN genera una etiqueta para cada uno de los puntos proporcionados pudiendo:

- Visualizar los datos de forma conjunta.
- Acceder a los servicios de consulta externa con un simple clic.
- Visualizar las imágenes existentes en Google.

Todos los servicios proporcionados en este módulo se encuentran complementados con los resultados obtenidos por el Sistema Multiagente que posteriormente comentaremos.

### 3. Comparación con los sistemas actuales

En el Capítulo II se realizó un análisis del problema y se establecieron el conjunto de requerimientos funcionales que tienen que satisfacer los sistemas. Tras el análisis de los sistemas *software* existentes en la actualidad se han



observando deficiencias en las características proporcionadas (ver Tabla 8 y Tabla 9) con respecto a BioMEN.

sistemas\ Características	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
Biota 2											
Specify											
Biotica											
Biolink											
Herbar											
Brahms											
Herbarium Access Management System											
Tracy											
PLabel											
VHE											
<b>BioMEN</b>											

Tabla 8. Tabla comparativa de los sistemas *software* existentes (ver leyenda en siguiente tabla) con BioMEN

Identificador	Significado
1.	Control de usuarios.
2.	Gestión de especímenes.
3.	Generación de Etiquetas.
4.	Consultas.
5.	Préstamos de material.
6.	Información multimedia.
7.	Referencias bibliográficas.
8.	Información taxonómica.
9.	Servicios Externos.
10.	Exportación de información.
11.	Distribución de Especímenes.

Tabla 9. Leyendas de tabla comparativa

De entre todos los sistemas evaluados, Biota 2 es el más completo, aunque presenta los inconvenientes descritos en el capítulo anterior. El principal problema en la utilización de Biota 2 reside fundamentalmente en que no se trata de un *software* de libre distribución por el que hay que adquirir unas licencias *software* para poder usarlo.

Por no ser un sistema de código abierto, no permite que se desarrollen herramientas sobre él para realizar estudios de biodiversidad, completado de datos, integración con fuentes heterogéneas, etc.

Debido a que ningún *software* reúne todas las características necesarias para la gestión global de las colecciones y de la información que éstas genera, se ha desarrollado BioMEN.

En la Tabla 8 podemos observar la comparación de BioMEN con el resto de los sistemas analizados. Como se puede observar BioMEN cumple todos los requerimientos necesarios para resolver el problema.

A lo largo de este capítulo se han mostrado las características de BioMEN y de qué forma se han solucionado cada una de ellas.

Una vez que BioMEN permite gestionar las colecciones y los datos para el Centro, existe la posibilidad de ampliar las cualidades de BioMEN teniendo en cuenta la aplicación de técnicas de inteligencia artificial para realizar estudios de biodiversidad.

En el siguiente capítulo se muestran todas las ventajas de la utilización de un Sistema Multiagente que gestiona la información contenida en las bases de datos de BioMEN.

De forma totalmente autónoma, el Sistema Multiagente se comunica con fuentes externas, integra la información proporcionada por las mismas y muestra la información filtrada al investigador.

## Capítulo IV. BioMEN: Un Sistema Multiagente

En el Capítulo III hemos tratado a BioMEN como un sistema de información que nos permite gestionar colecciones. No obstante, existen otras posibilidades de actuación por parte de las Ciencias de la Computación como puede ser la extracción de conocimiento (E. Mata Montero, 2006) que permitirían a los usuarios del sistema la toma de decisiones, la realización de análisis de biodiversidad, etc.

### 1. Introducción

Todas las herramientas evaluadas hasta el momento, realizan la gestión de un centro, en mayor o menor grado, como se muestra en la Tabla 8. Sin embargo, ninguna de ellas incorpora herramientas para el procesamiento de los datos y la obtención del conocimiento existente en las colecciones biológicas.

BioMEN reúne las características necesarias para proponerse como sistema base sobre el que construir otras herramientas:

- Permite gestionar toda la información de un centro. Información sobre los especímenes, préstamos de material, información taxonómica, accesos de usuarios, etc.
- Ha sido desarrollado bajo la filosofía GPL (*Generic Public License*) de código abierto.
- Está implementado en Java. Lenguaje extensible, modular y altamente portable.
- Es posible acceder al sistema desde cualquier parte.
- Está desarrollado utilizando una arquitectura de tres capas.

La utilización de un *software* viene limitada, principalmente, a:

1. El idioma de la interfaz de usuario.
2. Lo cómodo que se sienta el usuario al utilizar el sistema.

Por ello, el sistema debe de ser capaz de tener en cuenta las preferencias que el usuario indica de forma explícita al sistema (por ejemplo, el idioma). A su vez, para que el usuario se sienta cómodo con el sistema, éste debería de obtener las preferencias que puedan ser deducidas a partir de las operaciones realizadas por él mismo.

Supongamos que existe un usuario que está interesado en una especie concreta, *Alyssum spinosum*, y que siempre que accede al sistema consulta el material incorporado recientemente. El sistema puede obtener una relación de causalidad entre lo consultado por el usuario y el tipo de material que interesa al usuario.

El sistema debería procesar todas las actuaciones realizadas por el usuario y obtener las **preferencias** para poder presentárselas de forma automatizada y totalmente transparente.

Con la adaptación y personalización del sistema al usuario se lograría:

- Eliminar las fronteras idiomáticas.
- Personalizar el sistema proporcionando al usuario la información de su interés.

Estas tareas requieren de un procesamiento externo que sea capaz de tratar los datos asociados a los usuarios y que produzca la información necesaria para cada uno de ellos.

Además, existen otro tipo de problemas en los que sería posible actuar una vez gestionados todos los datos de las colecciones, como por ejemplo, el tratamiento de los sinónimos y la falta de información espacial en las muestras.

La existencia de sinónimos en los nombres científicos falsean el número de especies distintas existentes en un mismo hábitat. El investigador tiene que consultar la bibliografía existente para determinar los **nombres aceptados** y los **sinónimos** existentes para la especie o especies de las que desea realizar el estudio. En este sentido, el sistema puede **filtrar la información** nomenclatural existente en las bases de datos y proponer el nombre aceptado a cada uno de los especímenes teniendo en cuenta la información obtenida de fuentes externas y de los propios datos.

El conocimiento extraído del sistema tras determinar los nombres aceptados y los sinónimos, es importante para la realización de estudios de biodiversidad, como por ejemplo la diversidad alfa, beta, etc.

A nivel ecológico, la biodiversidad tiene dos expresiones bien definidas en el análisis de comunidades (Entomología, 2006; Halffter and Ezcurra, 1992) :

1. la diversidad presente en un sitio, o diversidad alfa: es una función de la cantidad de especies presentes en un mismo hábitat.
2. la heterogeneidad espacial o diversidad beta: es una medida del grado de partición del ambiente en parches o mosaicos biológicos, es decir, mide la contigüidad de *hábitats* diferentes en el espacio.

La información mínima con la que cuentan los datos sobre especímenes son las localidades. Sin embargo, para identificar a un espécimen en un área la forma más precisa es mediante sus coordenadas geográfica, expresada generalmente en latitud/longitud.

Debido a la aparición de nuevas preguntas de interés para la conservación de la biodiversidad y al desarrollo de nuevas herramientas de análisis, la georreferenciación de los especímenes es una tarea importante y relativamente nueva. Por todo ello, gran parte de las colecciones biológicas **no están georreferenciadas** (menos del 5% (Robert Guralnick and David Neufeld, 2005)). Esto implica la necesidad de analizar estos datos e intentar proporcionar una situación espacial que sea útil.

A modo de conclusión, podemos enumerar un conjunto de problemas que son importantes y que hacen el trabajo del investigador más fácil:

- Desarrollar el sistema en múltiples idiomas.
- Obtener las preferencias de uso del sistema de los usuarios.
- Tratar los nombres científicos aceptados y sinónimos.
- Incorporar algún mecanismo que permita completar la georreferenciación de las muestras.

La solución debe de ser suficientemente autónoma como para identificar el problema y actuar sobre él mismo. Para poder hacer esto, claramente el subsistema encargado de la solución deberá estar en continuo funcionamiento y además debe de estar orientado por metas.

Dado este conjunto de problemas y restricciones proponemos, como solución a los mismos, un Sistema Multiagente que trabajando de forma cooperante pueda conseguir las metas propuestas.

De esta forma la solución propuesta estará formada por un conjunto de agentes mediante los cuales se resolverá el problema a través de la resolución de los subproblemas en los que se descompone.

La construcción de un Sistema Multiagente se adapta perfectamente a nuestras necesidades del problema por las propiedades de los agentes:

- Son reactivos (detectar y actuar).
- Son sistemas autónomos.
- Están orientados por metas, es decir, persiguen un fin.
- Son temporalmente continuos; es algo persistente.

El Sistema Multiagente trabaja de forma independiente al usuario proporcionando resultados para cuando el investigador los solicite. A continuación se describe el Sistema Multiagente que realiza labores de análisis y de procesamiento de datos.

## **2. Sistema Multiagente Cooperativo Distribuido a través de Internet**

Una sociedad de agentes se configura como un conjunto de entidades autónomas con capacidad para:

- Generar objetivos propios en base a su percepción de la situación del medio en que operan.
- Usar conocimiento para plantear soluciones que permitan alcanzar sus objetivos.
- Apoyarse en unas reglas sociales que establecen las relaciones de información entre agentes de forma que permiten la cooperación.

La construcción de una sociedad de agentes persigue la consecución de un objetivo global: una adaptación a las necesidades del usuario y un estudio de los datos que permita, en este caso, realizar estudios de biodiversidad. Para lograr este objetivo, nuestro sistema tiene que estar organizado como una sociedad cooperativa que proporciona respuestas para tantos problemas como sus agentes son capaces de detectar.

El Sistema Multiagente ofrece la posibilidad de configurar el producto final gracias a la interacción autónoma entre él y los usuarios.

El principal estándar en la realización de sistemas Multiagentes es el desarrollado por la organización de la FIPA (IEEE Foundation for Intelligent Physical Agents, 2006) (ver Apéndice A).

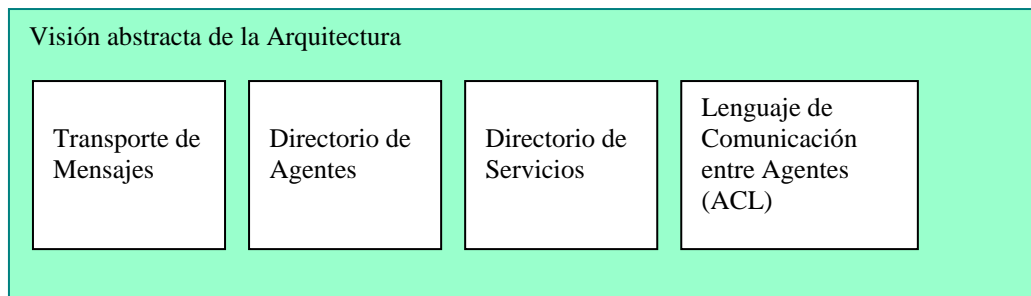


Figura 77. Visión abstracta de la Arquitectura del Sistema Multiagente

La arquitectura de nuestro Sistema Multiagentes viene determinada por la propuesta por la FIPA (ver Figura 77). La construcción en forma modular del sistema permite la incorporación y el intercambio de nuevos agentes con otras arquitecturas FIPA. Los agentes pueden usar diferentes métodos de transporte, diferentes lenguajes de comunicación, o diferentes lenguajes en el contenido del mensaje.

Las ventajas que aporta la utilización de un estándar como la FIPA para la construcción del Sistema Multiagente son las siguientes:

- Directorio de Agentes que tiene constancia de los agentes registrados en el sistema y las características de los mismos.
- Servicio de Directorio de Servicios en el que se registran los servicios que cada agente oferta a la comunidad. Mediante este directorio los agentes pueden describir sus servicios para que puedan ser localizados por otros agentes o servicios.
- Servicio de Transporte de Mensajes que realiza el transporte de mensajes entre agentes. Un agente envía un mensaje hacia otro agente y este servicio realiza las operaciones necesarias para que le llegue en el lenguaje adecuado.
- Interoperabilidad en el transporte de mensajes.
- Soporta varios tipos de lenguajes de comunicación.
- Soporta varios lenguajes en el contenido del Mensaje.



- Soporta varias representaciones en el directorio de servicios.

Para el desarrollo e implementación se ha utilizado la plataforma JADE (*Java Agent Development Framework*). Este proyecto de código abierto proporciona un marco de desarrollo en Java para la construcción de Agentes respetando las especificaciones de la FIPA.

### 3. Características

La utilización de un Sistema Multiagente permite realizar un tratamiento inteligente de las colecciones actuando de forma autónoma. Los agentes que forman el sistema persiguen la consecución de un conjunto de metas que forman el objetivo global del sistema.

El Sistema Multiagente lo constituyen cuatro grupos de agentes, según su temática, con las siguientes características:

- **Grupo de personalización:** conjunto de agentes que adaptarán el sistema a las preferencias detectadas para los individuos que interactúan en el sistema:
  - **Agentes Idioma:** son los encargados de adaptar el idioma de la interfaz a las preferencias del usuario.
  - **Agente Perfiles de Usuario:** es el encargado de obtener las preferencias del usuario. Actúa en cooperación con los Agentes Idioma para adaptar la interfaz del usuario.
- **Grupos de Georreferenciación:** tratan la información espacial de la muestra para intentar asignar una coordenada geográfica lo más precisa posible

- **Agentes Localidad:** permite procesar y tratar la información espacial de la muestra. Un estudio de la localidad de la muestra permite extraer información de su georreferenciación.
- **Agentes Asignación de Localidades:** en cooperación con los Agentes Localidad permite asociar información extra a una muestra. Para aquellas que no tengan este tipo de información, se realiza la georreferenciación de la muestra de forma totalmente independiente.
- **Grupo de Proyectos Nomenclaturales Externos:** grupo encargado de las comunicaciones con proyectos externos de nomenclatura.
  - **Agente Nombre Científico:** actúa como mediador entre las muestras y los agentes Nomenclaturales.
  - **Agentes Nomenclaturales:** son los encargados de las comunicaciones con proyectos externos para obtener información acerca del *status* de un nombre científico. Se establece un agente de comunicación por cada proyecto externo a integrar.
- **Estudio nomenclatural local:** Estudio de la asignación nomenclatural establecida en las muestras.
  - **Agente Campaña de recolección:** procesa las muestras existentes en el sistema e informa de las especies menos representadas en el sistema para orientar las campañas de recolección.
  - **Agente de Revisiones:** procesa las revisiones realizadas sobre las muestras para obtener información que nos indique lo estudiada que ha sido una muestra.
  - **Agente de Secuenciación Nominal:** encargado de realizar un estudio local de los nombres científicos asociados a las muestras

Debido a la estructura modular del sistema y a las ventajas proporcionadas por usar la estandarización propuesta por la FIPA es posible añadir tantos Agentes como sea necesario.

El proceso de incorporación de un nuevo Agente necesita:

- Realizar el análisis y diseño de sus características.
- Establecer los flujos de información entre el Sistema Multiagente
- Definir las ontologías que definen su dominio.

A lo largo de este capítulo se va a tratar la forma de actuar de los agentes que forman el Sistema Multiagente y los problemas que éstos solucionan.

El Sistema Multiagente se ha diseñado para actuar frente a los siguientes cuatro problemas:

1. **Personalización del interfaz del usuario:** la adaptación del idioma al usuario y la especialización de la información mostrada al mismo a través de su interacción con el sistema.
2. **Georreferenciación de Colecciones:** la ausencia de datos de georreferenciación en las muestras impiden su representación espacial y la realización de modelos de distribución.
3. **Comunicación con proyectos nomenclaturales:** la comunicación con proyectos externos permite obtener información actualizada de la relación entre los nombres aceptados y los sinónimos asociados a éste.
4. **Estudio de la Secuenciación Nominal:** existe información intrínseca asociada a las revisiones nomenclaturales realizadas a las muestras de las colecciones biológicas. El análisis de estos datos permitirá obtener la relación existente entre las revisiones realizadas a las muestras y el *status* de los nombres científicos.

Veamos a continuación la forma de resolver cada uno de los problemas expuestos anteriormente.

### 3.1. Personalización de la interfaz del usuario

Pretendemos que Biomen se especialice y adapte en función de las necesidades de sus usuarios. La utilización del sistema BioMEN por parte de investigadores implica la especialización del *software* y la adaptación del mismo a las necesidades de sus usuarios.

El problema de la personalización Web lo podemos estructurar en dos subproblemas:

- La adaptación de la interfaz al usuario.
- Adaptación de la información proporcionada en función de los intereses detectados en el usuario.

La **adaptación de la interfaz** del usuario está motivada, principalmente, por la utilización del idioma solicitado por el mismo. Mediante la interfaz de entrada al sistema, el usuario indica el idioma deseado. Esta preferencia es informada al Sistema Multiagente, y los agentes ocupados de resolver este problema realizan las tareas necesarias, para que la interfaz se presente en el idioma solicitado por el usuario.

La utilización de una arquitectura en tres capas de BioMEN permite tratar de forma independiente a cada uno de los usuarios del sistema. Para la construcción de la interfaz se ha utilizado la información proporcionada por la capa lógica. Ésta está estructurada en XML con un formato propio (*XML Schema*) que permite:

- Construir la interfaz de forma dinámica por el Sistema Multiagente.
- Intercambiar los ficheros de idioma entre usuarios.

El fichero en formato XML que define la interfaz de usuario está especificado mediante una ontología denominada ontología de idioma (ver Figura 78).

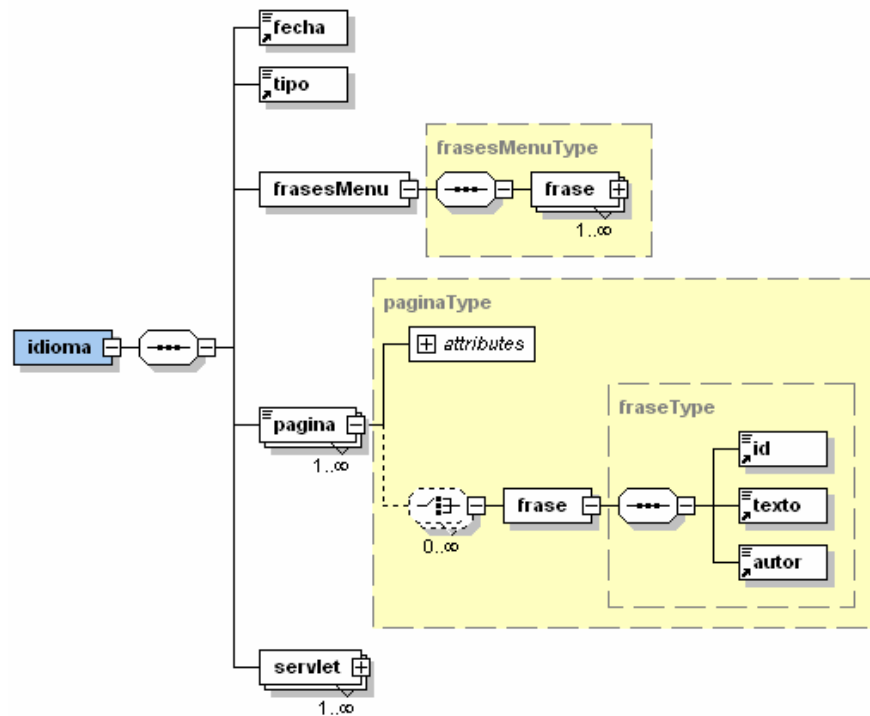


Figura 78. Ontología de idioma para BioMEN

La ontología de idioma especifica la estructura del fichero XML que define la información que será usada para la construcción de la interfaz.

