



Proyecto

**Soporte informático con el fin de desarrollar un prototipo de la
Plataforma para la Gestión de Conocimiento e Información Nacional
sobre Biodiversidad (PGCIB)**

**Documentación del primer prototipo del sistema
(versión 2)**

Documento elaborado por:

M.Sc. María Auxiliadora Mora Cross
M.Sc. Manuel Fernando Vargas Del Valle

18 de marzo de 2018

Contenidos

Siglas y acrónimos	2
Resumen ejecutivo	3
1. Introducción	4
2. Descripción del prototipo	6
3. Portal de datos	6
3.1. Arquitectura y funcionalidad	8
3.1.1. Características generales	8
3.1.2. Búsqueda de registros de presencia (Biocache)	10
3.1.3. Búsqueda de registros de colecciones, instituciones y conjuntos de datos (Collectory)	15
3.2. Herramientas informáticas utilizadas	18
3.3. Requerimientos para la instalación	19
4. Portal de comunicación	21
4.1. Características	21
4.2. Herramientas informáticas utilizadas	21
4.3. Requerimientos para la instalación	22
5. Glosario	23
6. Referencias	27
7. Apéndices	28
Apéndice I: Miembros del equipo de análisis de requerimientos de los usuarios del sistema (Earus)	28

Siglas y acrónimos

ACG	Área de Conservación Guanacaste
ALA	Atlas de la Biodiversidad de Australia (<i>Atlas of Living Australia</i>)
API	Interfaz de programación de aplicaciones (<i>Application programming interface</i>)
Ceniga	Centro Nacional de Información Geoambiental
Conagebio	Comisión Nacional para la Gestión de la Biodiversidad
CSS	Hojas de estilo en cascada (<i>Cascading Style Sheets</i>)
Earus	Equipo de análisis de requerimientos de los usuarios del sistema
GBIF	Infraestructura Mundial de Información en Biodiversidad (<i>Global Biodiversity Information Facility</i>)
GEF	Fondo Mundial para el Medio Ambiente (<i>Global Environmental Facility</i>)
HTML	Lenguaje de Marcas de Hipertexto (<i>HyperText Markup Language</i>)
ITCR	Instituto Tecnológico de Costa Rica
IU	Interfaz de usuario
Minae	Ministerio de Ambiente y Energía
Micitt	Ministerio de Ciencia, Tecnología y Telecomunicaciones
MNCR	Museo Nacional de Costa Rica
PGCIB	Plataforma para la Gestión de Conocimiento e Información Nacional sobre Biodiversidad
PNUD	Programa de las Naciones Unidas para el Desarrollo
Sinac	Sistema Nacional de Áreas de Conservación
Sinia	Sistema Nacional de Información Ambiental
TDWG	Estándares de Información sobre Biodiversidad (<i>Biodiversity Information Standards</i> , anteriormente <i>Taxonomic Database Working Group</i>)
WAR	<i>Web Application Archive</i>

Resumen ejecutivo

El proyecto "Soporte Informático con el fin de desarrollar un prototipo de la **Plataforma para la Gestión de Conocimiento e Información Nacional sobre Biodiversidad (PGCIB)**", liderado por la Comisión Nacional para la Gestión de la Biodiversidad (Conagebio) y el Centro Nacional de Información Geoambiental (Ceniga), tuvo como objetivo *"facilitar el acceso a la información sobre la biodiversidad nacional, desarrollando un prototipo de plataforma web que permita sistematizar, documentar y publicar esta información, con base en la estructura de plataformas similares ya desarrolladas"*. Entre los principales productos del proyecto pueden mencionarse el análisis y la documentación de los requerimientos de información de los usuarios de la PGCIB, el fortalecimiento de las capacidades a nivel nacional en temas de informática de la biodiversidad y el desarrollo de un primer prototipo de la PGCIB. La duración del proyecto fue de un año y dos meses, entre el 1 de febrero de 2017 y el 31 de marzo de 2018.

El presente documento describe el primer prototipo de la PGCIB. La funcionalidad priorizada para la implementación del prototipo fue seleccionada por el equipo de análisis de requerimientos de los usuarios del sistema (Earus). El Earus fue conformado por 19 personas de instituciones como el Área de Conservación Guanacaste (ACG), la Conagebio, el Instituto Tecnológico de Costa Rica (ITCR), el Ministerio de Ciencia, Tecnología y Telecomunicaciones (Micitt), el Museo Nacional de Costa Rica (MNCR) y el Sistema Nacional de Áreas de Conservación (Sinac). Los requerimientos de información de los usuarios del sistema, definidos y priorizados por el Earus, están disponibles en el documento "Casos de uso de la PGCIB". Entre los subsistemas documentados para formar parte de la PGCIB se encuentran: Integración de datos, Calidad de datos, Estadísticas, Consultas y visualización de datos, Gestión de modelos de distribución de especies, Modelización y análisis, Gestión de la investigación, Observaciones de especies, Herramientas para la identificación de especies, Comunicación y gestión institucional y Manejo de materiales de capacitación.

El primer prototipo está disponible en <http://biodiversidad.conagebio.go.cr/> e incluye la implementación de parte de los **subsistemas de "Consultas y visualización de datos (portal de datos)" y de "Comunicación y gestión institucional (portal de comunicación)"** de la PGCIB. El código fuente de estos desarrollos está publicado en <https://github.com/PGCIB>. El portal de datos integra registros de colecciones de biodiversidad realizadas en el ACG por el Dr. Daniel Janzen, la Dra. Winnie Hallwachs y su equipo de investigación, los cuales fueron seleccionados como una muestra apropiada para mostrar el potencial de la PGCIB. Este subsistema atiende las consultas realizadas por los usuarios con base en diversos criterios (ej. taxonomía, geografía, temporalidad) y retorna los datos correspondientes en los formatos requeridos (ej. tablas, mapas, gráficos estadísticos). El portal de datos de la PGCIB está basado en el software desarrollado por el Atlas de la Biodiversidad de Australia (ALA - *Atlas of Living*

Australia), un proyecto colaborativo que agrega datos de biodiversidad de ese país, provenientes de múltiples fuentes, y los presenta en Internet para su uso libre y abierto. Por su parte, el portal de comunicación contiene información general y noticias sobre la PGCIB. Utiliza múltiples medios (ej. textos, imágenes) para de una forma amigable informar e invitar a los visitantes a utilizar los servicios de la plataforma.

1. Introducción

La **Plataforma para la Gestión de Conocimiento e Información Nacional sobre Biodiversidad (PGCIB)**, ha sido conceptualizada por la Comisión Nacional para la Gestión de la Biodiversidad (Conagebio) y el Centro Nacional de Información Geoambiental (Ceniga) como un **sistema de información de acceso libre y gratuito, disponible en Internet y orientado a integrar datos e información sobre la biodiversidad de Costa Rica, promover su curación y brindar servicios de análisis y visualización a partir de los datos integrados**. En este contexto, el proyecto "Soporte Informático con el fin de desarrollar un prototipo de la PGCIB", que se desarrolló como una consultoría, estuvo orientado a apoyar a la Conagebio y al Ceniga en el diseño e implementación de este sistema. La consultoría fue parte del componente cuatro del proyecto "Promoviendo la aplicación del Protocolo de Nagoya a través del desarrollo de productos basados en la naturaleza, distribución de beneficios y conservación de la biodiversidad en Costa Rica", el cual es ejecutado a nivel nacional por Conagebio, financiado con recursos del Fondo Mundial para el Medio Ambiente (GEF - *Global Environmental Fund*) y administrado financieramente por el Programa de las Naciones Unidas para el Desarrollo (PNUD).

De acuerdo con sus términos de referencia, el objetivo de la consultoría fue *"facilitar el acceso a la información sobre la biodiversidad nacional, desarrollando un prototipo de plataforma web que permita sistematizar, documentar y publicar esta información, con base en la estructura de plataformas similares ya desarrolladas. Este prototipo servirá de modelo para el desarrollo final de la PGCIB. Este sistema, cuya coordinación general estará a cargo de Conagebio, operará en el contexto del Sistema Nacional de Información Ambiental (Sinia) del Ministerio de Ambiente y Energía (Minae) coordinado por el Ceniga"*. La consultoría tuvo una duración de un año y se ejecutó entre el 1 de febrero de 2017 y el 31 de marzo de 2018.

Con el fin de apoyar el proceso de desarrollo del prototipo, Conagebio conformó el **equipo de análisis de requerimientos de los usuarios del sistema (Earus)**. El Earus fue conformado por 19 personas de instituciones como el Área de Conservación Guanacaste (ACG), la Conagebio, el Instituto Tecnológico de Costa Rica (ITCR), el Ministerio de Ciencia, Tecnología y Telecomunicaciones (Micitt), el Museo Nacional de Costa Rica (MNCR) y el Sistema Nacional de Áreas de Conservación (Sinac). Los miembros del Earus participaron en las sesiones de análisis de requerimientos con el

fin de definirlos, evaluarlos y priorizarlos. El apéndice I contiene la lista completa de miembros del Earus.