Esta ontología está formada, principalmente, por cuatro términos:

- El componente **frase**: indica la equivalencia entre el identificador (id) dado por la capa lógica y el texto (texto) que tiene que mostrar al usuario.
- La información sobre el menú de operaciones que se muestra al usuario, denominado **frasesMenu**: está formado por una o más **frases**.
- Cada una de las **páginas** o pantallas que se presenta en el sistema y que forman la interfaz del usuario: están formadas por un conjunto de **frases**.

- Cada página Web es generada por un servlet. En ocasiones un **servlet** puede formar más de una página dependiendo de los parámetros con los que se ejecuta.

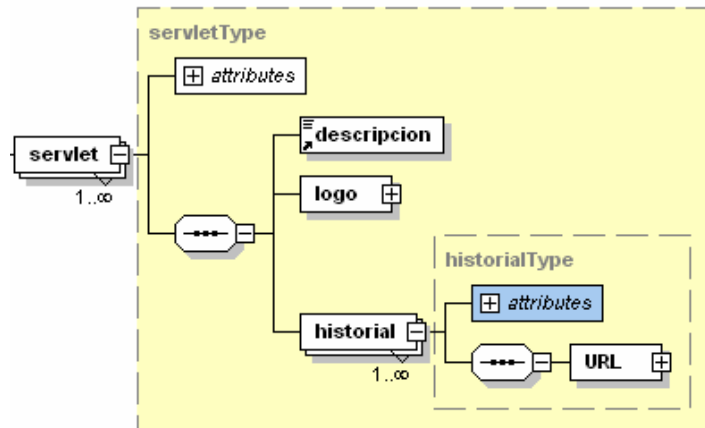


Figura 79. Parte que personaliza la página Web

Un servlet en el sistema puede cambiar el logotipo de la página y la barra situada a la izquierda, para personalizar la interfaz a las necesidades del módulo que se esté ejecutando, y a las necesidades del usuario (ver Figura 80).

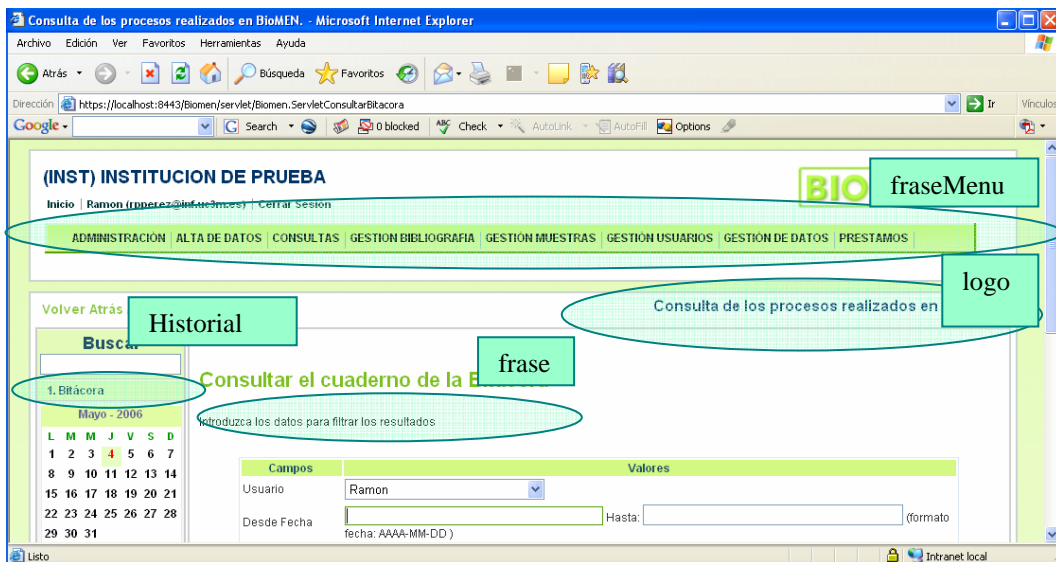


Figura 80. Partes personalizables de BioMEN

Ésta es la forma que tiene BioMEN de generar la interfaz para que se adapte a las preferencias del usuario. Ahora tenemos que tratar de qué forma el

Sistema Multiagente realiza su tarea cuando se selecciona un idioma que no está registrado en el sistema.

Cuando el usuario solicita acceder al sistema en un idioma para el cual éste no tiene el fichero correspondiente al idioma deseado, el agente Lenguaje reacciona a la petición generándolo.

En un primer paso, el Agente Lenguaje al no tener la interfaz solicitada, muestra al usuario el sistema en el idioma establecido como predeterminado y comienza a establecer las comunicaciones pertinentes para generar de forma automática, valiéndose de servicios externos, la interfaz para el nuevo idioma.

Los pasos que rigen la actuación del agente son los siguientes:

1. Establece como base uno de los formatos existentes en el sistema.
2. Mediante un proceso de negociación obtiene la información que necesita de los agentes presentes en el Sistema Multiagente. Para cada uno de los textos existentes en la instancia del idioma consulta a un servicio externo de traducción.
3. Construye el fichero XML correspondiente que servirá al usuario que lo solicitó para la próxima vez que se identifique en el sistema.

Veamos a continuación la forma de cooperar de los agentes para obtener su meta: establecer la interfaz del usuario.

Uno de los agentes del sistema es el encargado de conseguir el objetivo de traducir la interfaz del usuario. Para ello, utiliza los servicios proporcionados por otros agentes que, dado un texto, lo traducen.

El agente encargado de traducir toda la interfaz consulta el directorio de Agentes y el directorio de Servicios para localizar los Agentes que son capaces de traducir un texto (ver Figura 81).

```

DFAgentDescription template = new DFAgentDescription();
ServiceDescription sd = new ServiceDescription();
sd.setType("translate-language");
template.addServices(sd);
try {
    DFAgentDescription[] result = DFService.search(myAgent, template);
    System.out.println("Encontrados los siguientes agentes de traducción:");
    agentes = new AID[result.length];
    for (int i = 0; i < result.length; ++i) {
        agentes[i] = result[i].getName();
    }
}
catch (FIPAException fe) {
    fe.printStackTrace();
}

```

Figura 81. Consulta de los agentes que traducen texto (Translate-language)

Una vez que se identifican los agentes a los que puede solicitar el servicio, comienza un proceso de negociación entre los Agentes para establecer quién realiza el trabajo (ver Figura 82). El proceso se establece de acuerdo al protocolo de actuación denominado *FIPA Contract Net Interaction Protocol* (ver Apéndice A).

El agente lanza cfp (call for proposal) a “m” individuos.  
 - “j” agentes responden a la proposición con el texto traducido.  
 - “i” agentes rechazan la propuesta (refuse)

El agente inicializador recibe los textos traducidos de aquellos que pudieron resolverlo e informa de la realización de su consulta al primer Agente que resolvió la petición.

Figura 82. Algoritmo para la negociación de la traducción de un texto

La comunicación entre los distintos agentes se establece mediante la utilización de una ontología común.

El fin del Agente Lenguaje es la construcción del fichero que permita al usuario tener la interfaz en el idioma deseado. Para ello tiene que:

- Identificar qué partes son las posibles a traducir.
- Comunicarse con los agentes que se comunican con los servicios externos de traducción.



Los servicios externos usados para la traducción son:

1. El servicio de traducción de BabelFish (<http://www.xmethods.net/sd/2001/BabelFishService.wsdl>). Un ejemplo de los idiomas que traduce: inglés a chino simplificado, inglés a chino tradicional, inglés a neerlandés, inglés a francés, inglés a alemán, inglés a griego, inglés a italiano, inglés a japonés, inglés a coreano, inglés a portugués, etc.
2. El servicio de traducción de WebServiceX (<http://www.webservicex.net/TranslateService.asmx?WSDL>) que permite, a su vez traducir: inglés a chino, inglés a francés, inglés a alemán, inglés a italiano, inglés a japonés, inglés a coreano, inglés a portugués, inglés a español, chino a inglés, francés a inglés, francés a alemán, alemán a inglés, alemán a francés, ...

Este agente además actúa como pasarela para otros agentes que necesitan del servicio de traducción como por ejemplo el Agente que genera los perfiles de usuario.

El siguiente problema a resolver es la obtención de los perfiles de los usuarios teniendo en cuenta las operaciones que realizan los mismos. Si el sistema es capaz de obtener las preferencias, podrá informar a los usuarios de la incorporación de nuevo material que sea de interés para ellos. Para ello, se ha diseñado otro Agente que procesa los datos almacenados en el sistema y las operaciones solicitadas por el usuario.

Para llevar a cabo la obtención de los perfiles de los investigadores se han analizados dos alternativas:

- Filtrado Cooperativo (CF del inglés "*Collaborative Filtering*") (Y.H. Cho and L.K. Kim, 2004).
- Minería de Datos Web (Mobasher B., Dai, 2002)

Veamos en qué consiste cada una de estas técnicas, sus ventajas e inconvenientes.

### 3.1.1. Filtrado Cooperativo (CF)

El uso del sistema de filtrado cooperativo (*Collaborative filtering Systems*) (Paul Resnick, Neophytos Iacovou, 1994) nos permite obtener las recomendaciones para la personalización del sistema. Éstas vienen dadas por la similitud entre las preferencias del usuario con la de otros usuarios del sistema.

La forma de actuar de este proceso se puede resumir de la siguiente forma:

- El usuario indica sus preferencias.
- El sistema busca usuarios (vecinos) con preferencias similares formando una comunidad de vecinos (vecindario).
- Intenta personalizar la Web prediciendo el interés del usuario y basándose en el comportamiento de su vecindario. El problema a establecer las mejores recomendaciones: se denomina el problema de las **mejores-N recomendaciones** (*top-N recommendation problem*).

Este proceso tiene en cuenta las preferencias finales del usuario. Por tanto, el usuario tiene que indicar las preferencias de forma explícita. Este proceso plantea los siguientes problemas:

- **Poca densidad** (*Sparsity*): cuando existen muchas muestras es imposible que un sólo usuario las evalúe todas. Sabemos que el método de filtrado cooperativo funciona bien cuando existe un gran número de evaluaciones de usuarios en común. Cuando esto no sucede, no se forman bien los vecindarios, con lo cual el sistema no podrá realizar buenas sugerencias. Existen algunas soluciones a este problema:

- Aumentar la evaluación observando el comportamiento del usuario (*Implicit ratings*). Por ejemplo, usando técnicas de Minería de Datos se observa al usuario y se analiza su comportamiento en la visita recogiendo sus valoraciones implícitas.
  - Construir un método híbrido basado en el contenido del sistema y el filtrado cooperativo (*Hybrid filtering*).
  - Analizar al usuario para identificar relaciones entre las diferentes muestras consultadas y realizar una predicción de la puntuación para el usuario-muestra (*product-to-product*).
- **Problema de la Dimensión** (*Scalability*): tener muchos perfiles implica manejar una dimensión alta para formar el vecindario. Esto consume mucho tiempo para calcularlo.

### 3.1.2. Minería de Datos Web

La minería de datos Web (*Web Mining*) se refiere al proceso global de descubrir información o conocimiento potencialmente útil y previamente desconocido a partir de datos de la Web (U.M. Fayyad, 1996).

Dentro de la Minería de Datos Web existen distintas áreas dependiendo del objetivo deseado (Y.H. Cho and L.K. Kim, 2004):

- Minería Web del Contenido (*Web Content Mining*): extraer información del contenido de los documentos en la Web.
- Minería Web de la estructura (*Web Structure Mining*): descubrir un modelo a partir de la topología de enlaces de la red. Este modelo puede ser útil para clasificar o agrupar documentos.

- Minería Web de uso (*Web Usage Mining*): extraer información (hábitos, preferencias, etc. de los usuarios o contenidos y relevancia de documentos) a partir de las sesiones y comportamientos de los usuarios.

En la Figura 83 se muestran las principales áreas de aplicación de la minería de datos de la Web (Araya, Silva M., 2004).

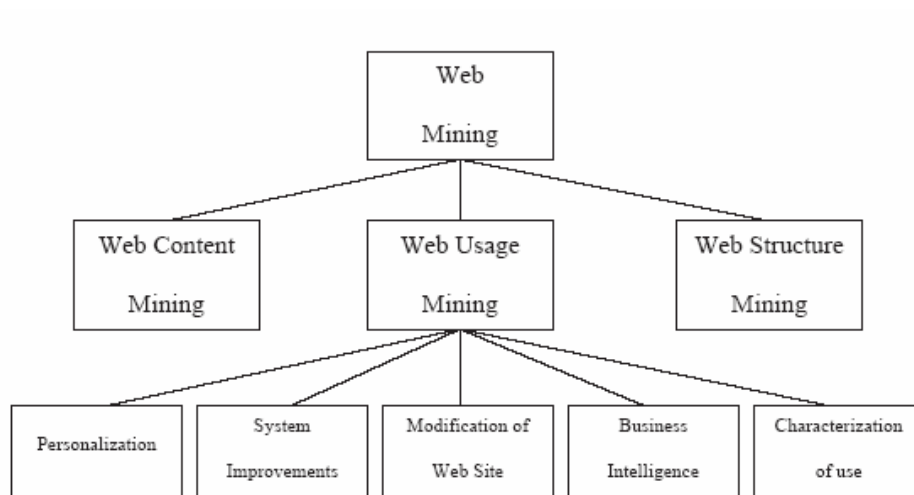


Figura 83. áreas de aplicación de Minería Web (Araya, Silva M., 2004)

El área de la minería de datos aplica técnicas de inteligencia artificial para personalizar el sistema y mejorar la interfaz para el uso del usuario.

El método propuesto por Araya (Araya, Silva M., 2004) está compuesto por el siguiente conjunto de fases:

1. identificación de objetivos y la disponibilidad de los datos: captar nuevos usuarios, proporcionar información interesante y de utilidad al usuario, ...
2. Selección: identificación del conjunto de datos con el que se va a trabajar (*Web data, Business data, meta data*).
3. Pre-procesamiento y transformación: cómo transformar la Web para el usuario.

4. Obtención de patrones e interpretación y evaluación de los resultados: según los objetivos, aplicar distintas técnicas de minería de datos.
5. Integración en el Negocio/sistema: Todo el conocimiento obtenido se tiene que ver reflejado en la integración con el sistema Web de forma que dinámicamente el sistema se personalice para el usuario y se consiga una mayor satisfacción del usuario con el sistema.

Este método propone añadir las fases 1 y 5 al proceso de extracción de conocimiento.

Una vez vistas las dos alternativas propuestas en este documento y teniendo en cuenta nuestro problema, observamos que si aplicamos el filtrado cooperativo (*Collaborative Filtering*) se presenta el problema de poca densidad debido a la gran cantidad de especímenes con los que trabaja un centro y a la imposibilidad de que todos los usuarios evalúen o consulten todos los especímenes. Por tanto, la alternativa a seguir será la propuesta por Araya (Araya, Silva M., 2004).

### **3.1.3. Aplicación de la Minería de Datos a la Web.**

En este apartado se ha implementado una adaptación del método propuesto por Araya (Araya, Silva M., 2004) para obtener las preferencias de los usuarios. Una vez que el Sistema Multiagente obtiene las recomendaciones para el usuario, éste emite boletines para que esté continuamente informado de las nuevas incorporaciones de material.

Veamos como abordar cada una de las fases del modelo elegido:

#### ***Datos disponibles para el sistema***

El sistema genera las páginas Web de una forma dinámica. Por tanto, existe información que no va a ser útil para el sistema, como por ejemplo: el

fichero de accesos, o el fichero de errores. Esto se debe a que el sistema no tiene información de las páginas visitadas por el usuario debido a que ésta no queda registrada.

Para solucionar este problema, se ha incorporado a BioMEN un módulo que registra el comportamiento del usuario, incorporando nuevas tablas en las bases de datos que proporcionarán información para el Agente Perfiles de usuarios.

En estas tablas se almacenará información referente a:

- Consultas realizadas: Especímenes (familia, género, especie, ... ), Provincias, Localidades, Determinadores, etc.
- Nuevos datos introducidos.
- Identificador del usuario.
- Hora, fecha de acceso.

### Selección

El sistema almacena toda la información del usuario y las operaciones en el seno del sistema. De esta información se obtiene la necesaria para la personalización.

### Preprocesamiento y transformación

El acceso a los datos por parte del usuario está perfectamente identificado puesto que existe un control exhaustivo de las tareas que éste puede realizar en el sistema. La incorporación de un bitácora evita la transformación y preprocesamiento de los ficheros de *log* almacenados en el sistema.

### Obtención de Patrones

Como ya se ha comentado anteriormente, nuestra intención es la de identificar patrones de comportamiento en el uso del sistema, para facilitar el trabajo de los investigadores mediante la previsión de sus necesidades.

Utilizando todos los datos indicados en las fases anteriores, se lleva a cabo un análisis de las consultas realizadas estableciendo un conjunto de recomendaciones.

Podemos intentar solucionar el problema de las **N-mejores recomendaciones** (*top-N recommendations*) teniendo en cuenta los datos indicados en las fases anteriores. La utilización de un cluster difuso, haciendo uso del algoritmo k-medias difuso, identifica los distintos grupos difusos de usuarios que existen en el sistema.

Una vez identificados los grupos tenemos dos formas de actuar para resolver el problema de las N-recomendaciones:

- Sin asignar al usuario al clusters de mayor grado de pertenencia.
- Asignando al usuario al clusters de mayor grado de pertenencia.

### Integración en el Negocio/sistema

Dadas las dos alternativas propuestas en la fase anterior: clusters o grados de pertenencia existen dos formas de actuar para solucionar este problema. Veamos cada una de ellas:

- Sin asignar al usuario al clusters de mayor grado de pertenencia:

Al realizar un clusters difuso obtenemos un conjunto de grados de pertenencia a cada uno de los clusters determinados. Teniendo en cuenta esta información podremos realizar las N recomendaciones.

Supongamos que:

1.  $R_k$  indica el número de recomendaciones existentes para el clusters  $k$ .
2.  $g_{ki}$  indica el grado de pertenencia al clusters  $k$  para un usuario  $i$ .

Podemos definir:

$$tg = \sum_{k=1}^c g_{ki}$$

$$lg_{ki} = \frac{g_{ki}}{tg}; \text{ tal que } \sum_{k=1}^c lg_{ki} = 1;$$

Por tanto,

$$N_{ki} = R_k * lg_{ki} \text{ de forma que } N_i = \sum_{k=1}^c N_{ki} = N\_recomendaciones$$

El valor  $N_{ki}$  nos indica el número de recomendaciones que tenemos que coger del clusters  $k$  para el usuario  $i$ . De esta forma obtendremos una lista de  $N$  recomendaciones haciendo uso de las listas de recomendaciones de cada uno de los clusters identificados en el sistema.

- Asignando un usuario al clusters con mayor grado de pertenencia

Si asignamos un usuario  $i$  al clusters  $k$  al que posee mayor grado de pertenencia, tan solo tendremos que realizar las  $N$ -recomendaciones existentes para dicho clusters.

De las alternativas anteriormente descritas parece más conveniente utilizar la primera de ellas, en las que no se hace una asignación del usuario al clusters sino que se tiene en cuenta la pertenencia a cada uno de ellos. Esta opción permitirá la adaptación del sistema al usuario sin tener que clasificarlo de forma clásica dentro de un grupo.



Tras estudiar la aplicación de esta técnica hemos observado la necesidad de realizar la generación de nuevos clusters de forma dinámica, para que el sistema se adapte a la incorporación de nuevos usuarios.

En este punto hemos realizado una primera aproximación al problema de la personalización de la Web. Haciendo uso de los datos obtenidos por las acciones del usuario, el sistema obtiene las preferencias de éste. Sin embargo, el principal propósito es mostrar al usuario todas las recomendaciones sin limitaciones, no solamente las N-mejores.