Los requerimientos de información del sistema, definidos y priorizados por el Earus, están disponibles en el documento "Casos de uso de la PGCIB"¹. Para documentar los requerimientos de información de los usuarios del sistema, se seleccionó la metodología de desarrollo de software conocida como **Proceso Unificado de Desarrollo de Software (Proceso Unificado)**. En esta, los requerimientos de información de los usuarios se especifican en forma de **casos de uso**, los cuales son listas de acciones o eventos que definen las interacciones de un sistema con sus usuarios o con otros sistemas para lograr un propósito determinado (ej. importar un conjunto de datos, desplegar un gráfico, mostrar un mapa) (Jacobson et al., 1999). Los casos de uso incluidos fueron organizados en **paquetes** y estos a su vez son parte de **subsistemas**. Los subsistemas que componen la PGCIB se listan a continuación:

1. Integración de datos
2. Calidad de datos
3. Estadísticas
4. Consultas y visualización de datos
5. Gestión de modelos de distribución de especies
6. Modelización y análisis
7. Gestión de la investigación
8. Observaciones de especies
9. Herramientas para la identificación de especies
10. Comunicación y gestión institucional
11. Manejo de materiales de capacitación

El presente documento describe el primer prototipo de la PGCIB. Los casos de uso tomados en cuenta para ser parte del prototipo fueron seleccionados por el Earus e incluyen la implementación de parte de los **subsistemas de consultas y visualización de datos (portal de datos) y de comunicación y gestión institucional (portal de comunicación)** de la PGCIB. El portal de datos integra registros de colecciones de biodiversidad realizadas en el ACG por el Dr. Daniel Janzen, la Dra. Winnie Hallwachs y su equipo de investigación, los cuales fueron seleccionados como una muestra apropiada para mostrar el potencial de la PGCIB. Este subsistema atiende las consultas realizadas por los usuarios con base en diversos criterios (ej. taxonomía, geografía, temporalidad) y retorna los datos correspondientes en los formatos requeridos (ej.

¹ Documentación de los casos de uso de la PGCIB:
<https://docs.google.com/document/d/1xC70Yao021-1-8P-niHCAIr4XGEqv7PEJO-5GGdjuwE/edit#>

tablas, mapas, gráficos estadísticos). El portal de datos de la PGCIB está basado en el software desarrollado por el Atlas de la Biodiversidad de Australia (ALA - Atlas of Living Australia), un proyecto colaborativo que agrega datos de biodiversidad de ese país, provenientes de múltiples fuentes, y los presenta en Internet para su uso libre y abierto. Por su parte, el portal de comunicación contiene información general y noticias sobre la PGCIB. Utiliza múltiples medios (ej. textos, imágenes) para de una forma amigable informar e invitar a los visitantes a utilizar los servicios de la plataforma.

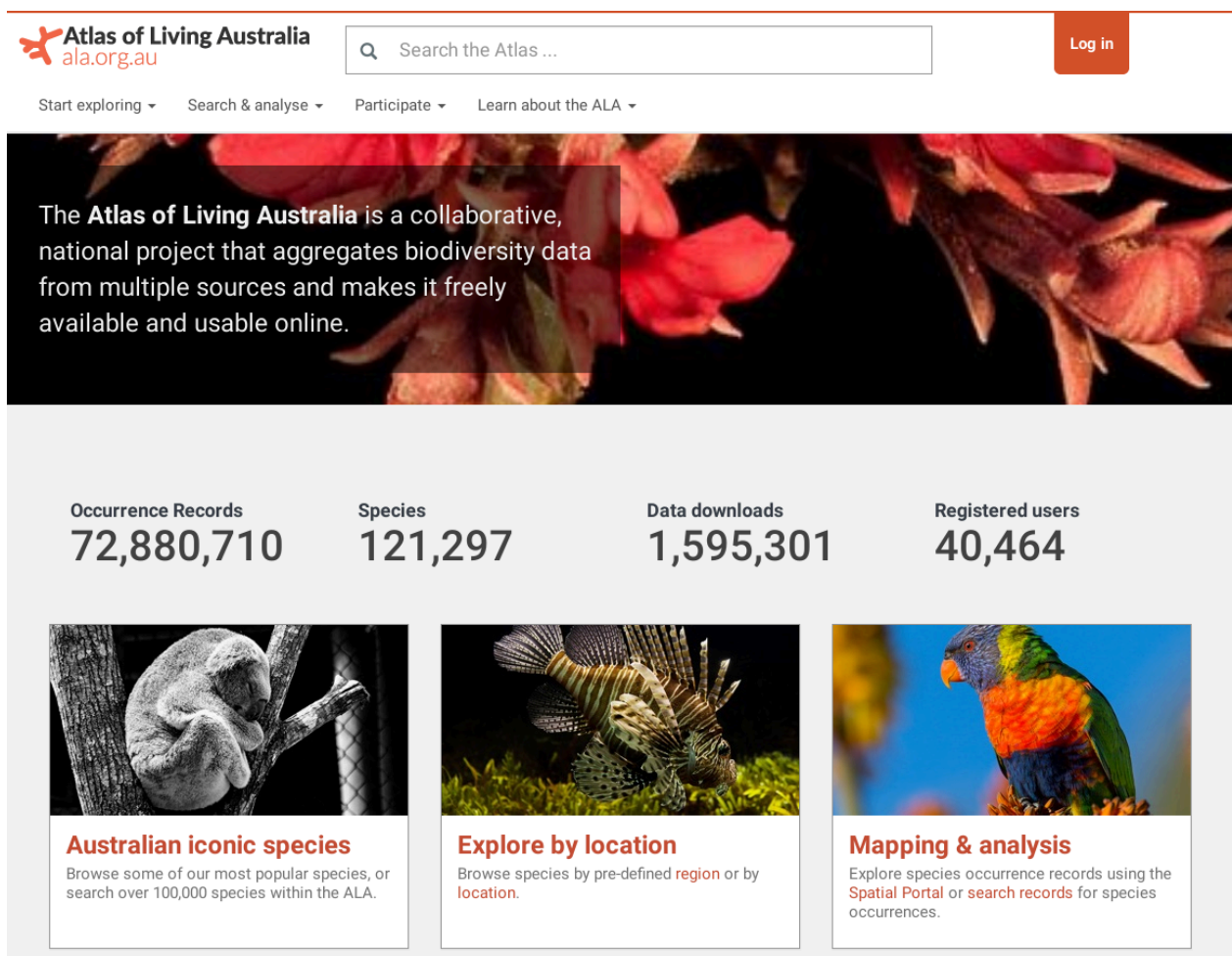
2. Descripción del prototipo

El prototipo comprende parte de la funcionalidad de dos de los subsistemas identificados y documentados durante el proceso de análisis de la PGCIB. Estos subsistemas son:

- **Subsistema 4: Consultas y visualización de datos (portal de datos).** Atiende las consultas realizadas por los usuarios con base en diversos criterios (ej. taxonomía, geografía, temporalidad) y retorna los datos correspondientes en los formatos requeridos (ej. tablas, mapas, gráficos estadísticos). Esta funcionalidad permite dar acceso a datos de presencia de especies y, en el caso del prototipo, a datos de colecciones de biodiversidad del ACG.
- **Subsistema 10: Comunicación y gestión institucional (portal de comunicación).** Proporciona información general de la PGCIB, acceso a noticias, documentos y otros temas de interés. De este subsistema, el prototipo incluye la funcionalidad de brindar al usuario información general y noticias relacionadas con la plataforma.

3. Portal de datos

El portal de datos de la PGCIB está basado en el software desarrollado por el Atlas de la Biodiversidad de Australia (ALA - *Atlas of Living Australia*), un proyecto colaborativo que agrega datos de biodiversidad de ese país, provenientes de múltiples fuentes, y los presenta en Internet para su uso libre y abierto. La página principal del sitio web del ALA se muestra en la figura 1.



Atlas of Living Australia
ala.org.au

Search the Atlas ...

Log in

Start exploring ▾ Search & analyse ▾ Participate ▾ Learn about the ALA ▾

The **Atlas of Living Australia** is a collaborative, national project that aggregates biodiversity data from multiple sources and makes it freely available and usable online.

Occurrence Records	Species	Data downloads	Registered users
72,880,710	121,297	1,595,301	40,464

Australian iconic species

Browse some of our most popular species, or search over 100,000 species within the ALA.

Explore by location

Browse species by pre-defined **region** or by **location**.

Mapping & analysis

Explore species occurrence records using the **Spatial Portal** or **search records** for species occurrences.

Figura 1. Página principal del sitio web del ALA².

Desde que fue inaugurado en 2010, producto de una inversión de €39 millones realizada entre 2007 y 2010 por el gobierno australiano, el portal de ALA se ha constituido en un ejemplo emblemático de aplicación de tecnología informática a las necesidades de un país de mejorar sus servicios de datos de biodiversidad y el acceso a estos³. En la actualidad, el portal de ALA es considerado “uno de los portales bioambientales más avanzados del mundo ... enfocado a audiencias que van desde estudiantes de primaria, pasando por ciudadanos naturalistas y llegando hasta investigadores de posgrado” (Belbin & Williams, 2015).