Para hacer llegar toda la información al usuario se puede:

1. Emitir boletines por E-mail cuando se dispongan de nuevas noticias.
2. Mostrar las preferencias al usuario mediante la interfaz del sistema.
3. Informar al usuario mediante el uso de un sistema de publicación de noticias.

De estas alternativas, la más conveniente es la utilización de un sistema de publicación ya que no limita el número de recomendaciones posibles.

El sistema de publicación RSS (*Really Simple Syndication*) es un sistema público que emite noticias. Para que esté personalizado, se ha incorporado una mejora que permite, que cada usuario acceda a la información obtenida por el Sistema Multiagente para su perfil.

#### **3.1.4. Aplicación de RSS como sistema de publicación de noticias.**

Mediante la utilización de la minería de datos obtenemos las preferencias del usuario procesando las peticiones que este, u otros con perfil similar, realizan en el sistema.

El sistema de publicación de noticias RSS se está convirtiendo en un estándar para la publicación de noticias. RSS pertenece a la familia del formato XML. Ha sido desarrollado específicamente, para ser utilizado para todo tipo de sitios que se actualicen con frecuencia. Por medio de RSS se puede compartir la información y usarla en otros sitios Web. A esto se le conoce como **sindicación**.

Los programas que leen y presentan fuentes RSS de diferentes procedencias se denominan *agregadores*. Gracias a los agregadores o lectores de noticias se puede obtener resúmenes de todos los sitios que se deseen: desde el escritorio del sistema operativo, programas de correo electrónico, aplicaciones Web que funcionan como agregadores. No es necesario abrir el navegador y visitar decenas de Web.

La idea es usar este canal de comunicación para transmitir información personalizada a los usuarios de BioMEN, ya que no la limita. El Sistema Multiagente existente en BioMEN realiza los procesos necesarios para obtener las preferencias del usuario y para generar información útil para éste. La información generada se publica mediante RSS.

Teniendo en cuenta que el sistema RSS no incorpora ningún mecanismo de identificación, tenemos que incluir algún procedimiento que permita identificar el usuario.

Para ello, el sistema de RSS incorporado en BioMEN permite indicar el identificador del usuario para obtener información personalizada:  
<https://localhost:8443/Biomen/servlet/Biomen.ServletRss?idusuario=ramon>.

De esta forma y haciendo uso de peticiones HTTP personalizadas, accedemos a la información que BioMEN procesa de forma automática.

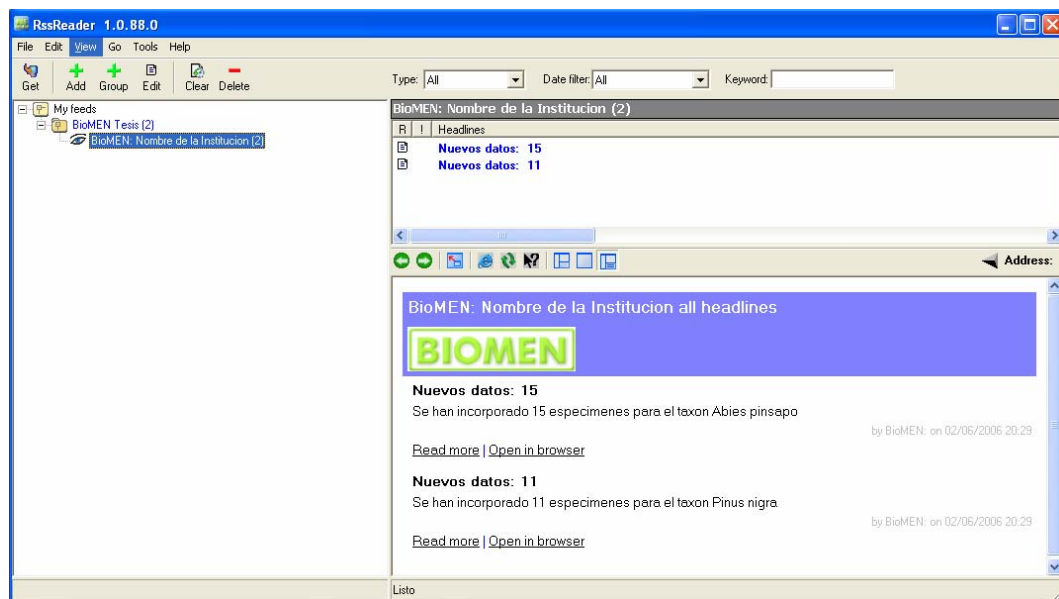


Figura 84. sistema de publicación de noticias personalizadas

Con la publicación de las noticias mediante RSS por el Sistema Multiagente se finaliza el tratamiento de BioMEN para la adaptación de la interfaz al investigador.

Veamos a continuación el siguiente de los problemas, el tratamiento de las colecciones y su georreferenciación.

### 3.2. Georreferenciación de colecciones

El hecho de mantener y disponer de datos correctos de localidad y georreferencia proporciona una herramienta valiosa que permite, a través del uso de Sistemas de Información Geográfica, la monitorización, análisis de tendencias, modelos de simulación, análisis de distribución de especies biológicas, etc, que apoyan la generación de nueva información y contribuyen a la conservación y al uso sostenible de la biodiversidad.

Las colecciones biológicas son la principal fuente de información para los estudios de biodiversidad (E. Mata Montero, 2006; Krishtalka and Humphrey, 2000). Se calcula que existen, hoy en día, más de tres billones (Robert Guralnick and David Neufeld, 2005) de especímenes almacenados en Museos de Historia

Natural y Herbarios. Sin embargo, se estima que tan solo el 5% de los especímenes están georreferenciados (Beaman and Conn, 2003).

La reciente incorporación de sistemas de posicionamiento global (GPS) permite asociar datos espaciales a la muestra. Sin embargo, existen muchas colecciones que no disponen de coordenadas espaciales para situarlas en un mapa geográfico.

No obstante, existe información en las colecciones que pueden ayudar a la georreferenciación, como por ejemplo la localidad, provincia y el país. La localidad de recolección, descrita en lenguaje natural, refleja una posición aproximada.

La utilización de gacetillas (*gazetter*) y de bases de datos de topónimos nos permite asociar una coordenada geográfica a una localidad. No obstante, analizar muestra a muestra para asignar esta referencia es un proceso muy laborioso.

Para alcanzar el objetivo de procesamiento de las localidades y obtener así una georreferenciación base sobre la que el usuario pueda empezar a trabajar, se ha creado un Agente Localidad y un grupo de agentes de comunicación exterior.

Este grupo de agentes analiza las muestras existentes en el sistema generando una georreferenciación para aquellas que no la tengan. Además, permite estandarizar y unificar la información geoespacial proporcionada mediante la transformación a coordenadas de latitud/longitud.

La forma de proceder de este grupo de agentes es la siguiente:

1. Para cada muestra de datos.
  - a. Si tienen datos espaciales, aplica un formato común.
  - b. Si no tiene georreferenciación, utiliza sus capacidades para comunicarse con otros agentes que puedan obtener su localidad y dar su georreferenciación teniendo en cuenta el país, provincia y localidad de la muestra.

Éste grupo de agentes permiten, de una forma fácil e independiente, **completar la información** existente en las bases de datos de las colecciones, sin que medie la intervención del usuario.

Éste es un paso importante para el tratamiento de las colecciones puesto que la existencia de datos sin georreferenciación impide la realización de modelos de distribución, monitorización y estudios de biodiversidad.

Los agentes hacen uso de servicios externos que permiten obtener información asociada a nombres de localidades. Utilizan un sistema de carácter especial, una base de datos de topónimos, que asocia lugares a coordenadas geográficas.

Actualmente, el proyecto denominado BioGeoMancer (BioGeoMancer, 2006) es la principal referencia para obtener este tipo de información. Es un caso muy particular de base de datos de topónimos. Admite consultar un conjunto de localidades concatenadas por comas. Éste es el formato habitual en la información de las localidades en los especímenes biológicos. BioGeoMancer devuelve las coordenadas geográficas para cada una de las localidades que reconoce.

BioGeoMancer provee un servicio de georreferenciación para colectores, conservadores y usuarios de especímenes de historia natural. La información sobre la localización de un espécimen está asociada a una descripción en lenguaje natural del lugar. Para ello, dado proyecto automatiza lo máximo posible el proceso y permite convertir descripciones de lugares a coordenadas latitud/longitud.

La comunicación con BioGeoMancer se realiza a través de peticiones http e indicando los parámetros necesarios para especificar la consulta. Una vez que se realiza la consulta, obtenemos los datos en formato XML. Estos tienen que ser procesados para obtener la información referente a la georreferenciación.

Veamos a continuación un ejemplo de petición en Java de las coordenadas para la localidad de Baza (Granada, España):

```

String strurl = "http://georef.peabody.yale.edu/cgi-bin/bgm-0.2/sample-cgi-form";
HttpClient client = new HttpClient();
PostMethod method = new PostMethod(strurl);
//Establecemos los parámetros
method.addParameter("country", "Spain");
method.addParameter("locality", "Granada, Baza");
method.addParameter("adm1", "");
method.addParameter("adm2", "");
method.addParameter("format", "xml");
method.addParameter(".cgifields", "format");
DefaultMethodRetryHandler retryhandler = new DefaultMethodRetryHandler();
retryhandler.setRequestSentRetryEnabled(true);
retryhandler.setRetryCount(5);
method.setMethodRetryHandler(retryhandler);
int statusCode = -1;
boolean done = false;
String Lat = null, Long = null, Utm = "";
try {
    statusCode = client.executeMethod(method);
    ...
}

```

Figura 85. Ejemplo de comunicación con BioGeoMancer

```

<SOAP-ENV:Envelope xmlns:xsi=http://www.w3.org/1999/XMLSchema-instance ...
<countryBoundingBox xsi:type="xsd:string">-18.1705589294434 27.6374969482422
4.31694412231445 43.772216796875</countryBoundingBox><matchedRecordCount
xsi:type="xsd:int">2</matchedRecordCount> ...
<featureName xsi:type="xsd:string">Granada</featureName><gazetteerSource
xsi:type="xsd:string">NIMA</gazetteerSource><InterpretedCoordinates
xsi:type="xsd:string"> POINT(-3.6000000 37.1833333) </InterpretedCoordinates> ...
<featureName xsi:type="xsd:string">Baza</featureName><gazetteerSource
xsi:type="xsd:string">NIMA</gazetteerSource><InterpretedCoordinates
xsi:type="xsd:string">POINT(-2.7666667 37.4833333) </InterpretedCoordinates>
...

```

Figura 86. Respuesta de BioGeoMancer

El Sistema Multiagente además almacena toda la información referente a la precisión con la que el *gazetter* es capaz de responder. Toda la información que el Agente obtiene sobre el espécimen es almacenada en el sistema de forma totalmente independiente, permitiendo al usuario modificarla o añadir una nueva.

El trabajo se realiza de forma automática y facilita la labor del investigador proporcionando los datos espaciales de aquellos especímenes que no tenían.

BioMEN utiliza la información obtenida y la visualiza mediante dos servicios externos:

- *MapQuest*: visualización de un solo punto sobre un mapa.
- *Google Earth*: permite visualizar un conjunto de muestras sobre un sistema de visualización geográfico interactivo.

El uso de *Google Earth* permite al investigador tener mejor perspectiva sobre la distribución de una especie.

En la Figura 87 se muestra la información obtenida por el Agente Localidad al consultar una muestra en BioMEN.

Campos	Etiqueta Original	Asignada 0
País	ESP	
Provincia	GR	
Localidad	Granada, Lanjarón, Tello.	Granada, Lanjarón, Tello.
Utm		30SVF5694885737
Latitud		36.9166667
Longitud		-3.4833333
Altura	0	
Fecha	26-IV-1943	
<b>Colector</b>		
Número Colectores		
Número contenedores	1	

Figura 87. Datos espaciales obtenidos por el Agente Localidad

Como se puede observar en la Figura 87 el Agente ha tratado de forma independiente la información obtenida para el espécimen (columna "Asignada 0"). La etiqueta original de la muestra recolectada en el 1943, no incluía georreferenciación, pero el grupo de agentes ha sido capaz de comunicarse con los servicios externos que le han provisto de los datos espaciales para su localidad (Granada, Lanjarón, Tello) (ver Figura 88).

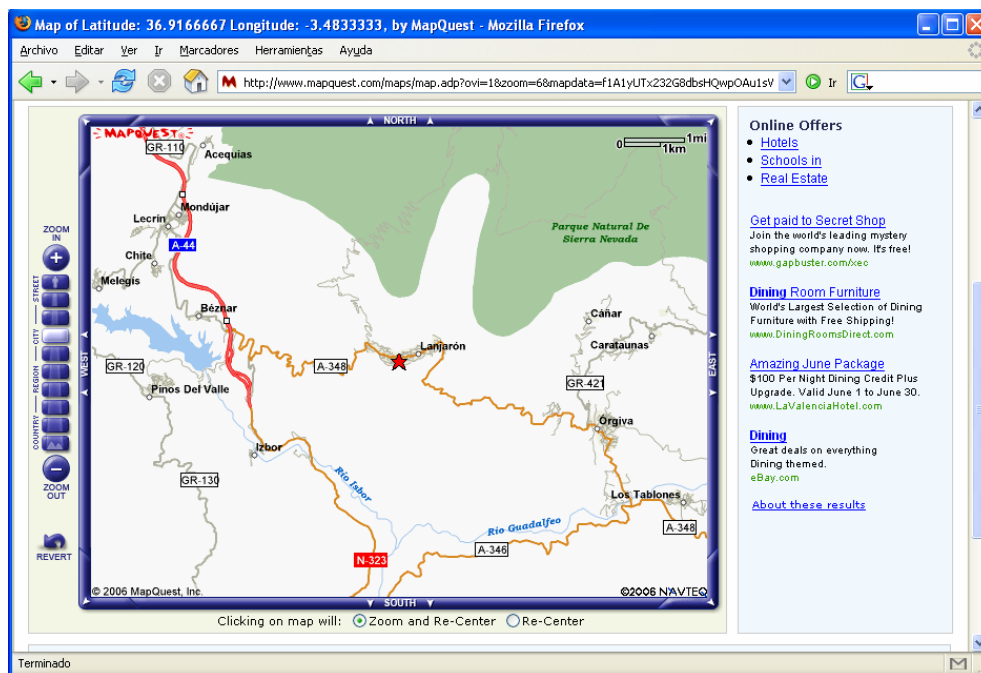


Figura 88. Visualización de la georreferenciación para una muestra

Para cada una de las muestras procesadas por el agente se almacena la siguiente información en el sistema:

- Localidad: datos usados para la obtener la georreferenciación.
- MGRS<sup>11</sup>: coordenada MGRS equivalente a la latitud/longitud.
- Latitud/Longitud: en formato decimal para las coordenadas.
- Nivel de Precisión: indica la precisión o el grado de error cometido en los datos espaciales.
- Validación: indica si la información proporcionada por el Agente ha sido validada por el investigador del sistema.

El Agente localidad permite obtener información espacial importante para la visualización de las muestras y el tratamiento espacial de las mismas.

<sup>11</sup> MGRS es el acrónimo de *Military Grid Reference System*.



El siguiente problema para poder continuar con el tratamiento y la gestión inteligente de las colecciones es el procesamiento de los nombres científicos de cara al filtrado de sinónimos.

### 3.3. Comunicación con proyectos nomenclaturales

Para poder completar la información existente en las bases de datos de colecciones, el sistema debe de ser capaz de tratar los nombres científicos asignados a las muestras. El reconocimiento de los sinónimos existentes en las colecciones permite contar el número real de especies distintas que existen en la misma y, por tanto, abordar estudios de biodiversidad.

Hasta el momento, cuando el usuario realiza una consulta al sistema por un nombre científico, éste muestra la información existente para el nombre científico solicitado. La respuesta que el sistema da al usuario puede no ser la correcta, debido a la existencia de sinónimos que ocultan al resto de especímenes de esa especie almacenados bajo un nombre científico distinto al consultado por el usuario.

La comunicación con los proyectos de tipo *checklist* permiten al sistema informar al usuario, de todas las muestras existentes para el nombre científico teniendo en cuenta tanto el nombre aceptado como los sinónimos facilitados por los proyectos externos.

La consulta a sistemas externos como *Species 2000* e *ITIS* obtiene información del nombre aceptado y sinónimos relacionados con el nombre indicado por el investigador.

La incorporación del grupo de agentes nomenclaturales y del Agente Nombres Científicos permite a BioMEN:

- Verificar la información nomenclatural antes de ser introducida en las bases de datos e informar al usuario, antes de proseguir, con la entrada de datos.

- Mejorar las respuestas a consultas sobre un *taxon*.
- Auto-completar bases de datos. Confrontando información con otras fuentes.

El agente Nombres Científicos actúa como mediador entre la Base de Datos y la información que necesita. Éste agente utiliza la capacidad de otros agentes para comunicarse con proyectos externos de nomenclatura. Veamos como actúan los agentes encargados de pedir información a los principales proyectos de tipo *checklist* (*ITIS* y *Species 2000*).

### 3.3.1. *ITIS*

La inexistencia de servicios Web por parte del proyecto *ITIS* hace que la comunicación no se realice en condiciones óptimas. No obstante, existe la posibilidad de enlazar las búsquedas mediante la utilización de peticiones http, y procesando las páginas Web (HTML) obtenidas como resultado de la petición.

Obtener el resultado en formato HTML implica que el Agente encargado de comunicarse con *ITIS*, tenga que extraer la información de una forma local. Es muy dependiente del código generado. Un cambio en la forma de la página Web impediría al Agente extraer la información que necesita.

El formato para solicitar información es el que se describe a continuación:

```
http://www.itis.usda.gov/servlet/SingleRpt/SingleRpt?  
search_topic=AAAA&search_value=BBBB&search_kingdom=CCCC  
&search_span=DDDD&categories=All&source=html&search_credRating=All
```

Donde:

- AAAA: puede tomar los valores: *Scientific\_Name*, *Common\_Name*, all y TSN (Número de identificación del *Taxon*) según el tipo de búsqueda que

se quiera realizar (Nombre Científico, nombre Común o vulgar, cualquiera o el numero de identificación del *Taxon*).

- BBBB: representa la cadena de búsqueda. Los caracteres especiales, así como los espacios, deben de ser tratados correctamente para que la dirección URL sea correcta.
- CCCC: hace referencia al reino a consultar; cualquiera, Animal, Planta, ...
- DDDD: condición de búsqueda; *exactly\_for* (igual), *containing* (contiene), *starting\_with* (comienza por) y *ending\_with* (termina con)

Veamos un ejemplo de comunicación con *ITIS*. Vamos a buscar los nombres científicos para cualquier reino, y que incluyan la cadena *Abies* dentro del nombre científico.

Url de petición:

[http://www.itis.usda.gov/servlet/SingleRpt/SingleRpt?search\\_topic=all&search\\_value=Abies&search\\_kingdom=every&search\\_span=containing&categories=All&source=html&search\\_credRating=All](http://www.itis.usda.gov/servlet/SingleRpt/SingleRpt?search_topic=all&search_value=Abies&search_kingdom=every&search_span=containing&categories=All&source=html&search_credRating=All)

Resultado:

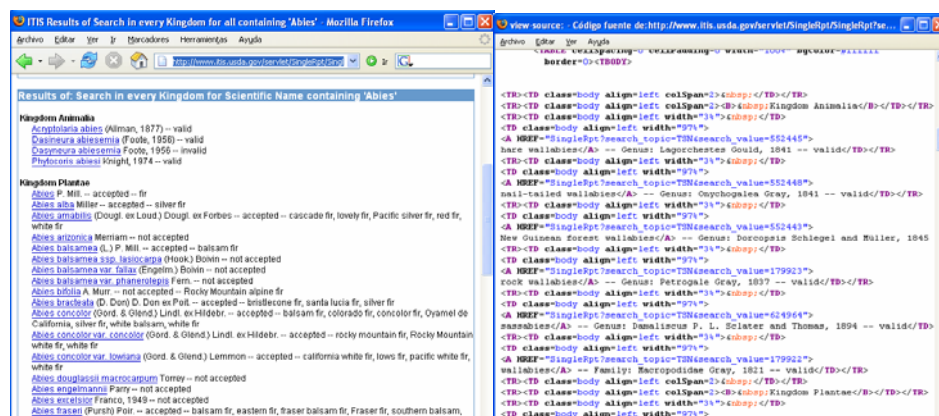


Figura 89. Página Web y código fuente de consultar a ITIS

El Agente tiene que procesar el código fuente generado en la comunicación con el proyecto *ITIS*. Esta solución es muy dependiente. Si se

realizan cambios en el Web de consulta de *ITIS* entonces se generarían fallos de comunicación.

La incorporación de Servicios Web facilitaría la labor del Agente y permitiría comunicarse entre aplicaciones de una forma más eficiente.

### 3.3.2. *Species 2000*

La comunicación con el proyecto *Species 2000* se realiza mediante Servicios Web. La incorporación de un servidor de servicios Web, por parte de *Species 2000*, permite la comunicación entre aplicaciones.

El acceso a *Species 2000* se realiza a través de la interfaz pública WSDL en:

<http://spice.sp2000europa.org/SPICE/services/CASWebService?wsdl>

Por parte de *Species 2000*, se han establecido un conjunto de tipos de datos propios (ver Figura 90 y Figura 91) que almacenan el resultado de las consultas en el sistema. Cuando el agente realiza la consulta al servicio Web, obtiene el conjunto de nombres que cumplen con la condición expuesta por el Agente.



Figura 90. Instancia de consulta a Species 2000 y *status* del nombre

El Agente encargado de comunicarse con Species 2000, consulta al Servicio Web el nombre científico *Pinardia anisocephala* y éste le informa que es un sinónimo (campo: *status*) (ver Figura 90). Por tanto, el usuario está consultando un nombre que no es el aceptado.



Figura 91. Instancia de Species 2000 donde se indica el nombre aceptado para una consulta

En la respuesta a la consulta del Agente se incluye información del nombre aceptado, *Heteranthemis viscidehirta*, establecido para el nombre científico consultado (Figura 91). De esta forma, el Agente obtiene la información necesaria para filtrar las consultas del usuario y muestra los datos que existen en el sistema combinando la respuesta de los proyectos externos (ver Figura 92).

**Resultado de la búsqueda**

Cadena de búsqueda: Pinardia anisocephala  
Estado del nombre: Sinónimo

Taxon	Provincia	País	Ver Con Google
Pinardia anisocephala	HUELVA	ESPAÑA	<a href="#">Ver Mapa</a>
Pinardia anisocephala	HUELVA	ESPAÑA	<a href="#">Ver Mapa</a>

**Fuente: Species 2000**

Nombre Aceptado: Heteranthemis viscedehista

Taxon	Provincia	País	Ver Con Google
Heteranthemis viscedehirta	HUELVA	ESPAÑA	<a href="#">Ver Mapa</a>
Heteranthemis viscedehirta	HUELVA	ESPAÑA	<a href="#">Ver Mapa</a>

Figura 92. Respuesta de BioMEN integrando fuentes externas

BioMEN permite la incorporación de información proveniente de fuentes externas facilitando la labor del investigador. De una forma rápida, BioMEN informa al usuario que realiza la consulta de todo el material existente en sus bases de datos teniendo en cuenta el *status* de los nombres científicos.

Hemos incluido dos agentes de comunicación externa, pero se podrían incluir tantos como proyectos deseemos contemplar. La comunicación con los agentes se realiza de forma dinámica y haciendo uso del Directorio de Servicios. Por tanto, se podrían incluir más agentes y se obtendrían sus resultados de forma inmediata. No obstante, los agentes tienen que ser diseñados de acorde a la ontología de comunicación existente.

El Agente Nombres Científicos consulta el Directorio de Servicio para obtener todos los agentes que proporcionan el servicio de nombres. Una vez que se obtienen los identificadores de los mismos, se lanza la petición de obtener el nombre aceptado y sus sinónimos a todos ellos (*call for proposal*).

Cada uno de los agentes se encarga de procesar la petición y de responder de acuerdo a su sistema, en un mismo formato, usando la ontología de comunicación desarrollada (ver Figura 93). De esta forma el Agente Nombres

Científicos obtiene todos los esquemas contemplados por los distintos proyectos, independientemente del formato de cada proyecto.

Para la comunicación entre los agentes se ha establecido una ontología que representa los conceptos y peticiones que puede realizar el agente (ver Figura 93).

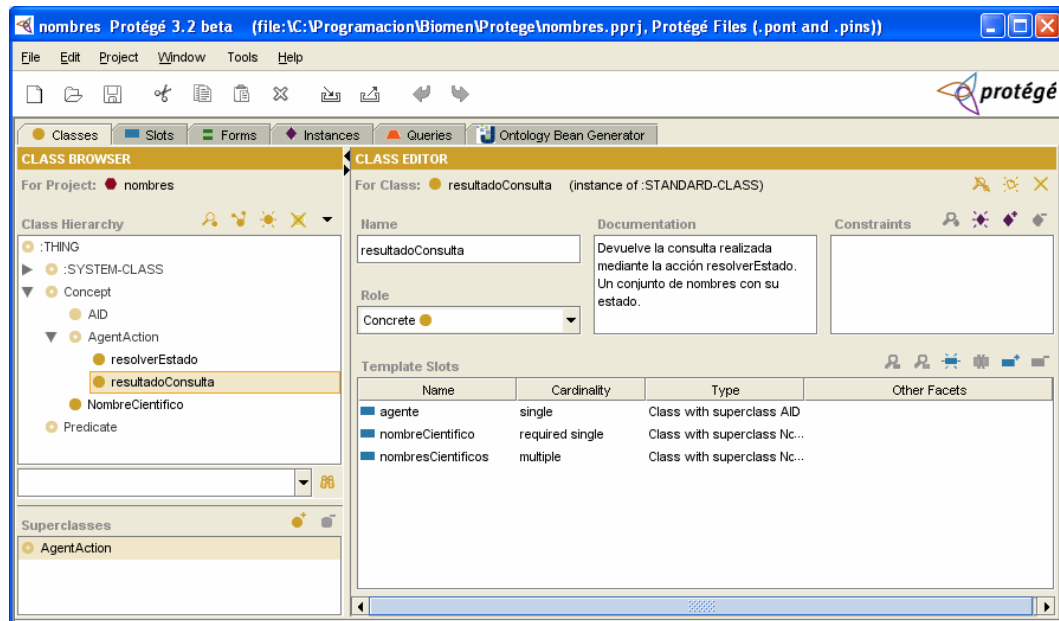


Figura 93. Ontología generada con Protégé para la comunicación entre los Agentes nomenclaturales.

El Agente establece sus necesidades de acuerdo a la ontología y obtiene el resultado de cada uno de los agentes consultados.

De esta forma, y mediante las consultas externas, resolvemos parte del problema de la sinonimia pero hay que tener en cuenta que estos proyectos no están completos y existe una ausencia de información para *taxa*.

Para ello, se ha tratado de obtener información sobre los nombres científicos, a partir de los datos existentes en las bases de datos del sistema.

### 3.4. Estudio de la secuenciación nominal

La principal fuente para estudios de Biodiversidad son los especímenes almacenados en los centros que albergan las colecciones biológicas. Las muestras

son estudiadas y revisadas por investigadores que completan y actualizan la información nomenclatural asociada.

- La existencia de los sinónimos en las determinaciones de los especímenes impide contabilizar las distintas especies que forman la colección.

Debido a la existencia de sinónimos, la información no está directamente accesible, para la realización de estudios de biodiversidad y análisis de los datos, como pueden ser:

- Estudios sobre la Complejidad Taxonómica (Claudia E. Moreno, 2001; Gonzalo Halffter, Claudia E. Moreno, 2001)
- Estudios de la Riqueza Específica (Claudia E. Moreno, 2001)
- Estudios de la Alfa/Beta/Gamma diversidad (Entomología, 2006).

Tomemos el siguiente ejemplo para mostrar el problema asociado. Supongamos que tenemos en el sistema los siguientes taxa: *Alyssum spinosum*, *Hormathophylla spinosa* y *Ptilotrichum spinosum*.

Si contásemos el número de especies distintas ateniéndonos al número de nombres diferentes en el sistema, obtendríamos tres especies distintas. Si consultamos Flora Ibérica (Castroviejo et al., 1996), los tres nombres mencionados hacen referencia a un mismo nombre científico, el aceptado (*Hormathophylla spinosa*). Flora Ibérica además nos muestra los estados por los que ha pasado el nombre. En orden cronológico, el nombre ha cambiado desde *Alyssum spinosum* a *Ptilotrichum spinosum* y por último a *Hormathophylla spinosa*.

Dada la importancia que tiene del problema que genera la sinonimia, vamos a tratar de solucionarlo mediante la utilización del Sistema Multiagente.

Para tratar el problema de la sinonimia tenemos tres formas de actuar:



1. Creación de una base de datos de sinónimos. Esta alternativa acelera el trabajo de procesamiento. Sin embargo, presenta una serie de inconvenientes:
  - El tamaño de la tabla de sinónimos es muy grande, debido al gran número de especies.
  - La tabla tendría que ser realizada por un experto, que tendría que hacer un trabajo repetitivo y tedioso.
2. Comunicación con proyectos externos: La existencia de proyectos internacionales (*Species 2000* e *ITIS*, por ejemplo) que trabajan para recopilar la lista de los nombres aceptados y los sinónimos, permiten tratar el problema pero el inconveniente de utilizar esta vía es:
  - Estos proyectos no están completos y existe una ausencia de información para *taxa*.
  - No se adaptan a la posible jerarquía taxonómica usada.
3. Mediante el estudio de la **secuenciación nominal**: el procesamiento continuo de las muestras y su actualización nomenclatural permiten asociar una secuenciación de los nombres a un cambio de *status* (de aceptado a sinónimo) en los mismos. Las ventajas de utilizar este método son las siguientes:
  - Esta tarea se puede realizar sin la necesidad de la actuación de un experto y nos permitirá obtener la secuencia del cambio en la determinación.
  - Se puede realizar de forma independiente al usuario mediante el Sistema Multiagente.

El estudio de la secuenciación nominal permite:

- Obtener los posibles nombres aceptados.
- Distinguir los sinónimos existentes en las bases de datos teniendo en cuenta la secuenciación nominal de las muestras.
- Facilitar informes al investigador que permiten realizar estudios de biodiversidad como por ejemplo la diversidad alfa/beta.
- Herramienta de limpieza y validación de los datos.

Para obtener el conocimiento citado, se han utilizado dos Agentes Inteligentes (Agente Revisiones y Agente Secuenciación) dentro del Sistema Multiagente establecido. Estos agentes están continuamente observando el medio (bases de datos), realizando un proceso de adquisición y procesamiento de la información para conseguir la información necesaria. Los agentes depositan la información en el sistema, mediante el uso de un sistema de la pizarra (Hayes-Roth, 1985), que permite a los usuarios acceder a la información generada por los Agentes.

La información obtenida por los agentes, Agente Revisiones y Agente Secuenciación, es utilizada por el Agente Campaña de Recolección. Éste analiza los datos de las muestras y emite un informe de aquellos especímenes poco representados en el Centro.

La solución propuesta presenta como principal novedad el tratamiento de los sinónimos existentes en la Base de Datos de una forma local, haciendo uso de la información procedente de las propias muestras de la colección. El método que proponemos no necesita de la intervención directa del experto porque se basa en el estudio de la historia nomenclatural de los especímenes contenidos en la base de datos.

En un paso posterior, el experto puede modificar el conocimiento extraído por los agentes alternado, en adelante, su comportamiento.

Definimos el concepto de secuenciación como el cambio en la determinación de un espécimen en el tiempo. Además, deberemos tener en cuenta que la secuenciación de un nombre no es común para todos los especímenes; en algunos ejemplares, esta secuenciación se encuentra interrumpida o es inexistente, por lo que cada uno de ellos puede estar almacenado en la base de datos con diferentes nombres, aún tratándose del mismo *taxon* (sinónimos).

Veamos los Agentes existentes en el Sistema Multiagente para poder estudiar la secuenciación nominal de cada muestra:

- Agente Revisiones: está encargado de estudiar la cantidad de revisiones que posee cada espécimen. De esta forma se obtiene el conjunto de material más estudiado.
  - La información obtenida por el Agente es depositada en la pizarra (H. Penny Nii., 1986a; H. Penny Nii., 1986b) haciendo uso de la tabla AgenteRevisiones.
- Agente Secuenciación Nominal: utiliza la información obtenida por el anterior agente para estudiar la secuenciación nominal de cada muestra.
  - La información obtenida por el Agente es depositada, igualmente, en la pizarra de intercambio.

A continuación se muestra un ejemplo que clarifica el procedimiento que realizan los agentes inteligentes para obtener los sinónimos:

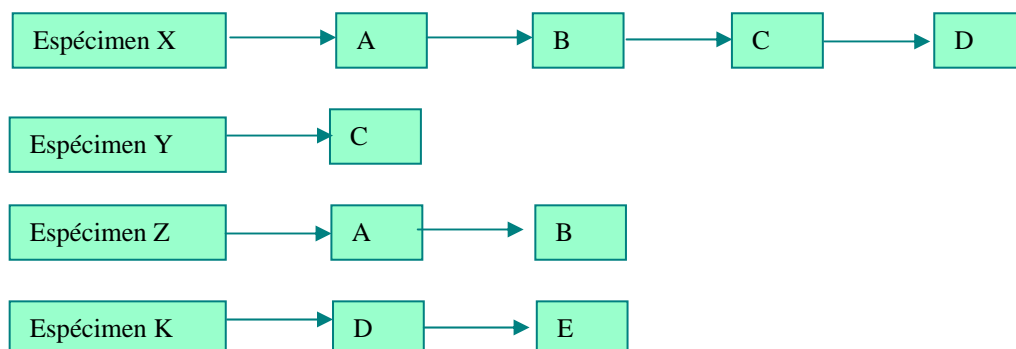


Figura 94. Ejemplo de sinonimia y secuenciación nominal

Los diferentes cuadros determinados como A, B, C, D y E, son las distintas determinaciones por las que pasa un espécimen. Las flechas indican el cambio de nombre en el tiempo. Veamos como se llevaría a cabo el ejemplo más detenidamente:

El espécimen 1 fue determinado como A pero más tarde, tras ser revisado por otros investigadores, es determinado como B, C y D, sucesivamente. Estas determinaciones son tratadas como secuenciaciones y por tanto, las determinaciones A, B y C son sinónimos de D. Estos valores son insertados en la tabla llamada *AlertasSecuenciación* indicando la siguiente información:

Especimen 1:  $A \rightarrow D$ , secuenciación 0.

Especimen 1:  $B \rightarrow D$ , secuenciación 0.

Especimen 1:  $C \rightarrow D$ , secuenciación 0.

El agente introduce los siguientes campos:

- **Identificador:** indica el número del espécimen. En este caso 1.
- **Antecedente:** Previas denominaciones del espécimen. Por ejemplo: A, B o C.
- **Consecuente:** indica la denominación siguiente. En este caso D.
- **Secuenciación.** Este campo puede tomar 2 valores:
  - Cuando secuenciación es 0: indica que el espécimen no será determinado de otra forma.
  - Cuando secuenciación es 1: indica que si el espécimen es estudiado, posiblemente cambie su determinación al indicado por el campo Consecuente. Esta información permite al investigador saber qué material tiene que ser estudiado ya que su nombre no está actualizado.

La última secuenciación nominal encontrada es establecida como el nombre aceptado. Esta información también es almacenada en la pizarra de intercambio. Particularmente, se almacena en la una tabla denominada taxones, para que pueda ser consultada por otros agentes. La información indicada es la siguiente:

- Espécimen 1: D.

Continuando con el ejemplo, la identificación del espécimen 2 es un sinónimo, ya que existe otro espécimen para el cual existe una revisión de C a D.

Por tanto, introduciríamos en la tabla *AlertasSecuenciación*:

- Espécimen 2:  $C \rightarrow D$ , secuenciación 1.

En este caso el valor secuenciación a 1 indica que si el espécimen 2 fuera estudiado seguramente se determinaría como D. Por tanto, lo insertamos como una posible secuenciación nominal.

El siguiente espécimen, espécimen 3, tiene algunas revisiones determinadas como A y B. Como en la tabla de *AlertasSecuenciación* tenemos almacenado que el nombre B puede pasar al nombre D, debido a la secuencia nominal del espécimen 1. Entonces, añadimos en la tabla *AlertasSecuenciación* el siguiente registro:

- Espécimen 3:  $B \rightarrow D$ , secuenciación 1.

Recordemos que en la tabla *taxones* existe un solo registro, que nos indica que de los especímenes estudiados sólo tenemos un *taxon* distinto.

En el siguiente espécimen, número 4, observamos que el nombre D es revisado al nombre E, por tanto el nombre D pasa a ser sinónimo de E y por consiguiente hay que realizar una revisión de los datos almacenados en la tabla *AlertasSecuenciación* y a su vez actualizar los taxones identificados. Por tanto, las tablas quedarían de la siguiente forma:

<p>tabla <b>taxones</b>: Especimen 4: E.</p> <p>tabla <b>AlertaSecuenciación</b>: Especimen 1: A→E, secuenciación 0. Especimen 1: B→E, secuenciación 0. Especimen 1: C→E, secuenciación 0. Especimen 1: D→E, secuenciación 1. Especimen 2: C→E, secuenciación 1. Especimen 3: B→E, secuenciación 1. Especimen 4: D→E, secuenciación 0.</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figura 95. Resultado de los agentes tras analizar las muestras del ejemplo

Una vez visto el ejemplo de forma general, vamos a utilizar el mismo procedimiento mostrado anteriormente, para ver cuál sería la forma de actuar de los agentes ante una situación real.

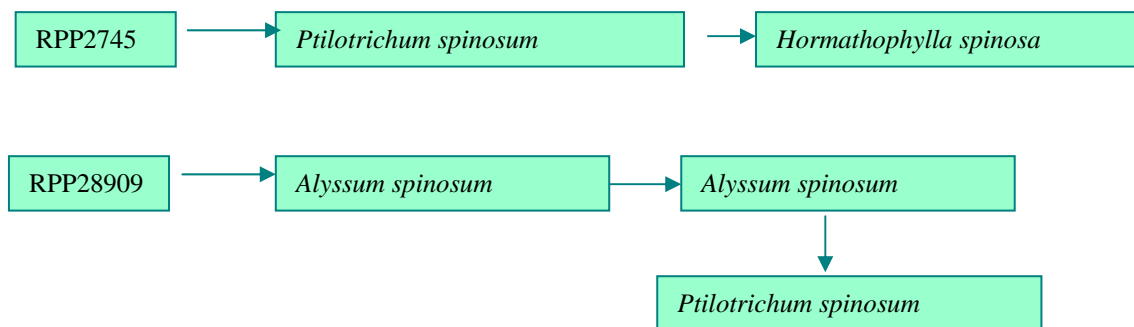


Figura 96. Ejemplo real para analizar su secuencia nominal

La información nomenclatural existente para el identificador RPP2745 es la siguiente:

- Revisión 1: *Ptilotrichum spinosum* (L.) Boiss.
- Revisión 2: *Hormathophylla spinosa* (L.) K pfer.