² Sitio web del *Atlas of Living Australia*: <https://www.ala.org.au/>

³ Artículo en el sitio de GBIF sobre la comunidad internacional de portales basados en el software de ALA: <https://www.gbif.org/programme/82953/living-atlases>

Además de Australia, las herramientas de ALA ya han sido instaladas y adaptadas en otros países (ej. España, Francia, Argentina, Brasil, Portugal, Reino Unido, Costa Rica, Suecia), los cuales han conformado una comunidad de programadores y usuarios con el apoyo de la Infraestructura Mundial de Información en Biodiversidad (GBIF - *Global Biodiversity Information Facility*). Esta colaboración es ahora un componente esencial de la red de GBIF, y de la comunidad global de informática de la biodiversidad, que facilita servicios y administración de datos de alta capacidad, tanto a nivel nacional como regional y también temático. La comunidad comenzó a formarse precisamente en Costa Rica en el año 2013, a raíz de un proyecto de apoyo al desarrollo de capacidades, financiado por GBIF, entre los nodos de Australia y Costa Rica. Desde entonces, al menos un taller de programación ha sido organizado anualmente en el contexto de iniciativas como GBIF⁴ y Estándares de Información sobre Biodiversidad (TDWG - *Biodiversity Information Standards*)⁵, a la vez que se han generado varios proyectos y acuerdos de colaboración enfocados en el mantenimiento, documentación e internacionalización del sistema⁶⁷.

3.1. Arquitectura y funcionalidad

En esta sección, se explican las características generales de la arquitectura del portal de ALA y se detallan las de los módulos que se implementarán para el prototipo de la PGCIB.

3.1.1. Características generales

El portal de ALA ha sido elaborado bajo los siguientes principios de arquitectura (Cavière et al., 2016):

1. Orientado a servicios: todos los datos presentados en el portal de ALA son accesibles a través de servicios web, los cuales han sido desarrollados mediante interfaces de programación (API - *Application Programming Interface*) públicas. Esta característica promueve la compartición de datos y permite que elementos del portal sean embebidos en otras herramientas y sitios web.

⁴ Taller internacional sobre ALA en Madrid, 2016:

<http://community.canadensys.net/2016/international-atlas-of-living-australia-workshop>

⁵ Presentación en la conferencia de TDWG 2017 sobre la comunidad internacional de portales basados en el software de ALA: <https://biss.pensoft.net/articles.php?id=20290>

⁶ Proyecto *Encounter Bay* de apoyo a la internacionalización y documentación del software de ALA: <https://www.gbif.org/project/82202/internationalization-of-the-ala-node-portal>

⁷ Proyecto *CoopBioPlat* de apoyo a la internacionalización y documentación del software de ALA: http://www.gbif.es/CoopBioPlat_in.php

2. Componentes modulares: el portal de ALA está formado por un conjunto de micro componentes de software, cada uno de los cuales cumple un rol específico, lo que ayuda a promover la reutilización y permite flexibilidad en los desarrollos.

3. Módulos reutilizables en la interfaz de usuario (IU): los componentes de la IU están basados en una arquitectura de complementos, lo que permite su reutilización y personalización. Los creadores del portal de ALA reconocen que brindar una marca personalizada es una parte importante de la implementación de un portal para una comunidad.

4. Portabilidad: el portal está construido exclusivamente con software de código abierto y multiplataforma, el cual facilita su instalación y migración entre las principales plataformas (ej. Linux, Microsoft Windows, Mac OS). Todas sus aplicaciones pueden ser instaladas mediante un *script* que ayuda también al mantenimiento de sus componentes.

El portal está compuesto por varios módulos que trabajan conjuntamente, pero que también pueden hacerlo individualmente o en grupos de módulos específicos. La figura 2 muestra la arquitectura completa del sistema y resalta en rojo los módulos implementados como parte del prototipo de la PGCIB.

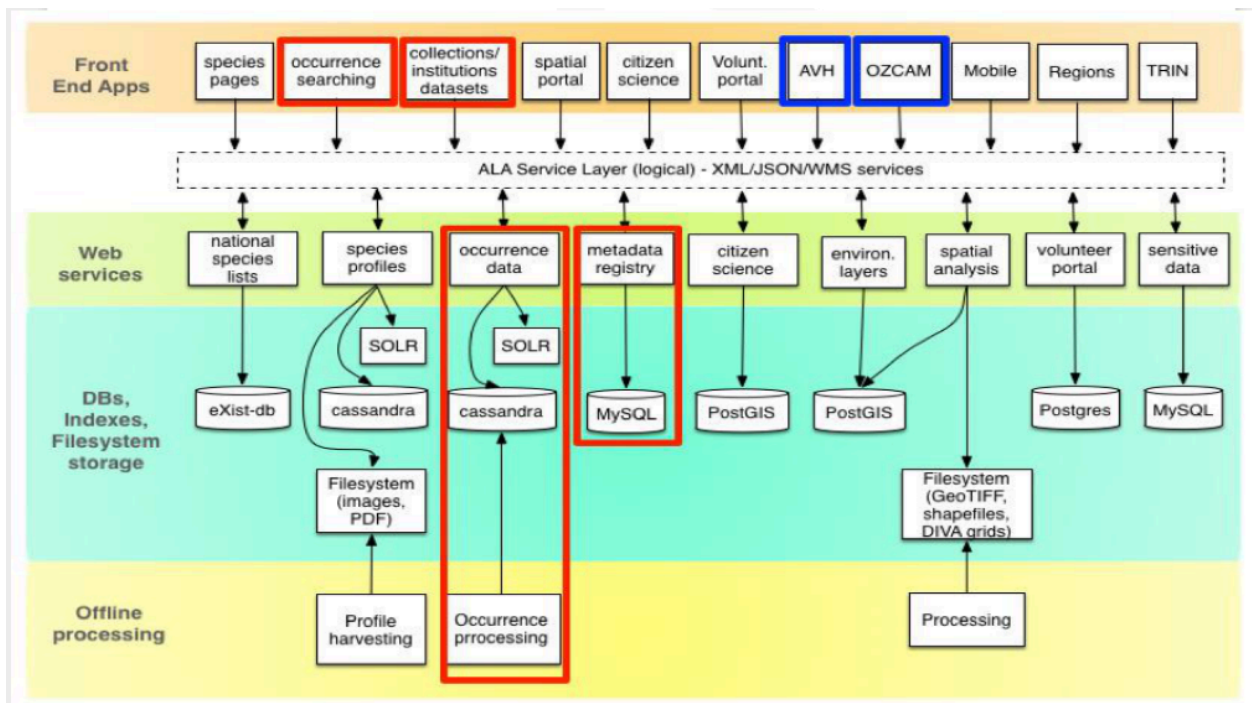
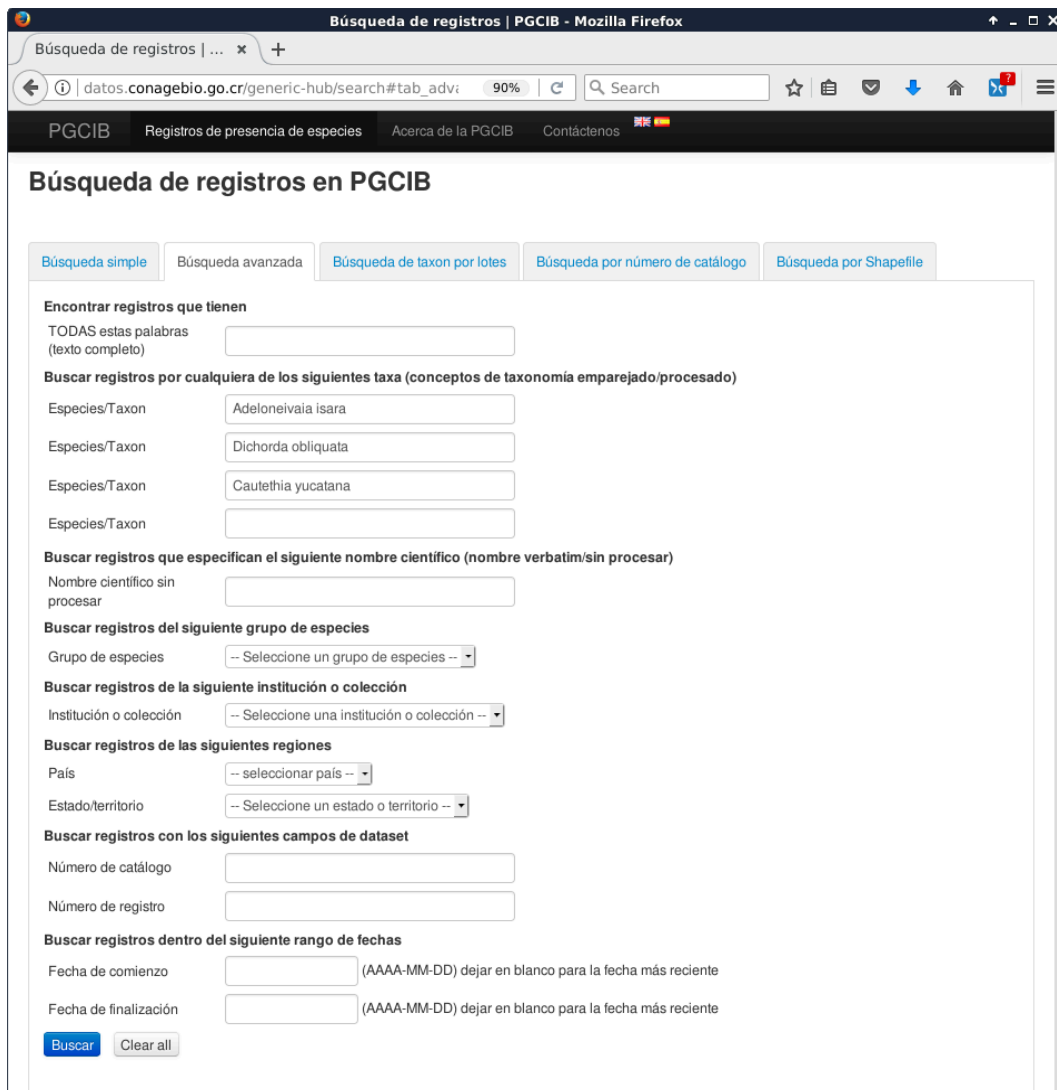


Figura 2. Arquitectura del portal de ALA (Cavière et al., 2016).

Seguidamente, se detallan las principales características de los módulos relacionados con la búsqueda de registros de presencia de especies y también de colecciones, instituciones y conjuntos de datos, que serán los que se implementarán como parte del prototipo de la PGCIB.

3.1.2. Búsqueda de registros de presencia (*Biocache*)



The screenshot displays the 'Búsqueda de registros en PGCIB' (Search for records in PGCIB) interface. The browser address bar shows the URL 'datos.conagebio.go.cr/generic-hub/search#tab_adv:'. The page features a navigation bar with 'PGCIB', 'Registros de presencia de especies', 'Acerca de la PGCIB', and 'Contáctenos'. Below the navigation bar, the main heading is 'Búsqueda de registros en PGCIB'. There are five tabs for search methods: 'Búsqueda simple', 'Búsqueda avanzada', 'Búsqueda de taxon por lotes', 'Búsqueda por número de catálogo', and 'Búsqueda por Shapetfile'. The 'Búsqueda avanzada' tab is selected. The search criteria are organized into several sections:

- Encontrar registros que tienen:** A text input field for 'TODAS estas palabras (texto completo)'. Below it, a section titled 'Buscar registros por cualquiera de los siguientes taxa (conceptos de taxonomía emparejado/procesado)' contains four 'Especies/Taxon' input fields, with the first three containing 'Adeloneivaia isara', 'Dichorda obliquata', and 'Cautethia yucatanana'.
- Buscar registros que especifican el siguiente nombre científico (nombre verbatim/sin procesar):** A text input field for 'Nombre científico sin procesar'.
- Buscar registros del siguiente grupo de especies:** A dropdown menu for 'Grupo de especies' with the option '-- Seleccione un grupo de especies --'.
- Buscar registros de la siguiente institución o colección:** A dropdown menu for 'Institución o colección' with the option '-- Seleccione una institución o colección --'.
- Buscar registros de las siguientes regiones:** Two dropdown menus for 'País' (option: '-- seleccionar país --') and 'Estado/territorio' (option: '-- Seleccione un estado o territorio --').
- Buscar registros con los siguientes campos de dataset:** Two text input fields for 'Número de catálogo' and 'Número de registro'.
- Buscar registros dentro del siguiente rango de fechas:** Two text input fields for 'Fecha de comienzo' and 'Fecha de finalización', both with the placeholder '(AAAA-MM-DD) dejar en blanco para la fecha más reciente'.

At the bottom left, there are two buttons: 'Buscar' (blue) and 'Clear all' (grey).

Figura 3. Pantalla de búsqueda avanzada.

La funcionalidad de búsqueda de registros de presencia es una de las principales del portal de datos. Hay diferentes formas de realizar estas búsquedas. Por ejemplo, en la

figura 3 se muestra la pantalla de búsqueda avanzada, en la que se realiza una búsqueda por tres diferentes especies. Adicionalmente, pueden utilizarse otros criterios de búsqueda como grupos de especies, instituciones que publican los registros, países/estados, números de catálogo y fechas de recolección, por mencionar algunos.

En la figura 4 se muestra la pantalla de búsqueda por área, que permite realizar búsquedas mediante la ubicación de un círculo en un mapa, para así visualizar los registros de presencia ubicados dentro del círculo y un resumen en formato tabular de estos. Otras interfaces permiten realizar búsquedas geoespaciales con base en otros tipos de polígonos, como por ejemplo rectángulos e incluso polígonos irregulares, ya sea dibujados manualmente sobre el mapa o provenientes de un archivo (ej. *shapefile*).

Explora Tu Área | Atlas of Living Australia - Mozilla Firefox

datos.conagebio.go.cr/generic-hub/explore/your-area?default=true#10

PGCIB Registros de presencia de especies Acerca de la PGCIB Contáctenos

Explora tu área

Ingrese su ubicación o dirección:

Por ejemplo, una dirección, nombre, código postal o coordenadas de GPS (como lat, largo)

Mostrando los registros de: **Unnamed Road, Provincia de Guanacaste, Costa Rica**

Mostrar los registros en un km de radio

Grupo	Especies	Especies : Nombre Común	Registros
All Species	1145	1. <i>Abablemma ulopus</i> Dyar, 1914	1
Animals	1145	2. <i>Accinctapubes albifasciata</i> Druce, 1902	6
Mammals	0	3. <i>Achalarus albociliatus</i> Mabilie, 1877	1
Birds	0	4. <i>Achlyodes busirus</i> Stoll, 1782	1
Reptiles	0	5. <i>Aclytia albistriga</i> Schaus, 1910	1
Amphibians	0	6. <i>Aclytia heber</i> Cramer, 1782	31
Fishes	0	7. <i>Aclytia punctata</i> Butler, 1876	1
Molluscs	0	8. <i>Aclytia ventralis</i> Guerin-Meneville, 1843	4
Arthropods	1145	9. <i>Acraga hamata</i> Schaus, 1910	3
Crustaceans	0	10. <i>Acronyctodes cautama</i> Schaus, 1901	2
Insects	1145	11. <i>Acroria chloographa</i> Hampson, 1908	4
Plants	0	12. <i>Acrosemia vulpecularia</i> Herrich-Schäffer, 1855	13
Bryophytes	0	13. <i>Acrotomia mucia</i> Druce, 1892	4
Gymnosperms	0	14. <i>Acrotomia viminaria</i> Herrich-Schäffer, 1856	3
FernsAndAllies	0	15. <i>Acrotomodes bola</i> Druce, 1892	2
Angiosperms	0	16. <i>Adeloneivaia boisduvalii</i> Doumet, 1859	1
Monocots	0	17. <i>Adeloneivaia jason</i> Boisduval, 1871/72	30
Dicots	0	18. <i>Adhemarius gannascus</i> Stoll, 1790	30
Fungi	0	19. <i>Adhemarius ypsilon</i> Rothschild & Jordan, 1903	8
Chromista	0	20. <i>Aellopos titan</i> Cramer, 1777	2
Protozoa	0	21. <i>Agathodes designalis</i> Guenée, 1854	5
Bacteria	0	22. <i>Aglaonice hirtipalpis</i> Walker, 1858	1
Algae	0		1

Consejos: puede ajustar la ubicación de la zona arrastrando el icono rojo

CONAGEBio COSTA RICA minae

This site is licensed under a [Creative Commons Attribution 3.0 Australia License](#). Provider content may be covered by other [Terms of Use](#).

Figura 4. Búsqueda por área.

La página de resultados incluye tres pestañas (*tabs*): una lista tabulada de todos los registros que coinciden con la búsqueda, un mapa con su ubicación y un conjunto de gráficos estadísticos. Desde estos resultados, se puede navegar al detalle de un registro para ver la totalidad de sus elementos de información, así como los datos de la institución y de la colección a la que pertenecen. Es importante resaltar que todos los componentes del portal están enlazados entre sí. En la figura 5 se presenta un ejemplo de resultados en formato tabular.