La informaci n nomenclatural existente para el identificador RPP28909 es la siguiente:

- Revisi n 1: *Alyssum spinosum* L.
- Revisi n 2: *Alyssum spinosum* L.

- Revisión 3: *Ptilotrichum spinosum* (L.) Boiss.

Al actuar el Sistema Multiagente obtendríamos la siguiente situación final:

- Tabla AlertaSecuenciación
  - Identificador RPP28909: *Alyssum spinosum* → *Hormathophylla spinosa*, secuenciación 0.
  - Identificador RPP28909: *Ptilotrichum spinosum* → *Hormathophylla spinosa*, secuenciación 1.
  - Identificador RPP2745: *Ptilotrichum spinosum* → *Hormathophylla spinosa*, secuenciación 0.
- Tabla taxones:
  - Identificador RPP2745: *Hormathophylla spinosa*

Como podemos observar, obtenemos el resultado deseado sin la necesidad de elaborar ninguna tabla que contenga estos sinónimos y que implique un gran trabajo por parte del investigador.

El proceso es sensible a aquellas muestras en las que ha habido un error en el proceso de determinación. Por ello, es posible que la información obtenida por los agentes no sea la esperada. El procedimiento seguido por estos es el correcto, pero los datos almacenados en las bases de datos llevan a no obtener la información. Para esto, se ha incorporado en BioMEN la posibilidad de que el experto retroalimente al sistema indicando si la información obtenida es cierta o no (ver Figura 97).

N° Muestra	Adic	Denominación Actual	Posible Denominación	Provincia	Actual Correcto
GDA28909	1	Alyssum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA10573	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA10574	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA11479	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA11480	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA13274	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA15447	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA18096	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA21445	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA21446	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA25108	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28480	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28481	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28482	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28483	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28484	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28485	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28486	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28487	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28488	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28489	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28490	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>
GDA28492	1	Ptilotrichum spinosum	Hormathophylla spinosa	GR	<input type="checkbox"/>

Figura 97. Revisión por parte del experto de la secuenciación nominal obtenida por los agentes.

Esta información se puede contrastar con los proyectos externos existentes pero, por tratarse de un mecanismo local, se ha permitido que el experto interactúe con el Sistema Multiagente verificando si el conocimiento extraído es correcto o no.

Esta información forma parte de la necesaria para el estudio de la complejidad taxonómica, la riqueza de especies y la orientación de las campañas de recolección. Además, es una herramienta muy valiosa para la limpieza y validación de los datos.

La información necesaria para la Orientación en las Campañas de Recolección está basada en el análisis del material existente en el centro y en el estudio de la secuenciación nominal. Se necesita saber qué especímenes están por debajo de unos mínimos establecidos en un área determinada. Esto permite al centro establecer las políticas de recolección para completar la información existente en las colecciones.



El sistema tiene que informar al usuario de los especímenes necesarios para tener un centro completo y bien representado. La información que el sistema puede proporcionar es:

- Especímenes poco representados y/o bajo mínimos.
- Representar, de forma visual, las zonas menos representadas.



## Capítulo V. Resumen y Conclusiones

El objetivo de esta tesis era construir un sistema de gestión que evitara o disminuyera los problemas encontrados en otros sistemas. Se ha analizado la situación actual para la gestión de colecciones complejas. Tras observar las necesidades de la realización de un sistema *software*, y determinar los problemas existentes, presentamos un sistema MultiAgente que permite gestionar las colecciones biológicas, denominado BioMEN.

BioMEN se presenta como una alternativa, de libre distribución, que permite:

1. Realizar las tareas y procedimientos necesarios para gestionar un centro.
2. Realizar estudios de biodiversidad integrando fuentes externas, y analizando los datos de las colecciones.
3. Personalizar el sistema dependiendo de las necesidades del usuario.

BioMen es un sistema totalmente operacional, que incorpora las siguientes tecnologías:

- Acceso al sistema mediante un Navegador Web.
- *Java™ Servlet technology* (Hall M., 2001; Sun Microsystems Corporation, 2001)
- *Apache Jakarta Project Tomcat*.
- *Apache AXIS* como servidor de servicios Web.
- Transformaciones XSLT para adaptar la interfaz a las necesidades del usuario.

- Plataforma JADE (*Java Agent Development Environment*)
- JDBC para la comunicación con las Bases de Datos.
- PostgreSQL: sistema Gestor de Bases de datos.

Podemos numerar las siguientes ventajas y utilidades de BioMEN:

1. Control de los usuarios del sistema y de las operaciones que éstos pueden realizar.
2. Incorpora todos los mecanismos para almacenar los datos necesarios de los especímenes.
3. Se ha incorporado un módulo de generación de etiquetas que permite incorporar códigos de barras unidimensionales para un tratamiento automatizado de la etiqueta.
4. La generación de códigos de barras multidimensionales en *PDF417* permite asociar toda la información de la muestra en una sola etiqueta.
5. La utilización de códigos de barras *PDF417* permite utilizar la propia etiqueta como mecanismo de copia de seguridad del sistema.
6. Hemos incorporado distintos módulos de consulta para que todos los datos estén accesibles al usuario.
7. Gestión de préstamos entre centros. El alta, modificación y devolución de los préstamos descarga de trabajo al personal del Centro. La solicitud *on-line* permite al usuario tramitar su propio préstamo.
8. Gestión de información multimedia: BioMEN permite asociar todo tipo de información a las muestras para que éstas estén lo más completas posibles en el sistema.

9. Permite gestión de la bibliografía del centro así como asociar la bibliografía utilizada para la determinación de un espécimen.
10. Tratamiento taxonómico y filtrado de los datos proporcionados por el usuario para mantener un mínimo de control.
11. BioMEN incorpora los mecanismos necesarios para que el Centro esté accesible a usuarios externos al Centro. La utilización de páginas Web de consulta para usuarios y los Servicios Web para las aplicaciones dotan a BioMEN de todas las herramientas necesarias para comunicarse.
12. Incorpora las herramientas necesarias para exportar datos en el formato deseado (*Darwin Core, ABCD Schema*).
13. La utilización de servicios externos de visualización espacial permite al usuario examinar la distribución geográfica de los datos almacenados en el sistema.
14. El Sistema Multiagente que incorpora BioMEN permite a su vez:
  - a. Personalizar la Web al usuario teniendo en cuenta sus preferencias.
  - b. Crear un sistema multi-idioma.
  - c. Utilizar un sistema de publicación de noticias de forma personalizada.
  - d. Georreferenciar colecciones. La utilización de bases de datos de topónimos permite asociar una coordenada espacial a una localidad.
  - e. Integrar información de proyectos externos de nomenclatura para obtener información sobre los nombres aceptados y los sinónimos.



## Trabajos Futuros

El acceso al conocimiento almacenado es un proceso complejo y muy estudiado. La gestión del conocimiento siempre ha estado concentrada en la codificación y en la propuesta de infraestructuras para el almacenamiento explícito del conocimiento. Sin embargo, existen problemas para la movilización del conocimiento.

Actualmente, se está potenciando el uso de dispositivos móviles para la toma de datos en el campo. Por tanto, sería necesaria la generación de una herramienta que permita:

- Trabajar con colecciones biológicas paralelas.
- La movilización del conocimiento a través de medios limitados de comunicación y de dispositivos de mano.
- Incorporar información de primera mano: coordenadas Latitud/Longitud de recogida del espécimen.
- Toma de imágenes que permitan observar al espécimen en su hábitat natural.
- Generar un registro de datos virtual independiente de las recolectadas para ser conservadas en el centro.
- Incorporar las herramientas necesarias para disponer de información de las especies existentes en un área concreta, para ser consultada directamente en el campo.





## Glosario

### Glosario de términos botánicos

#### **-B-**

##### **BASE DE DATOS TAXONÓMICAS:**

Base de datos en la que el criterio de organización es estrictamente taxonómico.

**BIODIVERSIDAD:** Variabilidad de organismos vivos en la Tierra. Incluye, tanto los ecosistemas terrestres, aéreos, marinos, acuáticos y otros complejos biológicos, como la diversidad dentro de cada especie, entre las especies y entre los ecosistemas y la diversidad genética.

#### **-C-**

**CINB:** acrónimo del Código Internacional de Nomenclatura Botánica.

**CINZ:** acrónimo del Código Internacional de Nomenclatura Zoológica.

**CLASIFICACIÓN:** Es la elaboración de un sistema lógico que agrupe a los organismos que presentan caracteres comunes. Las especies similares se agrupan en géneros, estos en familias, órdenes, etc.

##### **COLECCIONES BIOLÓGICAS:**

Conjunto de muestras de origen biológico perfectamente conservados, identificados y ordenadas que constituyen un registro permanente de la biodiversidad.

#### **-E-**

**ESPECIE:** Conjunto de individuos semejantes y aislados genéticamente de otros grupos próximos, que se pueden cruzar entre sí dando descendencia fértil. Es la categoría básica sobre la que se construyen las clasificaciones.

**-G-**

**GSD:** Acrónimo de *Global Species Databases*. Sistema utilizado por *Species 2000* para que cada grupo que trabaje en una base de datos de especies específicas.

**-I-**

**ICNB:** acrónimo de International Code of Nomenclature of Bacteria.

**IDENTIFICACIÓN,**

**DETERMINACIÓN:** Reconocimiento de los caracteres de un organismo, a la que se le aplica un nombre que ha sido dado con anterioridad a un organismo similar.

**INDIVIDUO:** Cada ser organizado respecto de la especie a que pertenece.

**-S-**

**SDD:** acrónimo de *Structure of Descriptive Data*. Estándar para información descriptiva.

**SINONÍMIA:** Nombre no aceptado para un *taxon*. Generalmente es un nombre legítimo y válidamente publicado pero posterior al nombre aceptado.

**SISTEMÁTICA:** Ciencia que se ocupa de la clasificación de los seres vivos. Se suele usar como equivalente a Taxonomía.

**SUBESPECIE:** Conjunto de individuos separados del resto de la especie por un conjunto de caracteres heredables, y que están aislados en el tiempo o en el espacio.

**-T-**

**TAXA:** Conjunto de categorías que componen la clasificación de los seres vivos.

**TAXON:** Cada una de las categorías o subdivisiones de la clasificación de los seres vivos, que se ordenan según jerarquías, por ejemplo: especie, género, familia, etc.

**TAXONOMÍA:** Ciencia que se ocupa de la descripción, denominación y clasificación de los seres vivos.

**TDWG:** Un grupo internacional de trabajo en bases de datos taxonómicas.

**-V-**

**VARIEDAD:** Semejante a la subespecie, pero los caracteres que definen la

variación no son heredables de forma constante.

*Glosario de términos Informáticos*

**-A-**

**APRENDIZAJE AUTOMÁTICO:** Subcampo de la Inteligencia Artificial que se ocupa de aquellos sistemas capaces de aprender a partir de la experiencia.

siendo participantes de una federación. Permiten la distribución parcial y controlada de sus datos. No hay un control centralizado, pues cada componente mantiene el control de sus datos.

**-B-**

**BASE DE DATOS DISTRIBUIDA:** Base de datos instalada en un entorno de red en el que sus componentes residen en más de un sistema. También permite el acceso, modificación y actualización de los datos que contiene, desde cualquiera de los sistemas conectados, de manera sincronizada.

**BIOINFORMÁTICA:** es el área interdisciplinar entre biología, computación y ciencias de la información necesaria para la gestión, procesamiento y entendimiento de los datos generados por entidades biológicas, independientemente de la escala de estudio.

**-C-**

**BASE DE DATOS FEDERADA:** Un sistema de base de datos federadas consta de componentes que son autónomos aún

**CGI:** *Common Gateway Interface.* Interfaz de intercambio de datos estándar en Internet a través del cual se organiza el envío de recepción de datos entre

navegadores y programas residentes en servidores Web.

### -G-

**GARP:** es el acrónimo de *Genetic Algorithm for Rule Set Production*. Es un sistema de predicción de distribuciones potenciales de entidades biológicas.

**GIS:** *Geographical Information System*. sistema de Información Geográfica que busca referenciar una base de datos con la cartografía de un territorio, ligando ambos conceptos.

### -R-

**RSS:** es el acrónimo de *Really Simple Syndication*. Sistema usado generalmente para la publicación de noticias.

### -S-

**SIG:** Sistema de Información Geográfica que busca referenciar una base de datos con la cartografía de un territorio, ligando ambos conceptos.

**SISTEMA CLIENTE/SERVIDOR:** Modelo lógico de una forma de proceso cooperativo, independiente de

plataformas hardware y sistemas operativos. El concepto se refiere más a una filosofía que a un conjunto determinado de productos. Generalmente, el modelo se refiere a un puesto de trabajo o cliente que accede mediante una combinación de hardware y *software* a los recursos situados en un ordenador denominado servidor.

**SISTEMA EXPERTO:** sistemas de Inteligencia Artificial que consiguen una capacidad similar a la de un experto en la resolución de problemas mediante la reproducción de un cuerpo de conocimiento.

**SOAP:** es el acrónimo de *Simple Object Access Protocol*. Define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos en XML.

### -U-

**UDDI:** Acrónimo de *Universal Description, Discovery, and Integration*. Protocolo para publicar la información de los Servicios Web.

**-W-**

**WSDL:** Acrónimo de *Web Services Description Language*, describe la interfaz pública a los Servicios Web.

**-X-**

**XML:** Acrónimo de *eXtensible Markup Language*, o lenguaje de anotación extensible. Realmente XML es un lenguaje de metamarcado que se puede utilizar para la especificar lenguajes de marcado.

**XSLT:** Acrónimo de *eXtensible Stylesheet Language Transformation*. Tecnología utilizada para la transformación de documentos.









## **Apéndice A. Estándar para la construcción de sistemas MultiAgentes: la FIPA.**

En este apartado trataremos de describir la arquitectura que propone la FIPA así como la forma de comunicarse entre los distintos Agentes, servicios que propone, etc...

### **1. *Arquitectura propuesta***

Comenzamos con una vista abstracta de la arquitectura que se propone (Foundation for Intelligent Physical Agents, 2001m). La principal función que se inserta es la semántica en el significado del intercambio de mensajes entre agentes los cuales pueden usar diferentes métodos de transporte, diferentes lenguajes de comunicación, o diferentes lenguajes en el contenido del mensaje.

Para el intercambio de mensajes deberemos tener en cuenta:

- Un modelo de servicios y un mecanismo que nos permita descubrir los servicios disponibles para agentes y otros servicios.
- Interoperabilidad en el transporte de mensajes.
- Soportar varias formas de lenguajes de comunicación.
- Varios lenguajes en el contenido del Mensaje.
- Varias representaciones en el directorio de servicios.

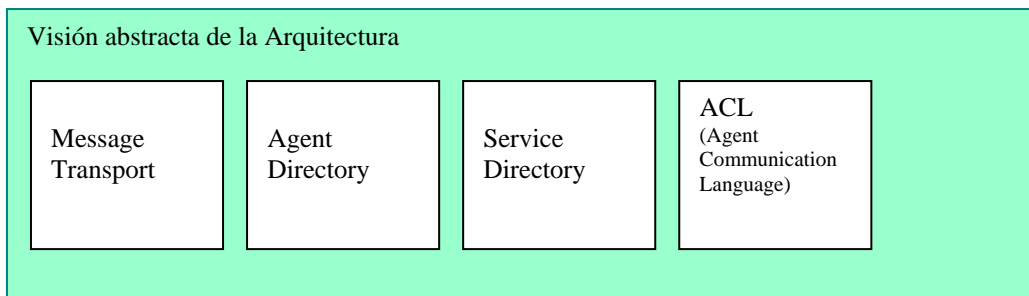


Figura 98. Arquitectura propuesta por la FIPA

Los agentes se comunican mediante el intercambio de mensajes, lo cual representa un acto de diálogo. Estos mensajes son codificados en un lenguaje etiquetado como “*agent-communicator-language*”.

Se proponen un conjunto de servicios, implementados de forma independiente a los agentes para asegurar que el servicio siempre está disponible para los demás agentes:

1. Directorio de Agentes “*Agent-directory-services*”
2. Servicio de Transporte de Mensajes “*Message-Transport-services*”: servicio para el transporte de mensajes entre agentes. Como veremos posteriormente, el agente envía un mensaje hacia un agente y este servicio hace las operaciones necesarias para que le llegue en el lenguaje adecuado.

El directorio de Agentes es un contenedor donde los agentes registran sus características. El directorio de Agentes almacena las entradas de cada agente. Estas entradas (“*agent-directory-entries*”) tienen la siguiente estructura:

- a. Nombre del Agente “*Agent-name*”
- b. Localizador del Agente “*Agent-Locator*”: este campo contiene la forma de localizar al agente teniendo una o más formas registradas mediante estructuras “*transport-descriptions*”.
  - *Transport-descriptions*: describe la forma de comunicarse con el agente: tipo de transporte (“*type-transport*”),

dirección (“*transport-specific-address*”) y una o más propiedades (“*transport-specific-properties*”).

- 1.- El agente se une a uno o más formas de transporte. (delegando en *Message-Transport-Service*).
- 2.- El agente incluye su *agent-directory-entries* y se registra en *agent-directory-service*

Pasos a seguir para registrarse un agente.

Al igual que existe un registro para los agentes existentes en la plataforma, existe también un directorio en el que se registran los servicios que cada agente oferta a la comunidad (“*Service-Directory-Services*”). Mediante este directorio los agentes y servicios pueden describir sus servicios para que puedan ser localizados por otros agentes o servicios.

La estructura de *Service-Directory-entries* consta de:

- *Service-name*
- *Service-type*: distintos tipos de servicios.
- *Service-Locator*: contiene los parámetros necesarios para localizar el servicio, así como la dirección, tipo y un conjunto de atributos asociados al servicio como pueden ser costos del mismo, restricciones del servicio, etc.

La comunicación entre agentes se realiza mediante el paso de mensajes. La FIPA divide el mensaje en tres partes:

- La estructura: es un conjunto de “*key-value-tupla*” y escrito en un “*agent-communication-language*”. El contenido del mensaje está expresado en algún “*content-language*” (p.e. KiF o SL). Entre estas tuplas de claves y valores tenemos información referente al nombre del agente y la ontología usada en el mensaje.
- Representación: cómo se forma el mensaje.

- Transporte: Cuando un mensaje es enviado, es codificado en un paquete (“*payload*”) e incluido en un mensaje de transporte (“*transport-message*”). El paquete se codifica de forma apropiada para el transporte (“*encoding-representation*”).

El mensaje de transporte está compuesto por dos partes: El “*payload*” y el “*envelope*” que incluye el destino y la fuente en el campo “*transport-descriptions*”. La estructura de descripción del transporte contiene información sobre cómo enviar el mensaje (qué transporte, qué dirección, detalle sobre cómo utilizar el transporte, etc.). También puede contener información referente a cómo se ha codificado el mensaje (“*encoding-representation*”), datos relacionados con la seguridad, ...

Un agente es identificado por su *Agent-Name* que es único y no intercambiable. Cada agente tiene uno o más descripciones de transporte las cuales son utilizados por otros agentes para enviar el mensaje de transporte. El conjunto de descripciones de transporte está almacenado en los localizadores del agente dentro del directorio de Agentes.

#### Envío de Mensajes entre Agentes

Como hemos visto anteriormente, el mensaje de transporte puede contener información referente a la seguridad del mensaje. Se reflejan dos tipos de seguridad en el mensaje:

- La validación del mensaje: los mensajes enviados se envían con parámetros de validación.
- Encriptación del mensaje: el mensaje es encriptado para que los no-autorizados no puedan entender el contenido del mismo.