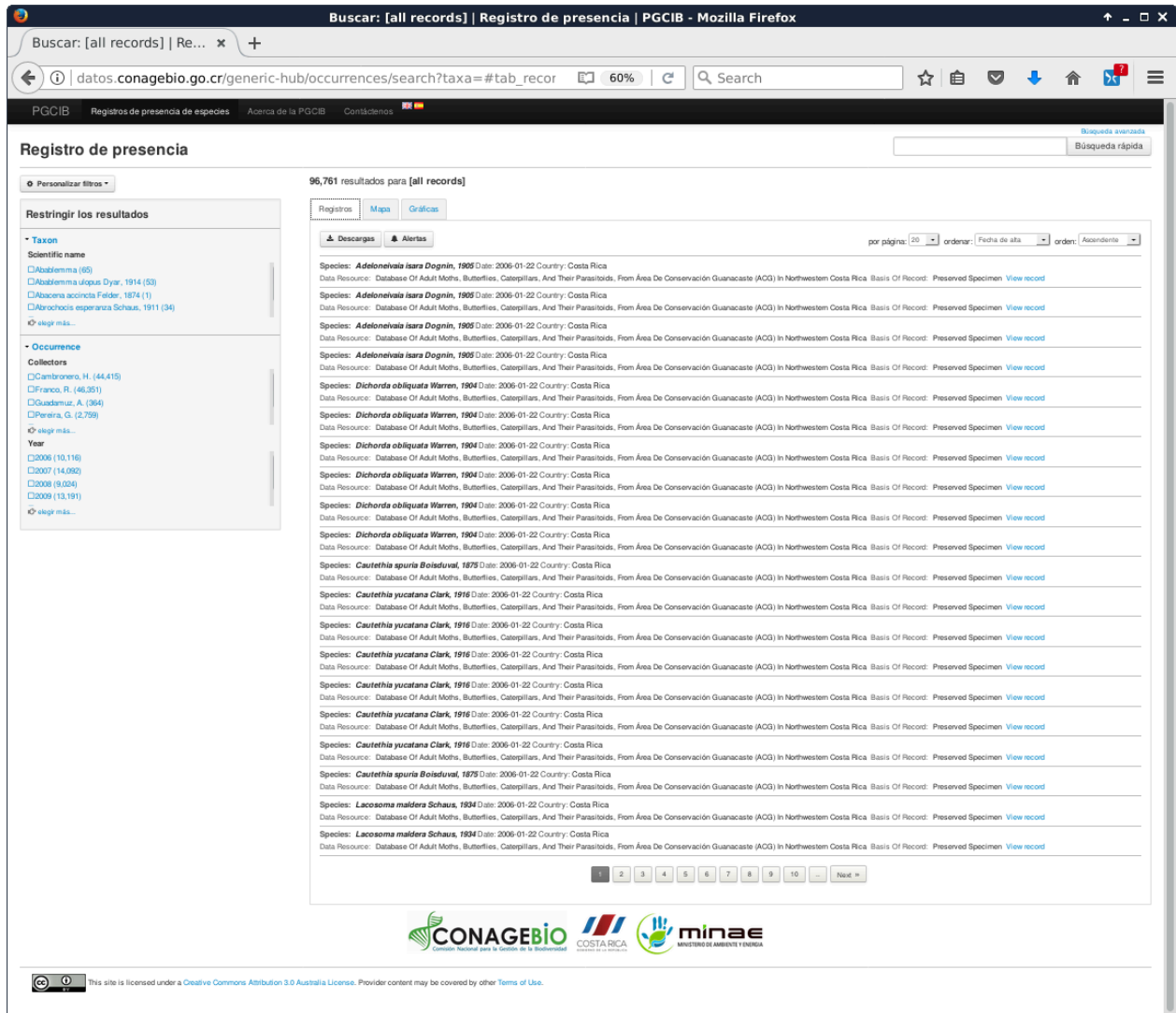


Figura 5. Resultados de una búsqueda en formato tabular.

En la figura 6 se muestran los gráficos estadísticos correspondientes a los resultados obtenidos mediante una consulta. Estos incluyen la distribución de los datos por tiempo (década y mes del año) y por taxonomía, así como una sumarización de los principales problemas o alertas sobre calidad que fueron encontrados (ej. fechas inválidas, georreferenciación inconsistente).

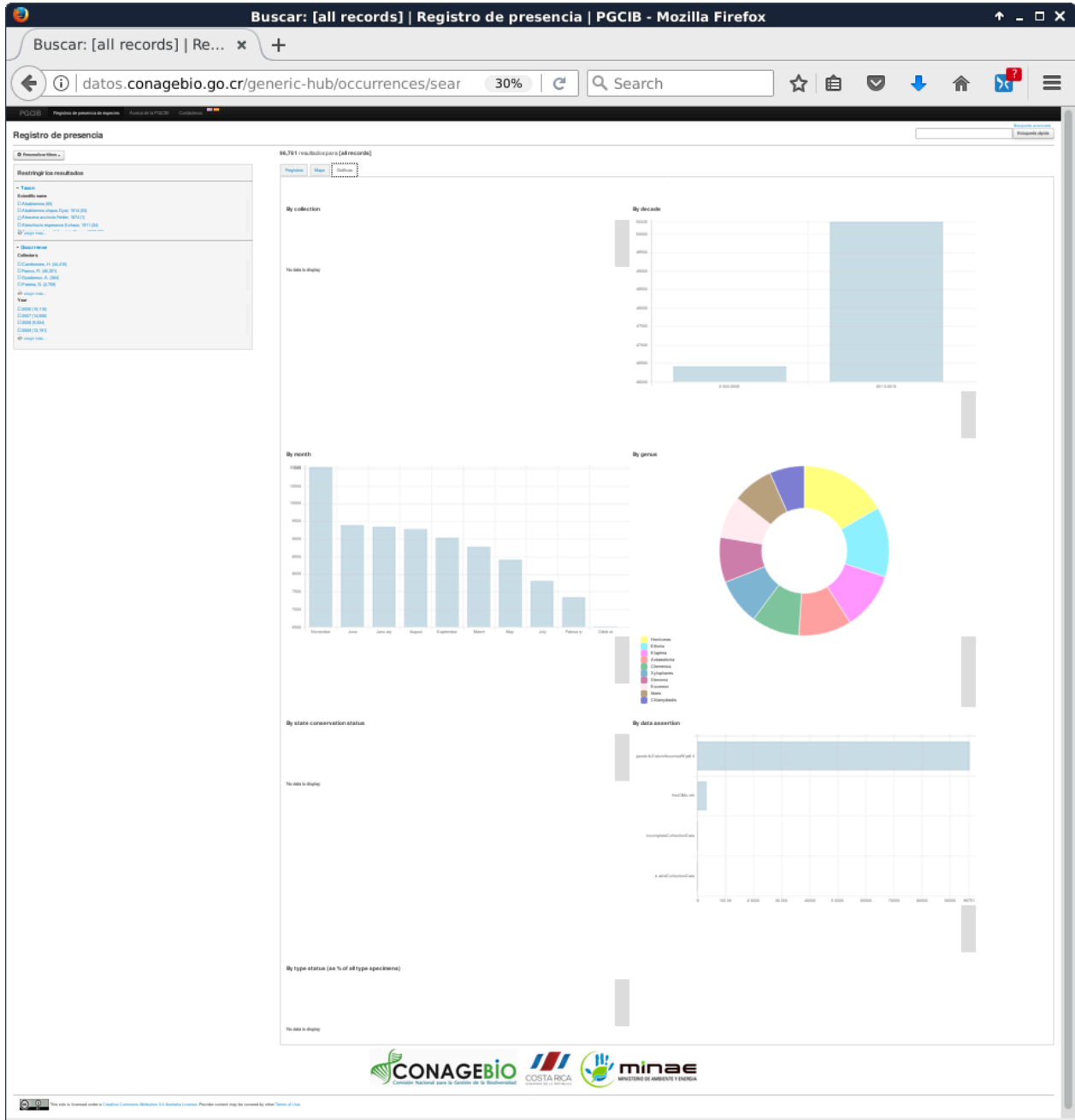


Figura 6. Resultados de una búsqueda en formato de gráficos estadísticos.

Desde el punto de vista técnico, la búsqueda de registros de presencia corresponde a los módulos que se resaltan en la figura 7 y que son parte de un conjunto llamado *Biocache*, el cual se encarga del manejo general de estos registros.

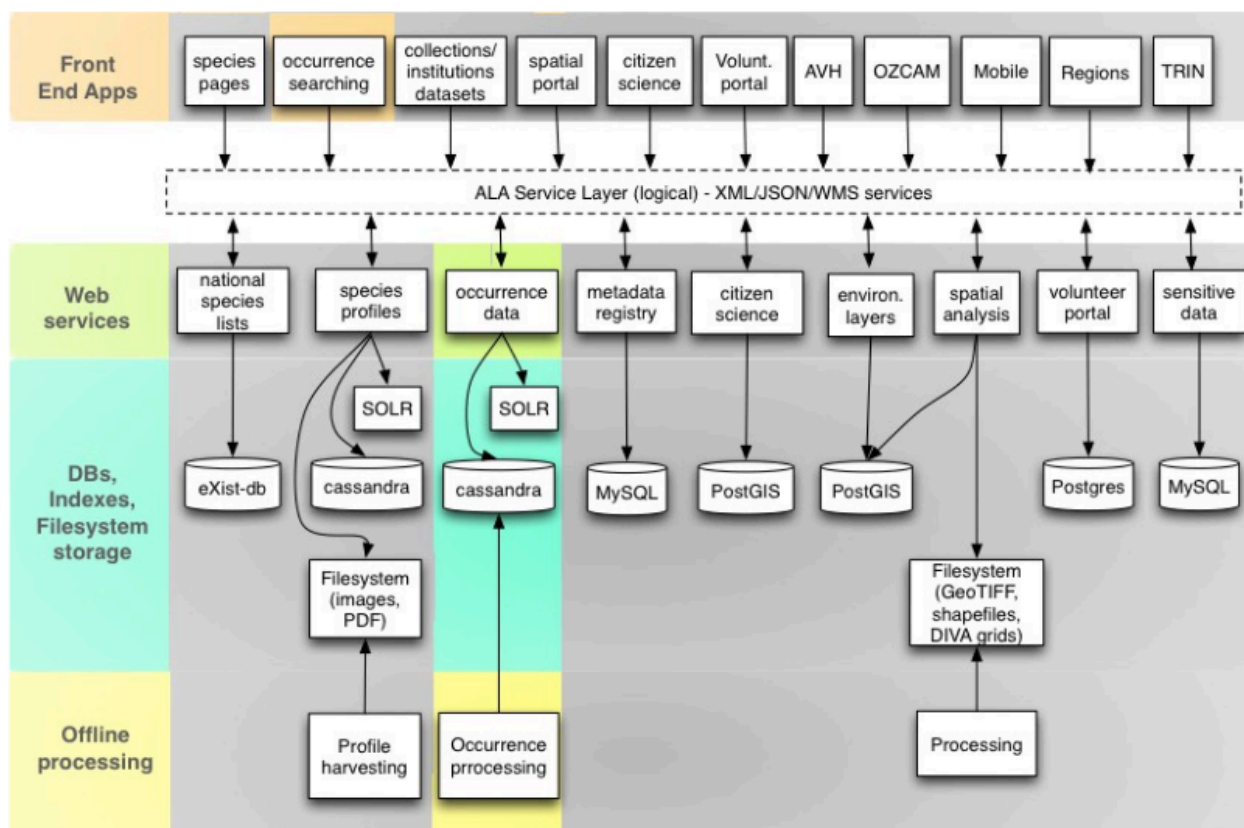


Figura 7. Módulos relacionados con la búsqueda de registros de presencia de especies (Cavière et al., 2016).

Los módulos principales que componen el *Biocache* son:

- El *biocache-store* (correspondiente a la caja *occurrence processing* en la figura 5).
- El *biocache-service* (correspondiente a la caja *occurrence data*).
- El *biocache-hubs* y el *generic-hub* (estos dos últimos corresponden a la caja *occurrence searching*).

El código fuente y la documentación de todos estos módulos está disponible en el sitio de ALA en la plataforma de desarrollo colaborativo GitHub⁸.

⁸ Sitio de ALA en GitHub: <https://github.com/AtlasOfLivingAustralia/>

El *biocache-store*⁹ es la implementación, en el lenguaje de programación *Scala*¹⁰, de la capa de datos del *Biocache*. Su código se encarga de la carga, muestreo, procesamiento e indexación de los registros de presencia de especies. Estos registros se almacenan en una base de datos *Apache Cassandra*¹¹ indexada mediante el motor de búsqueda *Apache Solr*¹². Esta indexación permite reducir considerablemente los tiempos de respuesta de las búsquedas. Además, el *biocache-store* cuenta con herramientas para apoyar la detección de casos atípicos (*outlier detection*) y de duplicados. El *biocache-service*¹³ es una capa de servicios web que acceden a los datos almacenados en el *biocache-store*.

Por su parte, el *biocache-hubs*¹⁴ es un complemento (*plugin*) de *Grails*¹⁵ que provee una interfaz gráfica web para el *biocache-service*, mientras que *generic-hub*¹⁶ es una implementación de ejemplo para el *biocache-hubs*. La idea es que *generic-hub* sirva como una plantilla a partir de la cual se creen implementaciones personalizadas para cada portal (ej, *crbio-hub*, *pgcib-hub*). Hay muy poco código contenido en el *generic-hub*, ya que la mayoría se "hereda" del *biocache-hubs*.

3.1.3. Búsqueda de registros de colecciones, instituciones y conjuntos de datos (*Collectory*)

El *Collectory*¹⁷ es un complemento de *Grails* que administra el registro de metadatos de instituciones, colecciones, conjuntos de datos, contactos y otras entidades relacionadas. Toda esta información se almacena en una base de datos *MySQL*¹⁸, como se muestra en la figura 8.

⁹ Repositorio de *biocache-store* en GitHub: <https://github.com/AtlasOfLivingAustralia/biocache-store>

¹⁰ Sitio web del lenguaje de programación *Scala*: <http://www.scala-lang.org/>

¹¹ Sitio web del sistema administrador de bases de datos *Apache Cassandra*: <https://cassandra.apache.org/>

¹² Sitio web del motor de búsqueda *Apache Solr*: <http://lucene.apache.org/solr/>

¹³ Repositorio de *biocache-service* en GitHub: <https://github.com/AtlasOfLivingAustralia/biocache-service>

¹⁴ Repositorio de *biocache-hubs* en GitHub: <https://github.com/AtlasOfLivingAustralia/biocache-hubs>

¹⁵ Sitio web del marco de trabajo para desarrollo de aplicaciones web *Grails*: <https://grails.org/>

¹⁶ Repositorio de *generic-hub* en GitHub: <https://github.com/AtlasOfLivingAustralia/generic-hub>

¹⁷ Repositorio de *collectory-plugin* en GitHub: <https://github.com/AtlasOfLivingAustralia/collectory-plugin>

¹⁸ Sitio web del sistema administrador de bases de datos *MySQL*: <https://www.mysql.com/>

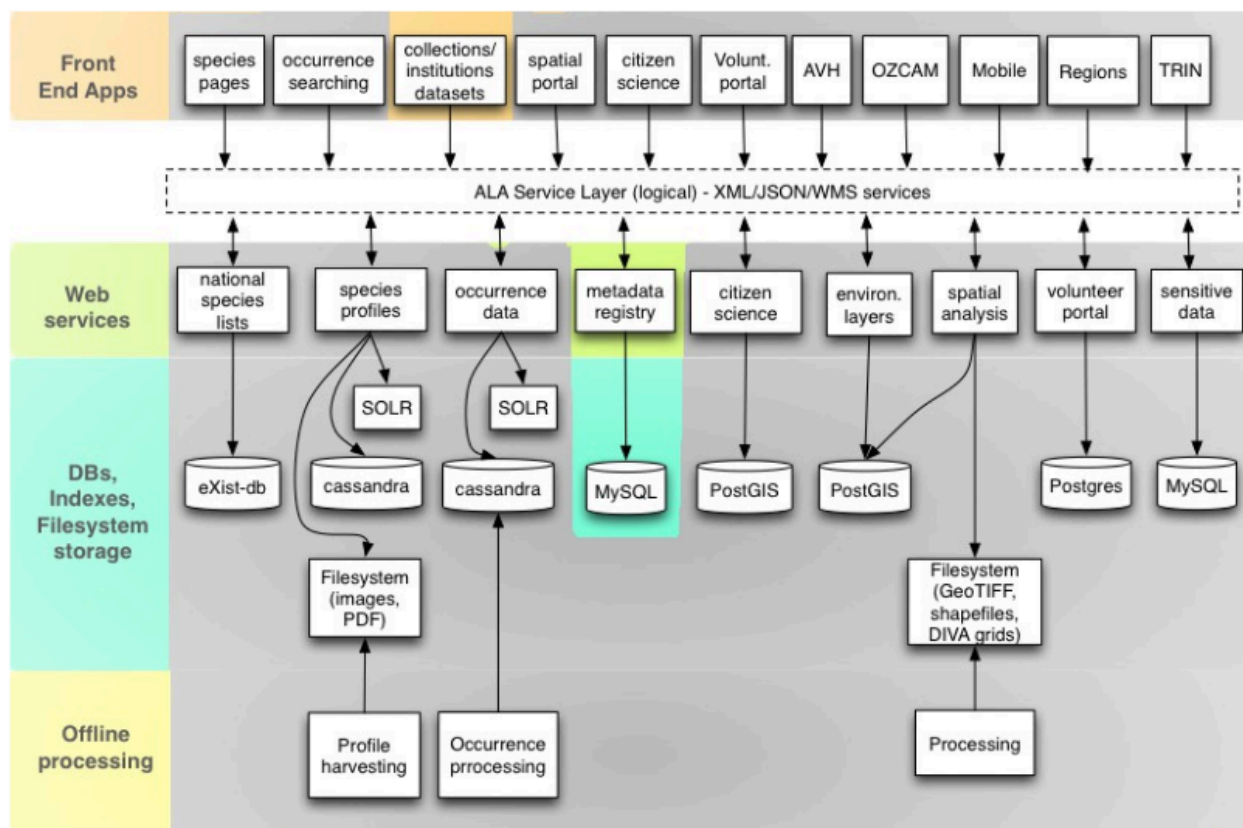


Figura 8. Módulos relacionados con la búsqueda de datos de colecciones, instituciones y conjuntos de datos (Cavière et al., 2016).

La interfaz de este módulo es relativamente simple. Por una parte, incluye funcionalidad para la búsqueda de colecciones, instituciones y conjuntos de datos mediante diferentes criterios. Además, provee herramientas para añadir, modificar y borrar estos tipos de entidades. Permite también importarlas desde el portal de GBIF.org. La pantalla principal es la que se presenta en la figura 9, mientras que en la figura 10 se muestra la interfaz para la búsqueda de conjuntos de datos.