Nombre	Descripción
<i>Agent</i>	Proceso computacional autónomo, funcionalmente comunicativo con un objetivo. Los agentes se comunican usando un lenguaje de comunicación.
<i>Agent-Attribute</i>	Uno de los atributos que pueden contener la estructura “ <i>agent-directory-entries</i> ”. Su objetivo es permitir una búsqueda del agente deseado dentro de “ <i>agent-directory-entries</i> ”. P.e. información sobre costes, características, ontologías que entiende el agente.
<i>ACL (Agent-Communication-</i>	Lenguaje en el cual las comunicaciones pueden ser expresadas y por lo tanto los mensajes construidos.

<i>Language</i> )	
<i>Agent-Directory-entry</i>	Es un conjunto de claves y valores (“ <i>key-value-tuple</i> ”) que contienen por ejemplo el nombre del agente, localizadores, ...
<i>Agent-Directory-Service</i>	Repositorio donde los agentes pueden registrar sus propiedades. Permite a otros agentes buscar información de su interés. Se soportan esquemas jerárquicos o federados para soportar la escalabilidad.
<i>Agent-Locator</i>	Un conjunto de “ <i>transport-description</i> ” los cuales son usados para comunicarse con el agente. Es usado por “ <i>message-transport-services</i> ” para seleccionar el transporte con el cual comunicarse con el agente.
<i>Content</i>	Es la parte de un mensaje. Está expresado en “ <i>content-language</i> ”.
<i>Content-languages</i>	Es el lenguaje usado para expresar el contenido de la comunicación entre agentes. Debe permitir el expresar proposiciones, acciones, términos, ... P.e. Fipa_SL, Fipa_RDF, Fipa-kif, Fipa-CCL.
<i>Encoding-Representation</i>	Una forma de representar un mensaje (“ <i>transport-encoding</i> ”) P.e. XML; <i>Bit-efficient encoding</i> , <i>serialized Java Object</i> . Este es aplicado al componente “ <i>payload</i> ” del mensaje de transporte.
<i>Encoding-service</i>	Es un servicio que permite la codificación de un mensaje. Este servicio también realizaría el servicio inverso, la de-codificación . Algunos servicios que soporta: Codificación y Decodificación: Pasar del mensaje al Payload y viceversa. Consulta sobre la codificación utilizada. Consulta sobre los métodos de codificación disponibles.
<i>Envelope</i>	En un “ <i>key-value-tuple</i> ” que contiene el reparto del mensaje así como la información de la codificación. Contiene descripciones sobre el transporte (“ <i>transport-description</i> ”).
<i>Message</i>	Es la unidad básica de comunicación entre dos o más agentes. Está escrito en un lenguaje de comunicación (ACL), contiene un “ <i>content</i> ”, una ontología, quién envía, uno o más receptores.
<i>Message Transport Service</i>	Es un servicio que soporta el envío y la recepción de “ <i>transport-message</i> ” entre agentes.
<i>Ontology</i>	Provee el vocabulario para representar y comunicar el conocimiento sobre algún tema y un conjunto de relaciones y propiedades que cumplen algunas de las entidades descritas en el vocabulario. Un mensaje tiene un atributo ontología que puede referenciar a uno o más ontologías.
<i>Payload</i>	Mensaje codificado que es incluido en el “ <i>transport-message</i> ”
<i>Service</i>	Conjunto de funciones y mecanismo que dan soporte a operaciones que puedan solicitar agentes y otros servicios.
<i>Service Address</i>	Es la dirección del servicio. Contiene un <i>Service-Type</i> específica la cadena que indica como unirse a un servicio en particular.
<i>Service Attributes</i>	Atributos adicionales en “ <i>service-directory-service</i> ” para un determinado servicio. Estos permiten buscar información sobre los servicios contenidos dentro de “ <i>service-directory-entries</i> ”.
<i>Service Directory Entry</i>	Es un Key-Value-Tupla que contiene “ <i>service-name</i> ”, “ <i>service-type</i> ”, “ <i>service-locator</i> ” y ningún o más “ <i>service-attributes</i> ”.
<i>Service Directory Service</i>	Es usado para registrar y localizar servicios. Contiene p.e. “ <i>message-transport-services</i> ”, “ <i>agent-directory-service</i> ”, ...
<i>Service Identifier</i>	Identificador de servicios. Sólo un nombre por espacio de nombres.
<i>Service Location Description</i>	Conjunto de <i>Key-Value-Tupla</i> que contiene: “ <i>signatura-type</i> ”, “ <i>service-signature</i> ” y un “ <i>service-address</i> ”.
<i>Service Locator</i>	Conjunto de “ <i>Service-Location-Description</i> ” los cuales podrán ser accesibles y hacer uso de un servicio.
<i>Service-Root</i>	Conjunto de “ <i>Service-directory-entries</i> ” disponibles para un agente. Debe de ser proporcionado por lo menos uno para poder formar parte del servicio.

<i>Service Signatura</i>	Firma obligatoria para un servicio.
<i>Service Type</i>	KVT ( <i>Key-Value-Tupla</i> ) que define el tipo de servicio.
<i>Signatura Type</i>	KVT describiendo el tipo de un “ <i>service-signature</i> ”. Permite la interpretación de un “ <i>service-locator</i> ”.
<i>Transport</i>	Es un servicio de entrega de mensajes.
<i>Transport Description</i>	KVT conteniendo “ <i>transport-type</i> ”, “ <i>transport-specific-address</i> ” y ninguno o más “ <i>transport-specific-properties</i> ”. Describe como se puede comunicar con un agente. Contiene: <i>payload</i> y <i>envelope</i> .
<i>Transport Specific Address</i>	Especifica la dirección para un particular “ <i>transport-type</i> ”. La forma de la dirección está especificada por el tipo.
<i>Transport Specific Property</i>	Propiedades asociadas con “ <i>transport-type</i> ”. Son propiedades de apoyo para el “ <i>transport-service</i> ” para construir las conexiones en el transporte.
<i>Transport Specific Type</i>	Describe el tipo de transporte asociado con un “ <i>transport-specific-address</i> ”.

Tabla 10. Conceptos de la FIPA.

Ya tenemos una visión global del funcionamiento de la arquitectura propuesta por la FIPA. A continuación vamos a detallar los distintos protocolos y especificaciones propuestas por la FIPA para los sistemas multiagentes.

## 2. FIPA SL Content Language Specification

Define una sintaxis concreta para la parte de contenido del mensaje (Foundation for Intelligent Physical Agents, 2001a). Tenemos dos partes una de expresiones y otras de fórmulas bien formadas:

- *Fipa SL Content Expresión*: puede ser usado como contenido de un ACL Mensaje. Existen tres tipos:
  - 2.1. Una Proposición: un valor booleano.
  - 2.2. Acción: que puede ser ejecutada.
  - 2.3. Una IRE (*identifying Referente Expressions*): para identificar a un objeto dentro de un dominio.
- Fórmulas bien formadas (“*Well-Formed Formulas*”): formada a partir de una fórmula atómica. Por ejemplo:

(not <Wff>), (and <Wff> <Wff>), (or <Wff> <Wff>), ...

Una fórmula atómica representa una expresión que tiene un valor de verdad en el dominio en el que se encuentra. Tres formas son definidas:

- Un símbolo definido como *false* o *true*.
- Dos términos podrían ser iguales bajo un contexto semántico.
- Un predicado es definido por ningún o más argumentos, los cuales son términos.

Un término está formado por sí mismo: constantes o variables, mediante la construcción recursiva.

Se definen a su vez una serie de operaciones entre los términos, denominadas operaciones de referencia. Algunos ejemplos:

1. *Iota*: define el alcance de una expresión.

i. Dada la base = [ P(a), Q(1,a), Q(1,b) ] (iota ?x (p ?x)) →  
((= (iota ?x (p ?x)) a))

2. *Any*: Operador usado para denotar cualquier objeto que satisfaga la proposición representada por la <fórmula>. Por ejemplo,

((any (sequence ?x ?y) (q ?x ?y))) → ((= (any (sequence ?x ?y) (q ?x ?y)) (sequence 1 a)))

3. *All*: Conjunto de todos los objetos que satisfacen la fórmula. Por ejemplo,

((all ?x (q ?x c))) → ((= (all ?x (q ?x c)) (set))), donde set es el conjunto vacío.

### 3. *FIPA Agent Management Specification*

La gestión de agentes provee un marco en el cual los agentes existen y operan (Foundation for Intelligent Physical Agents, 2001). La gestión de agentes está compuesto por:

1. *Agent*: proceso computacional, autónomo. Funcionalidad para la comunicación. Usa el lenguaje ACL. Es un actor en una plataforma de agentes.
2. *DF (Directory-Facilitator)*: Es una parte opcional. Proporciona un servicio parecido a las páginas amarillas y puede existir más de un DF en una misma plataforma.
3. *Agent Management System (AMS)*. Es obligatorio. Controla y supervisa los accesos y el uso de la plataforma (AP). Sólo uno por AP y mantendrá un directorio de AID (*Agent Identification*). Da un servicio a las páginas Blancas a otros agentes. Todos los agentes tienen que registrarse en el AMS para obtener su AID válido en el AP.
4. *Message Transport Service (MTS)*: mecanismo por defecto para comunicarse entre agentes, independientemente de dónde residan los otros agentes (en distintas APs).
5. *AP (Agent Platform)*: provee una infraestructura física en el cual los agentes pueden ser desarrollados. Gestiona los ciclos de vida de los agentes y registra a éstos en el sistema.



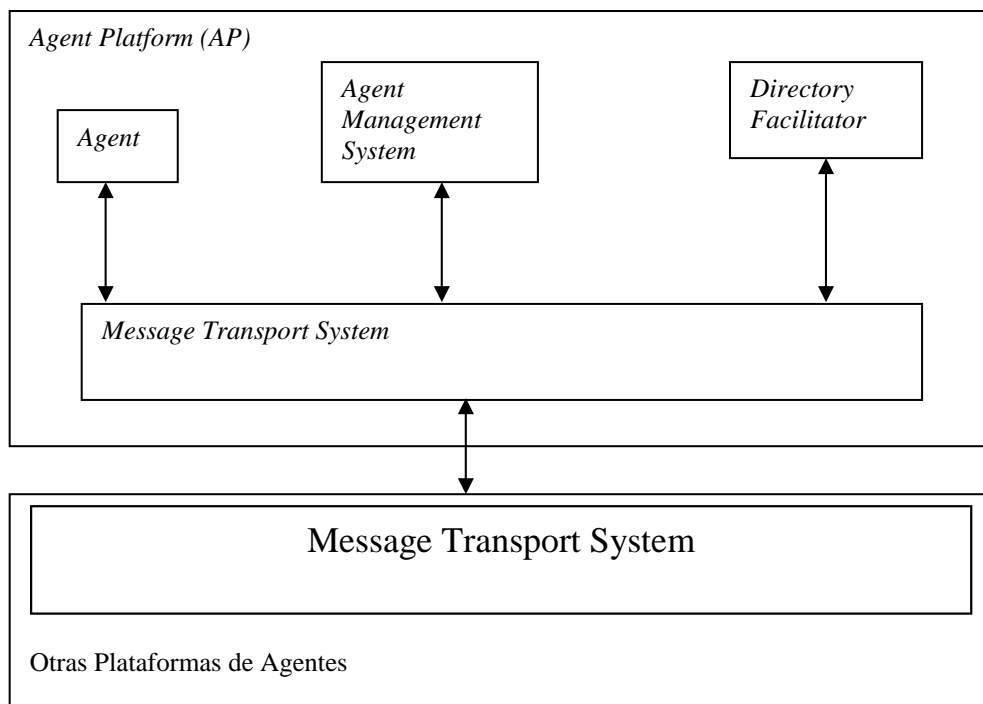


Figura 99. Comunicación entre plataformas.

#### **4. FIPA Request Interaction Protocol Specification**

Este protocolo permite que un agente solicite a otro agente el ejecutar una acción (*fipa-request*) (Foundation for Intelligent Physical Agents, 2001b).

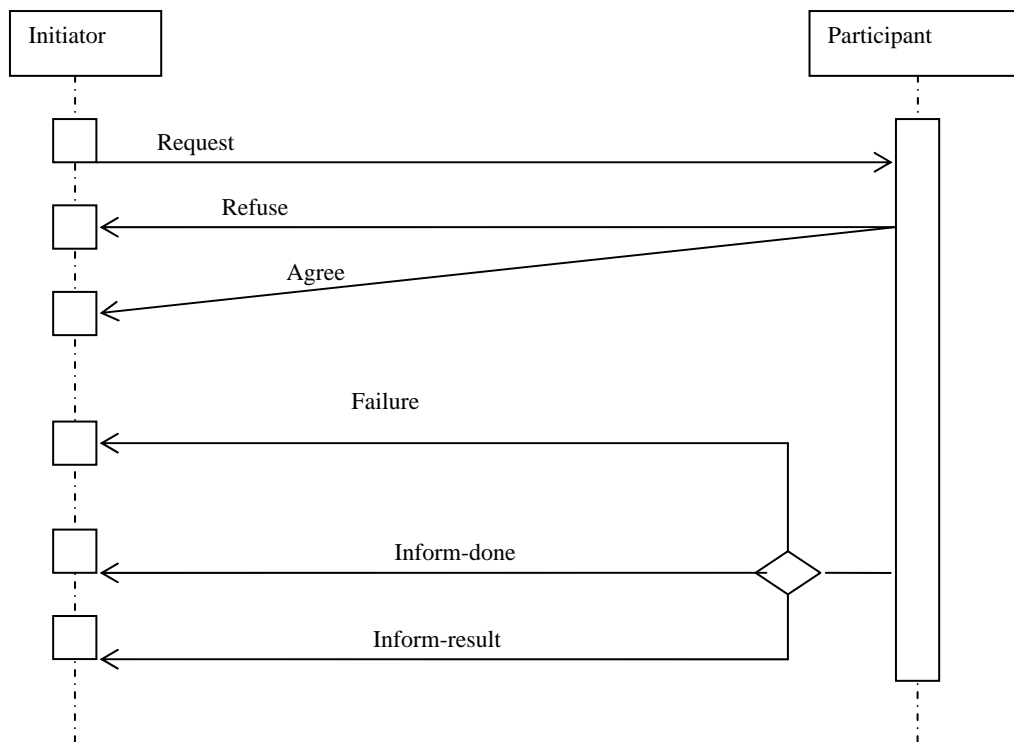


Figura 100. Protocolo de petición de acciones.

Si el iniciador decide suspender la comunicación por cualquier motivo (por ejemplo, porque el otro agente no le entienda) existe otro protocolo de actuación.

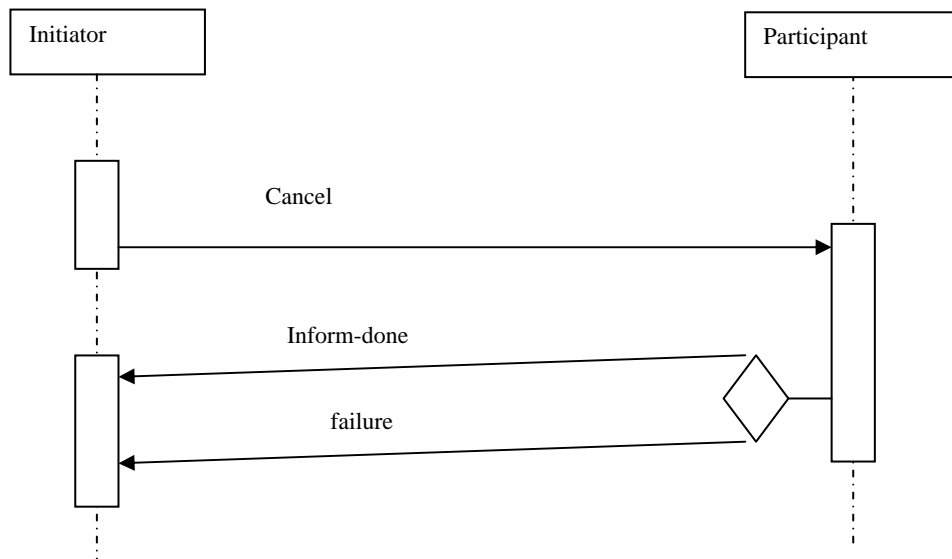


Figura 101. Protocolo de cancelación.

## 5. FIPA Query Interaction Protocol Specification

Protocolo que permite a un agente solicitar o ejecutar alguna acción sobre otro agente (*fipa\_query*) (Foundation for Intelligent Physical Agents, 2001d).

Se definen dos tipos de consultas:

1. *Query-if*: pregunta si una proposición es cierta o no (*true / false*)
2. *Query-ref*: pregunta por algún identificador de objeto.

El protocolo de actuación es el siguiente:

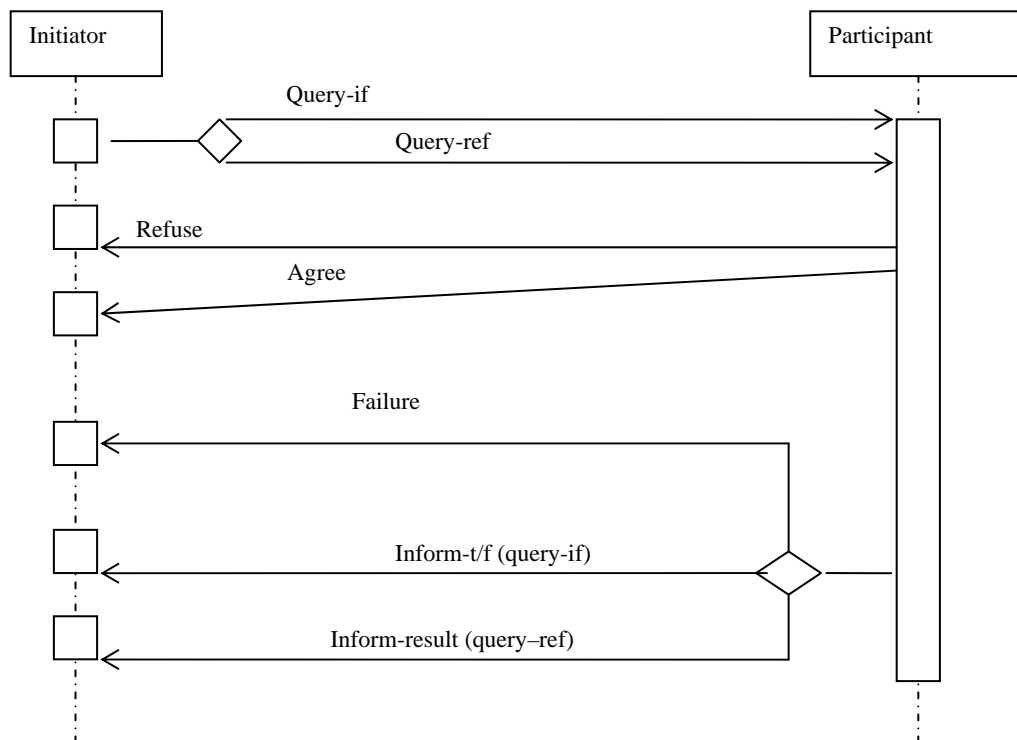


Figura 102. Protocolo de Consulta

## 6. FIPA Contract Net Interaction Protocol Specification

Un agente (initiator) es el gestor y desea que se ejecuten una tarea por uno o más agentes (Foundation for Intelligent Physical Agents, 2001f). Además, la tarea se tiene que realizar optimizando algunas de sus características (por ejemplo, un precio).

El agente lanza una propuesta a los agentes y comienzan los actos de negociación (ver Figura 103).