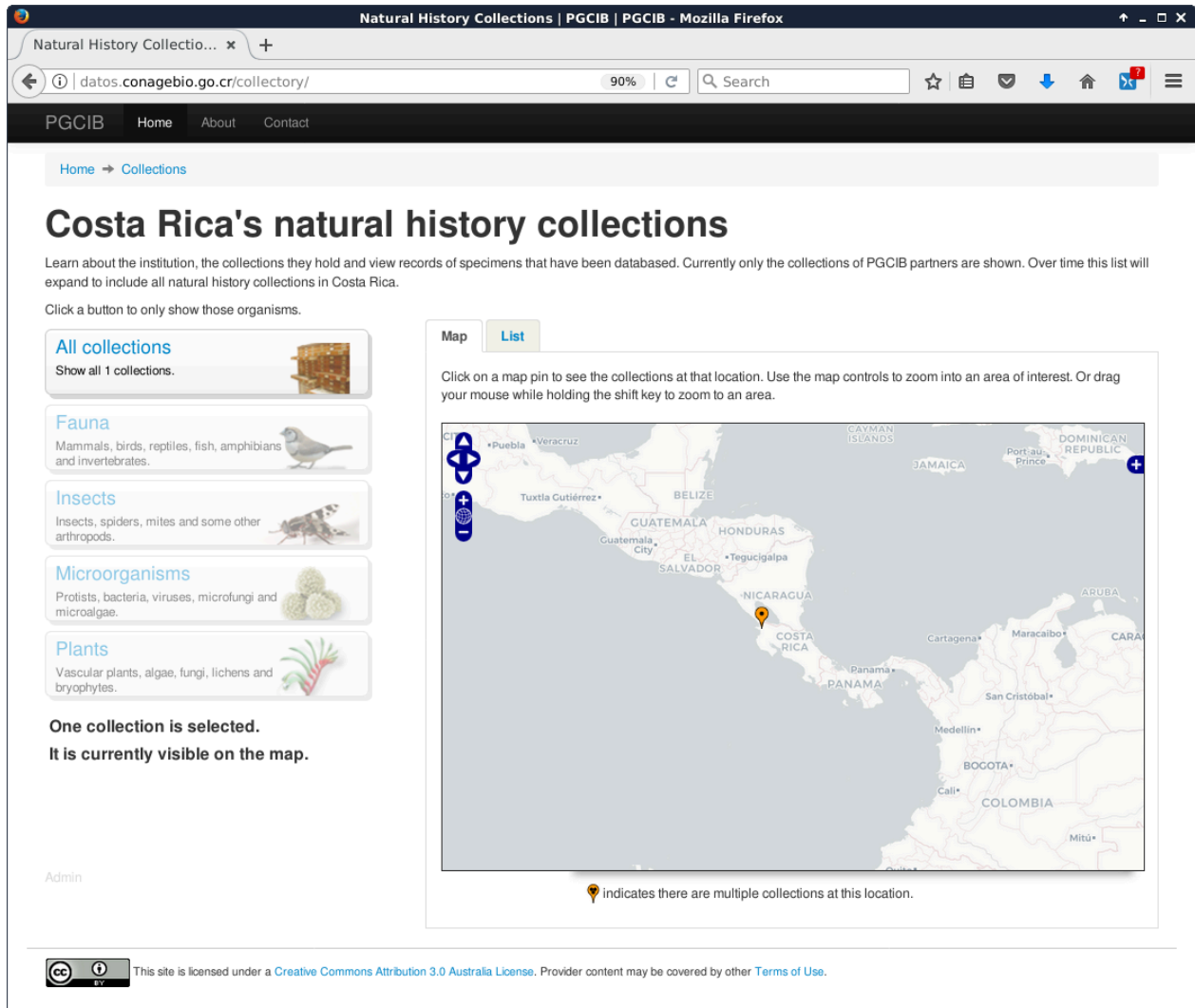


Figura 9. Pantalla principal del *Collectory*.

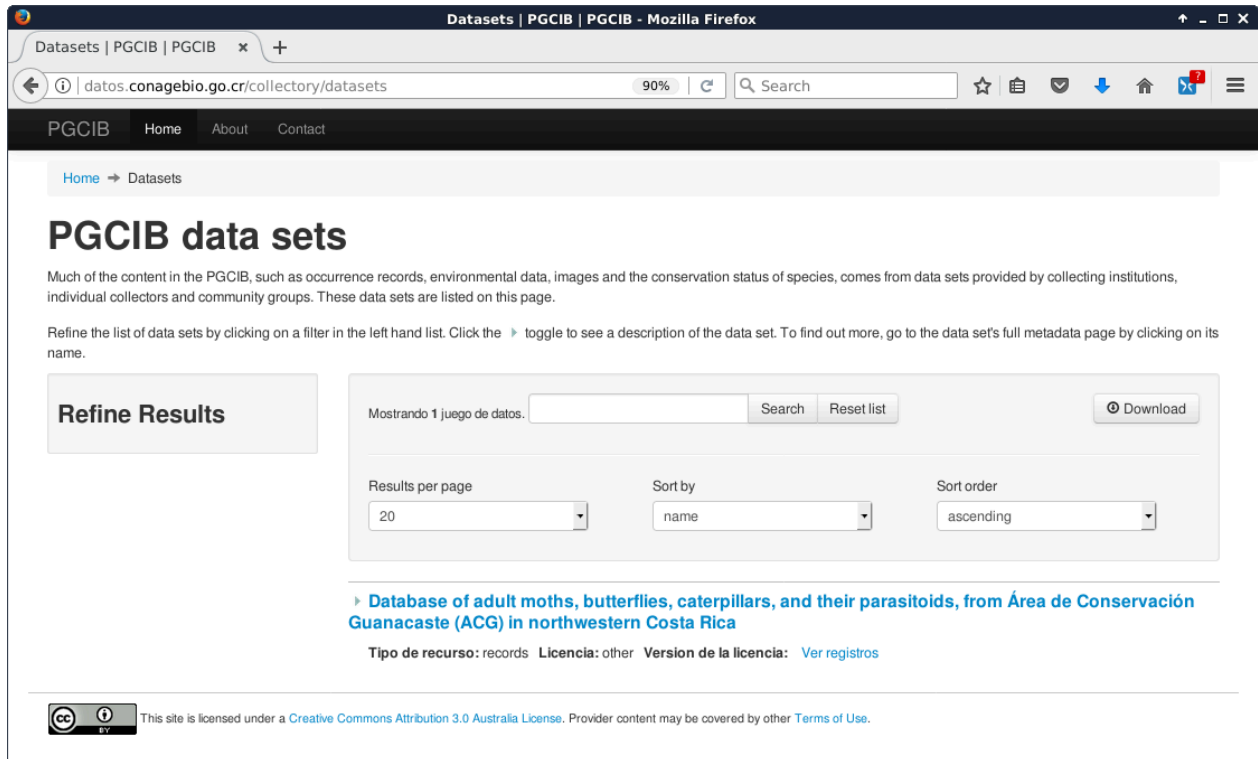


Figura 10. Interfaz de búsqueda de conjuntos de datos.

3.2. Herramientas informáticas utilizadas

Tanto en el *Biocache* como en el *Collectory*, se utiliza el marco de trabajo (*framework*) para desarrollo de aplicaciones web llamado *Grails*, el cual está basado en el lenguaje de programación *Groovy*¹⁹. *Groovy* se ejecuta en la máquina virtual de *Java*²⁰ y utiliza *Hibernate*²¹ para realizar el mapeo entre el modelo de objetos y la base de datos. Los módulos se distribuyen como archivos *WAR* (*Web Application Archive*) y se ejecutan en el contenedor de *servlets* de *Java* denominado *Apache Tomcat*²² (versión 7).

Como se ha mencionado, el motor de bases de datos que usa el *Biocache* es *Apache Cassandra* y el que usa el *Collectory* es *MySQL*.

Si se desea configurar o estilizar las aplicaciones web del portal, además de las herramientas ya mencionadas, es conveniente tener conocimientos del lenguaje de marcas de hipertexto (*HTML - HyperText Markup Language*) y de la sintaxis de hojas de

¹⁹ Sitio web del lenguaje de programación *Groovy*: <http://groovy-lang.org/>

²⁰ Sitio web del lenguaje de programación *Java* y su máquina virtual: <https://www.oracle.com/java/>

²¹ Sitio web de la herramienta de mapeo objeto-relacional *Hibernate*: <http://hibernate.org/>

²² Sitio web del contenedor de *servlets* de *Java* *Apache Tomcat*: <http://tomcat.apache.org/>

estilo en cascada (CSS - *Cascading Style Sheets*). Se recomienda también conocer el marco de trabajo *Bootstrap*²³ para desarrollo de aplicaciones web, el cual brinda, entre otras ventajas, la capacidad de que las aplicaciones del portal se adapten a diferentes tipos y tamaños de pantallas (ej. computadoras, tabletas, teléfonos).