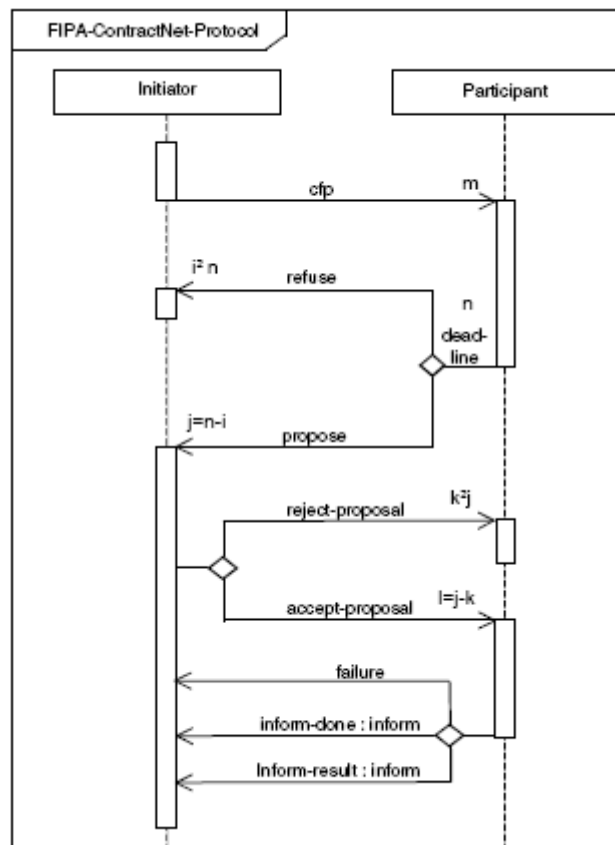


Figura 103. Protocolo de negociación (Foundation for Intelligent Physical Agents, 2001f)

Proceso a seguir:

1. El agente lanza cfp (*call for proposal*) a "m" individuos.

2. “j” agentes responden a la proposición con un conjunto de precondiciones (tiempo, costes, ...)
3. “i” agentes rechazan la propuesta (*refuse*)
4. El agente inicializador estudia las proposiciones para la ejecución y decide a cuál o cuales de ellos ejecutan la proposición.
5. Existe un tiempo máximo para mandar las propuestas.

## **7. FIPA Iterated Contract Net Interaction Protocol Specification**

Modificación o extensión sobre el anterior protocolo “*Contract net*” difiriendo en que se permiten múltiples rondas para la adjudicación de la propuesta (Foundation for Intelligent Physical Agents, 2001e). Se establece un número máximo de rondas por el agente que propone y modifica la cfp (*call for propose*) para adaptar sus proposiciones.

## **8. FIPA Brokering Interaction Protocol Specification**

Diseñado para hacer de intermediario entre sistemas multiagentes (Foundation for Intelligent Physical Agents, 2001h). El intermediario (*broker*) conoce ciertos servicios y los oferta a otros agentes.

Un agente también puede dar un mensaje al *broker* para que se lo entregue a otros agentes que lo conocen.

Si el *broker* no encuentra al agente o los agentes, devuelve un mensaje de error. Pero si el *broker* encuentra a los destinatarios utiliza un protocolo de comunicación distinto con los agentes destino del mensaje.

## **9. FIPA Recruiting Interaction Protocol Specification**

Este protocolo actúa de forma similar al anterior (Foundation for Intelligent Physical Agents, 2001c). En este caso, el agente de reclutamiento no actúa de mediador y no modifica la respuesta de los agentes, enviando directamente los mensajes a los destinos, actuando de forma parecida a un repetidor.

## **10. FIPA Communicative ACT Library (CAL) Specification**

En este apartado vamos a tratar la estructura y la semántica en la forma de comunicarse (Foundation for Intelligent Physical Agents, 2001g).

Los objetivos para definir el CAL:

1. Ayudar a garantizar la interoperabilidad mediante un conjunto de estándares de composiciones y series de actos de comunicación derivados de primitivas de comunicación de la FIPA.
2. Para facilitar la reutilización de complejos actos de comunicación.
3. Para dar un proceso bien-definido para mantener los actos de comunicación y etiquetas de acción en la FIPA ACL.

Acciones definidas en la librería:

1. *Accept Proposal*: cuando se acepta la acción previamente enviada.
2. *Agree*: acción de aceptar la ejecución de una acción.
3. *Cancel*: Petición de un agente sobre otro para cancelar la petición previamente solicitada.

4. *Call for Proposal*: se propone una acción (establecidas unas determinadas precondiciones).
5. *Confirm*: el agente fuente informa al receptor que una determinada proposición es cierta.
6. *Disconfirm*: Informar de que una determinada proposición es falsa, donde el receptor cree que es verdadera.
7. *Failure*: comunicar a otro agente que se intentó realizar la acción pero que ésta falló.
8. *inform*: informa que una proposición es cierta.
9. *Inform-if*: respuesta condicional dada por un agente, respondiendo *true* si cree que es cierta y *false* en caso contrario.
10. *Inform-ref*: se devuelve un objeto como respuesta a una petición.
11. *Not-Understood*: Informa que no entiende lo que tiene que hacer (la acción).
12. *Propagate*: un agente *i* envía un mensaje a un Agente *j*, para que *j* lo propague como si lo enviara el agente *j* (algo parecido a un Proxy). Lo propaga a los destinatarios denotados por el descriptor.
13. *Propose*: acción de enviar una propuesta para ejecutar una cierta acción, dadas unas precondiciones.
14. *Proxy*: el agente *i* quiere que el agente *j* envíe un mensaje a los destinatarios dados en la descripción.
15. *Query-if*: acción para preguntar a otro agente si una proposición es cierta o no.
16. *Query-ref*: acción de preguntar a un agente por los objetos disponibles dada una expresión.

17. *Refuse*: acción de rechazar la ejecución de una acción dando la explicación del por qué.
18. *Reject Proposal*: rechazar alguna proposición a ejecutar o alguna acción durante una negociación.
19. *Request*: Acción de solicitar la realización de alguna acción.
20. *Request when*: Agente i solicita a j ejecutar alguna acción cuando las condiciones sean ciertas.
21. *Request Whenever*: Agente i solicita a un agente j ejecutar una acción siempre que ocurra la condición.
22. *Subscribe*: solicita que se le notifique el valor de un objeto, y que se le informe cada vez que el objeto referenciado cambie.

Un mensaje ACL contiene un conjunto de parámetros. Estos parámetros son necesarios para hacer efectivas las comunicaciones. Los parámetros variarán según la situación:

- El tipo de acción (*Performative*): las vistas anteriormente.
- *Sender, receiver, reply-to, content, language, encoding, ontology, protocol, conversation-id, reply-with, in-reply-to, reply-by*.

## **11. FIPA Agent Message Transport Service Specification**

Mecanismo para transferir mensajes ACL (Foundation for Intelligent Physical Agents, 2001i). Servicio que añade el “*Envelope*” al “*Payload*”. Ofrece mecanismos para:

- Actualizar el *Envelope*.
- Añadir nuevos parámetros en el *Envelope*.



Se utilizará un canal para transmitir el mensaje (ACC: *Agent Communication Chanel*). Este canal interpreta el *Envelope* para determinar cómo, para quién, quién envía el mensaje, ...

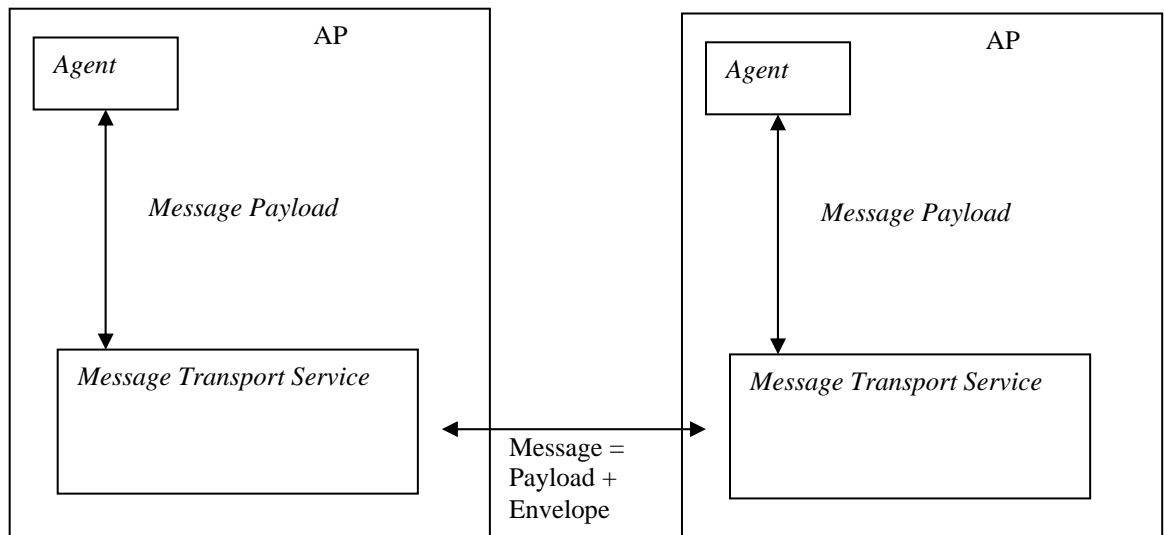


Figura 104. Mensaje entre agentes.

La FIPA define las formas de usar este servicio de Mensajes:

1. Enviado mensaje (tres opciones):
  - a. Agente A envía el mensaje a su local ACC usando una interface estándar. ACC toma el mensaje y lo envía a la ACC correspondiente usando el adecuado MTP.
  - b. Agente A envía directamente el mensaje al ACC donde reside el Agente B (destino).
  - c. El agente A envía directamente el mensaje a B por mecanismos privados (modo no cubierto por la FIPA).
2. Recibiendo el mensaje: El agente recibe *Envelope+Payload*. El agente tienen acceso a toda la información del transporte incluida en el *Envelope* (Encriptación, tipo de acción ACL, ...)

## **12. FIPA Agent Message Transport Protocol for IIOP Specification**

Se define la forma de realizar el transporte de mensajes entre agentes usando IIOP (*Internet Inter-Orb Protocol*) (Foundation for Intelligent Physical Agents, 2001j).

Este MTP está basado en la transferencia de un OMG IDL estructura que contienen el *Envelope* y una secuencia de octetos representando el cuerpo del mensaje ACL. *Envelope+Payload* son transmitidos juntos por IIOP.

## **13. FIPA Agent Message Transport Protocol for http Specification**

Este protocolo incluye el *Envelope* en una solicitud http (Foundation for Intelligent Physical Agents, 2001k). La transferencia http se realiza en dos partes:

1. La fuente hace una solicitud http *request*.
2. Después de recibir los datos, el destino envía un http *response*.

Una petición http (*request*) contiene los siguientes parámetros:

1. *Request live*: POST, identificación mediante URI, versión http.
2. *Request headers*: debe de contener los siguientes parámetros:
  - a. *Content-type*: multipart/mixed y debe de tener un límite.
  - b. *Host*: de la forma hostname y el número de puerto.
  - c. *Cache-control*.
  - d. *Mime-versión*:1.0

- e. Valor opcional de la longitud del cuerpo del mensaje (*request body*).
3. *Request Body*: Contiene el mensaje del agente (*Envelope+Payload*).

Una respuesta http contiene los siguientes parámetros (*response*):

1. *Response Line*: http /1.1
2. *response headers*: *Content-type, cache-control, content-length, etc.*
3. *Response body*: mensaje de respuesta y el content puede ser texto, en formato binario.



## **Apéndice B.      Introducción al Lenguaje XML**

### **1. ¿Qué es XML?**

XML significa *eXtensible markup language*, o lenguaje de anotación extensible. Ya conocemos el lenguaje HTML (*hypertext markup language*), lenguaje de anotación para páginas Web que permite navegación tipo hipertexto; sin embargo, XML no es sólo un lenguaje, es una forma de especificar lenguajes, de ahí lo de extensible. Por lo tanto, XML no es un lenguaje para hacer mejores páginas Web, sino un lenguaje para información auto-descrita.

### **2. Historia de XML.**

XML se inició como un subconjunto de SGML (*Structured Generalized Markup Language*), un estándar ISO para documentos estructurados sumamente complejo para servir documentos en la Web. XML es algo así como SGML simplificado (Melero, 2006).

Este lenguaje tiene gran número de aplicaciones. La mayor parte de los portales y sitios de noticias ya están basados en XML porque permite estructurar la información y posteriormente aplicar de forma sencilla transformaciones para presentarlo. Lo habitual es que la información almacenada en una base de datos se convierta a XML y luego se transforme para servirla al cliente.

### **3. XML bien formado.**

Como lenguaje de anotación, las sentencias en XML consisten en una serie de etiquetas (llamadas elementos) con una serie de modificadores (llamados atributos). Las etiquetas pueden estar anidadas unas dentro de otras, pero toda

etiqueta que se abra se tiene que cerrar, y siempre en el mismo orden. En caso de que un elemento no tenga pareja (por no tener ningún contenido dentro), se le denomina elemento vacío y se indica con un “/” al final. Los elementos se agrupan en documentos, tales como el de la Figura 105 que describe las estancias de una casa.

```
<?xml version="1.0" encoding="iso-8859-1" ?>
<micasa>
  <habitacion id="comedor">
    <mueble>aparador</mueble>
    <mueble>sofá</mueble>
    <puerta a="balcón" />
  </habitacion>
</micasa>
```

Figura 105. Ejemplo básico de XML.

Todos los documentos XML deben estar bien formados, eso significa que se debe cumplir lo siguiente (García Arenas, 2006):

- Si no se utiliza DTD, el documento debe comenzar con una declaración de documento *standalone*, tal como la que se pone en la primera línea del ejemplo.
- Todas las etiquetas deben estar equilibradas: esto es, todos los elementos que contengan datos de tipo carácter deben tener etiquetas de principio y fin.
- Todos los valores de los atributos deben ir entrecomillados.
- Cualquier elemento vacío (por ejemplo, aquellos que no tienen etiqueta final como “<IMG>”, “<HR>”, y “<BR>” y otros de HTML) deben terminar con “/>” o convertirse en “no vacíos” añadiéndoles una etiqueta de fin.
- No debe haber etiquetas aisladas (“<” ó “&”) en el texto (por ejemplo, debe escribirse como “&lt;” y “&amp;”), y la secuencia “]]>” debe escribirse como “]]&gt;”.
- Los elementos deben anidar dentro de sí sus propiedades.

- Los ficheros bien-formados sin DTD pueden utilizar atributos en sus elementos, pero éstos deben ser todos del tipo CDATA, por defecto. El tipo CDATA (*character DATA*) son caracteres.
- Los nombres de las etiquetas pueden ser alfanuméricos, comenzando con una letra, e incluyendo los caracteres “-“ y “:”, aunque este último tiene un significado especial.

En un documento XML, aparte de elementos y atributos, puede haber otros elementos como:

- Entidades. Representan símbolos "atómicos" que habitualmente deben ser entendidos por el navegador. Las entidades van encerradas entre los símbolos “&” y “;”. La Tabla 11 muestra algunos ejemplos de entidades XML.

ENTIDAD	CARACTER
&amp;	&
&lt;	<
&gt;	>
&apos;	'
&quot;	"

Tabla 11. Entidades XML.

- Comentarios, que se procesan de forma diferente al texto, y que, tal como en HTML, van precedidos por “<!-- “ y acaban con “-->”.S
- Secciones CDATA. Sirven para extraer del documento XML una sección que va a ser interpretada tal cual, sin hacer ninguna modificación. Puede servir, por ejemplo, para incluir HTML "mal-formado" dentro de un documento XML. La Figura 106 incluye todos los elementos descritos.

```

<?xml version="1.0" encoding="iso-8859-1"?>
<!-- Descripción de los elementos de una casa soñada -->
<micasa>
  <habitacion id="comedor">
    <mueble>aparador</mueble>
    <mueble>sofá "de época"</mueble>
  </habitacion>
  <habitacion id="cocina">
    <mueble><![CDATA[
<p>En la pared de la derecha hay un frigorífico <p>Y en
la de la izquierda, sólo mugre
]]></mueble>
    <mueble>fregadero</mueble>
  </habitacion>
</micasa>

```

Figura 106. Ejemplo de documento XML completo.

#### 4. XML namespaces

Si todo el mundo definiera sus etiquetas, un documento acabaría siendo un caos de diferentes etiquetas procedentes de diferentes sitios, y, lo que es peor, de etiquetas con el mismo nombre que, en realidad, significan cosas diferentes. El concepto de espacios de nombres (*namespaces*) permite dividir el conjunto de todos los nombres posibles, de forma que se pueda definir a qué zona de ese espacio corresponde una etiqueta. De este modo, etiquetas con el mismo nombre, pero definidas por dos autores diferentes, pueden diferenciarse en el espacio de nombres. El espacio de nombres no es esencial en todos los documentos, pero resulta útil cuando se usan etiquetas de diferente procedencia (por ejemplo, etiquetas nuevas dentro de un documento XML), o etiquetas que se quieren procesar de forma diferente. El espacio de nombres de una etiqueta se indica con un prefijo y “:: <namespace:etiqueta” (ver Figura 107).

```

<mc:micasa xmlns:mc="http://www.geneura.org/micasa">
  <mc:habitacion mc:id="comedor">
    <mc:mueble>aparador</mc:mueble>
    <mc:mueble>sofa de época</mc:mueble>
  </mc:habitacion>
</mc:micasa>

```

Figura 107. Ejemplo de documento XML con namespaces.



En este documento, se utiliza la primera línea para declarar el prefijo del espacio de nombres mediante el atributo “xmlns” (*XML namespace*). En este caso, hemos elegido el prefijo “mc”. El espacio de nombres tiene que tener asignado un URI (*Universal Resource Identification*), que es un identificador único en el documento.

Un documento XML puede tener tantos espacios de nombres como se quieran declarar, y se pueden mezclar elementos de diferentes espacios de nombres, e incluso sin ningún espacio

## 5. XML y diccionarios de datos.

En algunos casos es necesario validar que un documento XML es correcto, es decir, que las etiquetas que se usan son correctas y que están anidadas de la forma adecuada. Para ello se pueden utilizar dos herramientas:

- DTD, o *data type dictionary*. El DTD de la Figura 108 indica que es una habitación tiene uno o varios muebles, y una o varias puertas (que se indica con “+”); a su vez, “micasa” puede tener una o más habitaciones, y cada uno de los elementos pueden tener los atributos que se indican con la sentencia ATTLIST. Como se puede ver, no se trata de XML, aunque se parezca.

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT mc:habitacion (mc:mueble+)>
<!ATTLIST mc:habitacion mc:id CDATA #REQUIRED>
<!ELEMENT mc:micasa (mc:habitacion)*>
<!ATTLIST mc:micasa xmlns:mc CDATA #REQUIRED>
<!ELEMENT mc:mueble (#PCDATA)>
```

Figura 108. Ejemplo de DTD.

- *XML Schema*, el equivalente a un DTD, pero en XML. Un XML Schema describe la sintaxis correcta de un documento XML. El esquema de la Figura 109 es el equivalente al DTD del ejemplo anterior.

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
  elementFormDefault="qualified">
  <xs:complexType name="habitacionType">
    <xs:sequence>
      <xs:element name="mueble" ref="mueble" minOccurs="1" maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="id" type="xs:string" use="required"/>
  </xs:complexType>
  <xs:element name="micasa">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="habitacion" type="habitacionType"/>
      </xs:sequence>
      <xs:attribute name="mc" type="xs:anyURI" use="required"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="mueble" type="xs:string"/>
</xs:schema>
```

Figura 109. Ejemplo de esquema XML.

## 6. Lectura y manipulación de un documento XML.

El W3C ha especificado dos normas para trabajar con documentos XML. Esta especificación permite que cualquier lenguaje de programación pueda navegar a través de un documento XML y manipularlo.

La importancia de estas normas tiene una gran repercusión. Si cada lenguaje de programación implementase su vía especial para navegar a través de los documentos XML, sería caótico: ciertos lenguajes sólo permitirían realizar ciertas operaciones y no habría consistencia alguna. Las normas creadas por el W3C aportan una interfaz independiente del lenguaje de programación en todos los documentos XML bien formados. Ésta interfaz es conocida como una *API* (Application Program Interface, interfaz de programación de la aplicación). Estas *API* describen una jerarquía de objetos, con métodos y atributos que simplifican las tareas relativas al recorrido y acceso a las partes del documento. Estos dos mecanismos se denominan *SAX* (Simple API for XML Parsing) y *DOM* (Document Object Model).

- SAX se utiliza para hacer un recorrido secuencial de los elementos del documento XML. Por este motivo, utiliza mucha menos memoria que DOM, es más rápido porque no invierte tiempo construyendo el árbol y porque es escalable a cualquier cantidad arbitraria de información.
- DOM implica la creación de un árbol en memoria que contiene todo el documento XML. Un nodo DOM de nivel 1 contiene 9 punteros a otros nodos. Cada puntero ocupa 4 bytes lo que hace un total de 36 bytes de punteros por nodo. Además de esta, contiene otros tipos de información como el nombre y el tipo del nodo.



# Bibliografía

1. Alroy, J., *How many named species are valid?* Proc. Natural Academic Science USA, 2002. **99**: p. 3706-3711.
2. Araya, S., Silva M., and Weber R., *A methodology for web usage mining and its application to target group identification.* Fuzzy Sets and Systems, 2004. **148**(1): p. 139-152.
3. Balmin, Y. and Papakonstantinou, Y., *Incremental Validation of XML Documents.* ACM Transactions on Database Systems, 2004.
4. Beaman, R. and Conn, B., *Automated geoparsing and georeferencing of Malesian collection locality data.* Telopea, 2003. **10**: p. 43-52.
5. Bedathur, S.J. and Haritsa, J.R. *Design and implementation of a biodiversity information management system.* in *International Conference on Management of Data.* 2000: Proceedings of the Tenth International Conference on Management of Data.
6. Berendsohn, W., *A Common Datastructure for European Floristic Databases (CDEFD).* Biotechnology, 1997a(Hoeveler, A. & Cresti, M.): p. 251-253.
7. Berendsohn, W., *A taxonomic information model for botanical databases: The IOPI model.* Taxon, 1997b. **46**(2): p. 283-309.
8. Berendsohn, W., *Resource Identification for a Biological Collection Information Service in Europe (BioCISE).* . 2000, Berlin. 76.
9. Berendsohn, W., *Contributions to Global Change Research: A Report by the German National Committee on Global Change Research.* Biodiversity Informatics., 2001(Bonn): p. 89-94.
10. Berendsohn, W. *Software for Biological Collection Management.* TDWG Subgroup on Accession Data 2006 February 28, 2006 [cited 10/04/2006]; Available from: <http://www.bgbm.fu-berlin.de/TDWG/acc/Software.htm>.
11. Berendsohn, W., Döring, M., Gebhardt, M., and Güntsch, A. *BioCASE - A Biological Collection Access Service for Europe.* in *Trends and Developments in Biodiversity Informatics.* 2002. Indaiatuba, SP, Brasil 2002.

12. BioGeoMancer. *BioGeoMancer*. 2006 [cited 04/05/2006]; Available from: <http://georef.peabody.yale.edu/>.
13. Biótica. *Sistema de Información Biótica*. 2005 23/01/2004 [cited 2006 01/04/2006]; Available from: [http://www.conabio.gob.mx/informacion/biotica\\_espanol/doctos/distribuv4.1.html](http://www.conabio.gob.mx/informacion/biotica_espanol/doctos/distribuv4.1.html).
14. Bridson, D. and Forman, L., *The Herbarium Handbook*. 1992: Royal Botanic Gardens Kew.
15. Brisby, F., *The Quiet Revolution: Biodiversity Informatics and the Internet*. Science, 2000. **289**: p. 2309-2312.
16. Castroviejo et al., *Flora ibérica : plantas vasculares de la Península Ibérica e Islas Baleares.*, ed. Real Jardín Botánico, C. Vol. IV. 1996.
17. Cedric Raguenaud, Jessie Kennedy, and Barclay, P.J., *Database Support for Taxonomy*, in *Prometheus technical report #1*, Computing, S.o., Editor. 1999, Napier University.
18. Claudia E. Moreno, *Métodos para medir la biodiversidad*, ed. SEA, M.T.-M.T. Vol. 1. 2001.
19. Colwell, R.K. *Biota 2: The Biodiversity Database Manager*. 2004 [cited 2006 10/04/2006].
20. Commonwealth, S. and Organisation, I.R. *Biolink*. 2005 27/10/2005 [cited 2005 15/10/2005]; Available from: <http://www.biolink.csiro.au/>.
21. Dallwitz, M.J., *A flexible computer program for generating identification keys*. Systematic Zoology, 1974. **1**: p. 50-57.
22. DiGIR. *DiGIR: Distributed Generic Information Retrieval*. 2006 9/6/2005 [cited 01/04/2006]; Available from: <http://digir.sourceforge.net>.
23. Döring, M. *A first TAPIR implementation - the BioCAsE PyWrapper serves a new protocol*. in *Taxonomic Databases Working Group 2005 Annual Meeting*. 2005. St Petersburg, Russia.
24. Döring, M., et al. *The integration of DiGIR and BioCAsE*. in *Meeting Taxonomic Databases Working Group (TDWG)*. 2004. Christchurch, New Zealand 2004.
25. E. Mata Montero. *Informatización del conocimiento para la conservación de la biodiversidad*. Ecosistemas 2006 [cited 2006 20/05/2006]; 1-8]. Available from: [www.revistaecosistemas.net](http://www.revistaecosistemas.net).

26. ENHSIN. *European Natural History Specimen Information Network*. 1999 [cited 12/04/2006]; Available from: <http://www.bgbm.org/BioDivInf/projects/ENHSIN/>.
27. Entomología. *Qué es la biodiversidad?* 2006 [cited 23/05/2006]; Available from: <http://www.geocities.com/CollegePark/Classroom/7370/pagina5.htm>.
28. Fermanian, T. and Michalski, R.S., *Weeder: An Advisory System for the Identification of Grasses in Turf*. *Agronomy Journal*, 1989. **81(2)**: p. 313-316.
29. Filer, D.L., *BRAHMS - Botanical Research and Herbarium Management Systems. A pocket introduction and Demostration Guide*. Oxford Forestry Institute, 1994: p. 28.
30. Foundation for Intelligent Physical Agents. *FIPA SL Content Language Specification*. 2001a [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00008/index.html>.
31. Foundation for Intelligent Physical Agents. *FIPA Request Interaction Protocol Specification*. 2001b [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00026/index.html>.
32. Foundation for Intelligent Physical Agents. *FIPA Recruiting Interaction Protocol Specification*. 2001c [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00034/index.html>.
33. Foundation for Intelligent Physical Agents. *FIPA Query Interaction Protocol Specification*. 2001d [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00027/index.html>.
34. Foundation for Intelligent Physical Agents. *FIPA Iterated Contract Net Interaction Protocol Specification*. 2001e [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00030/index.html>.
35. Foundation for Intelligent Physical Agents. *FIPA Contract Net Interaction Protocol Specification*. 2001f [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00029/index.html>.
36. Foundation for Intelligent Physical Agents. *FIPA Communicative Act Library Specification*. 2001g [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00037/index.html>.
37. Foundation for Intelligent Physical Agents. *FIPA Brokering Interaction Protocol Specification*. 2001h [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00033/index.html>.

38. Foundation for Intelligent Physical Agents. *FIPA Agent Message Transport Service Specification*. 2001i [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00067/index.html>.
39. Foundation for Intelligent Physical Agents. *FIPA Agent Message Transport Protocol for IIOP Specification*. 2001j [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00075/index.html>.
40. Foundation for Intelligent Physical Agents. *FIPA Agent Message Transport Protocol for HTTP Specification*. 2001k [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00084/index.html>.
41. Foundation for Intelligent Physical Agents. *FIPA Agent Management Specification*. 2001l [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00023/index.html>.
42. Foundation for Intelligent Physical Agents. *FIPA Abstract Architecture Specification*. 2001m [cited 20/05/2005]; Available from: <http://www.fipa.org/specs/fipa00001/index.html>.
43. García Arenas, M.I. *Curso XML 1ª Edición*. 2006 [cited 05/12/2005]; Available from: <http://geneura.ugr.es/~maribel/xml/introduccion/index.shtml>.
44. GBIF. *Global Biodiversity Information Facility*. 2001 [cited 23/03/2006]; Available from: [www.gbif.org](http://www.gbif.org).
45. Gibaja, E., *Modelos de Representación del Conocimiento para la Identificación Taxonómica y Aplicaciones.*, in *Dpto. Ciencias de la Computación e I. Artificial*. 2004, Universidad de Granada: Granada, España. p. 321.
46. Gonzalo Halffter, Claudia E. Moreno, and Eduardo O. Pineda, *Manual para evaluación de la biodiversidad en Reservas de la Biosfera*, ed. MT&SEA. Vol. 2. 2001.
47. Govindaraj, T., *Manufacturing and Networked Information Systems for Life Support*. Darwish M. K. Al Gobaisi ed, ed. (EOLSS), E.o.L.S.S. 2001.
48. H. Penny Nii., *Blackboard Systems: Blackboard Application Systems, Blackboard Systems from a Knowledge Engineering Perspective (part two)*. . The AI Magazine, 1986a. 7(3): p. 82-106.
49. H. Penny Nii., *Blackboard Systems: The Blackboard Model of Problem Solving and the Evolution of Blackboard Architectures*. . The AI Magazine, 1986b. VII(2): p. 38-53.
50. Halffter, G. and Ezcurra, E., *¿Qué es la Biodiversidad?*, in *La diversidad biológica de Iberoamérica I.*, Mexicana, A.Z., Editor. 1992. p. 1-389.



51. Hall M., *Core Servlets and JavaServlet pages.*, ed. Englewood Cliffs, N.S.M.P.P.H. 2001.
52. Hayes-Roth, A *Blackboard Architecture from Control.* Artificial Intelligence. Vol. 3. 1985. 251-321.
53. Herbario de la Universidad de Granada. *Bases de datos del Herbario de la Universidad de Granada.* 2006 [cited 06/04/2006]; Available from: <http://drimys.ugr.es>.
54. IEEE Foundation for Intelligent Physical Agents. *The Foundation for Intelligent Physical Agents.* 2006 [cited 2006 12/05/2006]; Available from: <http://www.fipa.org/>.
55. Index, I.P.N., *International Plant Names Index.* 2004.
56. Integrated Taxonomic Information System, *ITIS: Integrated Taxonomic Information System.* 2006.
57. James L. Edwards, Meredith A. Lane, and Ebbe S. Nielsen, *Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop.* Bioinformatics for Biodiversity, 2000. **289**: p. 2312-2314.
58. Jennings, M.D., *Natural terrestrial cover classification: Assumptions and definitions.* Gap analysis technical, U.S. Fish and Wildlife Service, Moscow, ID., 1993. **2**.
59. John D. and Catherine T. MacArthur Foundation. *Virtual Herbarium Express: A Tool for Capturing and Sharing Herbarium Specimen Information Electronically.* 1996 [cited 2006 12/04/2006]; Available from: <http://www.nybg.org/bsci/vh/>.
60. Krishtalka, L. and Humphrey, P.S., *Can Natural History Museum Capture the Future?* BioScience, 2000. **50**: p. 611-617.
61. Lawrence G.H.M., *Taxonomy of vascular plants.* 1951: The Macmillan Company, New York. 228-262.
62. Lesk, A.M., *Introduction to Bioinformatics,* ed. Cambridge, U.o. 2002: 14/02/2002. 320.
63. M. Delgado, et al. *Informática para la Biodiversidad.* Curso de Formación Continua de la Universidad de Granada 2006 [cited 20/05/2006]; Available from: [http://continua.ugr.es/data\\_html/cursomas.php?cod=06/CA/045&nombre=INFORMATICA%20PARA%20LA%20BIODIVERSIDAD%20\(1%AA%20ED\)](http://continua.ugr.es/data_html/cursomas.php?cod=06/CA/045&nombre=INFORMATICA%20PARA%20LA%20BIODIVERSIDAD%20(1%AA%20ED)).

64. Melero, J. *Tutorial de XML*. 2006 10/10/2004 [cited 13/12/2005]; Available from: <http://geneura.ugr.es/~jmerelo/xml/>.
65. Minnigerode, M.D. *Tracy: a Herbarium Management System*. 1997 [cited 01/05/2006]; Available from: <http://www.csd.tamu.edu/FLORA/input/inputsys.html>.
66. Mobasher B., Dai, H., Luo, T., and Nakagawa, M., *Discovery and Evaluation of Aggregate Usage Profiles for web Personalization*. Data Mining and Knowledge Discovery, 2002. **6**: p. 61-82.
67. Morse, L.E., *Computer aids to plant identification*. American Journal of Botany, 1970. **57**(6): p. 754-760.
68. Nicholas M Luscombe, Dov Greenbaum, and Mark Gerstein, *What is bioinformatics? A proposed definition and overview of the field*. Method Inform, 2001. **4**: p. 346-358.
69. Oliver, I., et al., *Virtual biodiversity assessment systems*. BioScience, 2000. **50**: p. 441-449.
70. Pando, F. *El Herbario de Criptógamas del Real Jardín Botánico y sus bases de datos*. in *IX Simposio Nacional de Botánica Criptogámica*. 1991. Salamanca.
71. Pando, F. *Herbar: una aplicación en MS-Access para la gestión de herbarios*. 1996 [cited 04/04/2006]; Available from: <http://www.gbif.es/herbar/herbar.php>.
72. Pankhurst, R.J., *A computer program for generating diagnostic keys*. Computer Journal, 1970. **13**(2): p. 145-151.
73. Pankhurst, R.J., *Practical Taxonomic Computing*. Cambridge University Press. 1991.
74. Paul Resnick, et al. *GroupLens: An Open Architecture for Collaborative Filtering of Netnews*. in *Conference on Computer Supported Cooperative Work*. 1994.
75. Pressman, R., *Ingeniería del Software.*, ed. Hill, M. 1997.
76. Pullan, M.R., et al., *The Prometheus Taxonomic Model: a practical approach to representing multiple classifications*. Taxon, 2000. **49**(1): p. 55-75.
77. Robert Guralnick and David Neufeld, *Challenges building online GIS Services to Support global biodiversity mapping and Analysis: lessons from the mountain and plains database and informatics project*. Biodiversity Informatics., 2005. **2**: p. 56-69.

78. Scott, M., Tear, T., and Davis, F., *Gap Analysis. A Landscape approach to biodiversity planning*. Bethesda, MD: The American Society for Photogrammetry and Remote Sensing. 1996.
79. Soberón, J.M. and Peterson, A.T., *Biodiversity informatics: Managing and applying primary biodiversity data*. . Phil. Trans. Roy. Soc. B, 2004. **359**: p. 689-698.
80. Species 2000. *Species 2000*. 2005 [cited 01/05/2006]; Available from: <http://www.species2000.org>.
81. Specify. *Specify Software Project*. 2005 [cited 10/04/2006]; Available from: <http://www.specifysoftware.org/>.
82. Stanley D. Blum. *An Overview of Biodiversity Informatics*. California Academy of Sciences [HTML] 2000 07/10/2000 [cited 07/3/2006]; Available from: [http://www.calacademy.org/research/informatics/sblum/pub/biodiv\\_informatics.html](http://www.calacademy.org/research/informatics/sblum/pub/biodiv_informatics.html).
83. Stockwell, D.R.B. *Overview of Computational Biodiversity Research*. 1997 01/04/2006 [cited 01/03/2006]; Available from: <http://biodi.sdsc.edu/Doc/BIS/overview.html>.
84. Stockwell, D.R.B. and Peters, D.P., *The GARP modelling system: problems and solutions to automated spatial prediction*. International F. Geographic Information System., 1999. **13**: p. 43-158.
85. Suarez, A.V. and N.D. Tsutsui, *The value of museum collections for research and society*. BioScience, 2004. **54**: p. 66-74.
86. Sun Microsystems Corporation. *Servlet Specification v 2.4*. 2001 [cited 26/04/2006]; Available from: <http://www.sun.com/servlet/>.
87. Taxonomic Databases Working Group. *TAPIR Protocol: TDWG Access Protocol for Information Retrieval*. 2004 [cited 15/05/2005]; Available from: <http://ww3.bgbm.org/protocolwiki/>.
88. Taxonomic Databases Working Group. *ABCD Schema - Task Group on Access to Biological Collection Data*. 2006 [cited 20/05/2006]; Available from: <http://www.bgbm.org/TDWG/CODATA/Schema/>.
89. TDWG. *Desarrolladores del estándar Darwin Core*. 2005 [cited 01/04/2006]; Available from: <http://darwincore.calacademy.org>.
90. the Apache Software Foundation. *The Apache Projects*. 2006 [cited 12/04/2006]; Available from: <http://www.apache.org>.

91. Tormo Molina, R. *Lecciones Hipertextuales de Botánica*. Proyecto de Innovación Educativa 2006 [cited 2006 23/03/2006]; Available from: <http://www.biologie.uni-hamburg.de/b-online/ibc99/botanica/botanica/presenta.htm>.
92. U.M. Fayyad, *Data mining and knowledge discovery: making sense out of data*. IEEE Expert, Intelligent Systems and their Applications, 1996: p. 20-25.
93. Unidad de Coordinación de GBIF en España. *GBIF.ES - Software. Software avalado por la Unidad de Coordinación de GBIF.es*. 2006a [cited 06/04/2006]; Available from: <http://www.gbif.es/software.php>.
94. Unidad de Coordinación de GBIF en España. *GBIF.ES - Recurso. Estructuras y contenidos para el intercambio de datos. Darwin Core*. 2006b [cited 02/01/2006]; Available from: <http://www.gbif.es/Recursos.php#Formatos>.
95. Unidad de Coordinación de GBIF en España. *GBIF.ES - Información general sobre GBIF*. 2006c [cited 02/01/2006]; Available from: <http://www.gbif.es/gbif.php>.
96. University of California, B. *DLI2 Annual Report 4/1/99 - 3/31/00*. 2000 01/04/2006 [cited Digital Library Project; Available from: <http://elib.cs.berkeley.edu/admin/rpts/DLI2/AR2000/index.html>.
97. University of California Davis. *University of California Davis Herbarium Management System*. 2004 Febrero, 2004 [cited 2006 01/05/2006]; Available from: <http://herbarium.ucdavis.edu/database.html>.
98. University of Florida Herbarium. *PLabel: Herbarium Label Program*. 2000 17/10/2005 [cited 03/05/2006]; Available from: <http://www.flmnh.ufl.edu/herbarium/pl/>.
99. University of Oxford. *BRAHMS. Botanical Research And Herbarium Management System*. 1985 Marzo, 2006 [cited 2006 04/04/2006]; Available from: <http://www.brahmsonline.com/>.
100. WWS. *February 2005 Web Server Survey Finds 40 Million Sites on Apache*. NetCraft 2005 01/05/2006 [cited 01/05/2006]; Available from: [http://news.netcraft.com/archives/2005/02/01/february\\_2005\\_web\\_server\\_survey\\_finds\\_40\\_million\\_sites\\_on\\_apache.html](http://news.netcraft.com/archives/2005/02/01/february_2005_web_server_survey_finds_40_million_sites_on_apache.html).
101. Y.H. Cho and L.K. Kim, *Application of Web usage minig and product taxonomy to collaborative recommendations in e-commerce*. Expert Systems with Applications, 2004. **26**: p. 233-246.