Todo el código del portal de datos de la PGCIB está disponible en la plataforma de desarrollo colaborativo *GitHub*²⁴, en la dirección <https://github.com/PGCIB>. Ahí pueden encontrarse tres repositorios:

1. **pgcib_install** (<https://github.com/PGCIB/pgcib-install>): *scripts* de instalación de las herramientas.
2. **pgcib_hub** (<https://github.com/PGCIB/pgcib-hub>): código fuente del *Biocache*.
3. **pgcib_collectory** (<https://github.com/PGCIB/pgcib-collectory>): código fuente del *Collectory*.

3.3. Requerimientos para la instalación

El portal de ALA requiere de una gran cantidad de componentes de software, algunos de los cuales fueron mencionados en la sección anterior. Con el fin de facilitar su instalación, ALA provee un conjunto de *playbooks* (archivos similares a *scripts*) de la herramienta de automatización *Ansible*²⁵, la cual es una plataforma de software libre para configurar y administrar computadoras. *Ansible* combina instalación multi-nodo, ejecuciones de tareas *ad hoc* y administración de configuraciones.

Estos *playbooks* de *Ansible* se localizan en el repositorio *pgcib-install* del sitio de la PGCIB en *GitHub*. *pgcib_install* es una bifurcación (*fork*) de *ala-install*²⁶. Se recomienda clonar este repositorio, mediante el protocolo de control de versiones *Git*²⁷, a una estación de trabajo, para ejecutarlo desde ahí. Esta estación de trabajo debe emplear un sistema operativo tipo *Unix* (ej. *Linux* o *Mac OS*), que son en los que puede ejecutarse *Ansible*. A la fecha de escritura de este documento (marzo de 2018), *Ansible* no está disponible para ejecutarse en *Microsoft Windows*, aunque sí puede configurar máquinas que usen ese sistema operativo²⁸.

²³ Sitio web del marco de trabajo para desarrollo de aplicaciones web *Bootstrap*: <http://getbootstrap.com/>

²⁴ Sitio web de la plataforma de desarrollo colaborativo *GitHub*: <https://github.com/>

²⁵ Sitio web de la herramienta de automatización *Ansible*: <https://www.ansible.com/>

²⁶ Repositorio de *ala-install* en *GitHub*: <https://github.com/AtlasOfLivingAustralia/ala-install>

²⁷ Sitio web del protocolo de control de versiones *Git*: <https://git-scm.com/>

²⁸ Documentación de *Ansible*, soporte para *Microsoft Windows*: http://docs.ansible.com/ansible/latest/intro_windows.html#windows-support

La documentación del portal de ALA en *GitHub*²⁹ describe dos opciones principales para realizar la instalación:

1. En una máquina virtual en una estación de trabajo local. Esta opción es apropiada para efectos de desarrollo y pruebas. Para este caso, se describe el procedimiento para la instalación en una máquina virtual del hipervisor *VirtualBox*³⁰ configurada mediante la herramienta *Vagrant*³¹, para la administración de máquinas virtuales.
2. En una máquina en la nube (ej. un servicio de alojamiento en Internet [*hosting*]). Esta opción es más apropiada para un ambiente de producción.

En cualquier caso, se recomienda que la máquina en la que instale el portal use el sistema operativo *Ubuntu 16.04 (Xenial Xerus)*³², que es el que ha sido mejor probado, aunque pueden encontrarse configuraciones de máquinas virtuales para otras versiones de Ubuntu en el sitio de ALA en GitHub. Para un servidor de pruebas, se recomiendan 100 GB disponibles de espacio en disco (preferiblemente de estado sólido), 32 GB de memoria RAM y 2 CPU. Para un ambiente de producción, pueden probarse diferentes configuraciones, que incluso pueden separar los diferentes componentes (ej. motores de bases de datos, índices, servidores web) en servidores separados. Pueden encontrarse varios ejemplos de estas configuraciones en el documento “*Atlas of Living Australia: Key Technical Document*” (Cavière et al., 2016).

²⁹ Documentación del portal de ALA en *GitHub*:
<https://github.com/AtlasOfLivingAustralia/documentation/wiki>

³⁰ Sitio web del hipervisor *VirtualBox*: <https://www.virtualbox.org/>

³¹ Herramienta para la administración de máquinas virtuales *Vagrant*: <https://www.vagrantup.com/>

³² Sitio web del sistema operativo *Ubuntu 16.04*: <http://releases.ubuntu.com/16.04/>

4. Portal de comunicación

4.1. Características

El portal de comunicación de la PGCIB contiene información general y noticias sobre la iniciativa. La herramienta informática utiliza múltiples medios (ej. textos, imágenes) para de una forma amigable informar e invitar a los visitantes a utilizar los servicios de la plataforma. El menú principal del sitio está presente en todas sus páginas y presenta las siguientes opciones:

- Datos: esta opción contiene un hipervínculo que conduce al usuario al portal de datos.
- Acerca de: Contiene el siguiente submenú.
 - ¿Qué es la PGCIB?: Texto introductorio sobre la PGCIB que incluye temas como gobernanza, contexto nacional, misión y servicios que brinda la iniciativa.
 - ¿Cómo usar los datos? Contiene documentación para orientar a los usuarios al respecto de los tipos de datos integrados y cómo aprovechar los datos y servicios disponibles en al PGCIB.
 - ¿Cómo publicar sus datos en la PGCIB? Incluye los lineamientos que debe seguir una institución o investigador para publicar sus datos con éxito en la PGCIB.
 - Contáctenos: Informa al usuario sobre las distintas formas de contactar o enviar sus consultas y comentarios a los representantes de la PGCIB.
 - Noticias: Brinda un espacio para mantener informados a los usuarios sobre acontecimientos de importancia para la PGCIB.

Requerimientos generales del portal de comunicación:

- La interfaz del portal es Bilingüe, está disponible en español e inglés. Cuando se disponga de los textos traducidos al inglés serán incorporados al sistema.
- La página principal incluye los logos de las instituciones que lideran y financian el proceso de implementación del proyecto.

4.2. Herramientas informáticas utilizadas

Para la implementación del Portal de comunicación se seleccionó el administrador de contenidos (CMS por sus siglas en inglés) WordPress que permite crear sitios web con

múltiples servicios como blogs, chats, noticias, entre otros. WordPress es software de código abierto desarrollado por una comunidad global de colaboradores. Documentación de la herramienta y funcionalidad complementaria (*plugins*) está disponible en el sitio web para desarrolladores³³. A nivel global, WordPress es el CMS más utilizado, más de 19.5 millones de sitios lo utilizan, lo que representa aproximadamente un 35% del mercado de los CMS³⁴.

Para ejecutar WordPress, se requiere utilizar las siguientes herramientas informáticas:

- Procesador de hipertexto PHP (*PHP: Hypertext Preprocessor*) versión 7.0 o superior: PHP es un lenguaje de programación basado en *scripts* (i.e. secuencias básicas de comandos interpretados por un servidor web con un módulo de procesador de PHP) que permite a los desarrolladores escribir páginas web generadas dinámicamente. El código fuente de PHP es abierto y especialmente adecuado para el desarrollo de aplicaciones web ya que puede integrarse fácilmente al HTML³⁵. WordPress fue desarrollado utilizando PHP.
- MySQL versión 5.6 o superior o MariaDB versión 10.0 o superior: MySQL y MariaDB son servidores de base de datos que permiten almacenar los componentes web de los sitios implementados con WordPress. El que los componentes sean almacenados en una base de datos hace que el desarrollo y la administración de los sitios web sea flexible y eficiente, por ejemplo, los privilegios de acceso de los administradores del sitio se pueden administrar de forma modular por medio de perfiles de usuario.
- Servidor web (HTTP): Cualquier servidor HTTP que tenga soporte para PHP puede ser utilizado, sin embargo, los requerimientos descritos en la documentación de WordPress³⁶ recomiendan Apache o Nginx como servidores robustos, seguros y eficientes que implementan el estándar HTTP. Para la instalación del portal de comunicación de la PGCIB se seleccionó Apache porque es el servidor web más utilizado a nivel global.

4.3. Requerimientos para la instalación

Los requerimientos de memoria, disco y CPU de WordPress son mínimos, cualquier servidor moderno puede ser utilizado con éxito para implementar un portal de comunicación con WordPress. Sin embargo, los requerimientos de hardware y software definidos para la PGCIB se basan en la cantidad de recursos que se espera utilice la aplicación en sus primeras etapas de desarrollo, cuando se inicie con las pruebas de

³³ <https://developer.wordpress.org/>

³⁴ <https://trends.builtwith.com/cms>

³⁵ <http://php.net/manual/en/preface.php>

³⁶ <https://wordpress.org/about/requirements/>

integración al portal de otros tipos de datos como imágenes, páginas de especies, referencias bibliográficas, entre otros. El servidor de la PGCIB posee 200 GB de disco duro, 6 GB de memoria RAM, es accesible por Internet y utiliza el sistema operativo Ubuntu.

5. Glosario

Application Programming Interface (API, Interfaz de Programación de Aplicaciones)

Conjunto de subrutinas, estructuras de datos, variables y otros elementos de programación que se pone a disposición de los programadores (ej. por medio de una biblioteca contenida en un archivo o de una interfaz web) para que construyan aplicaciones de software.

Caso de uso

En el contexto de la ingeniería de software, es un conjunto de acciones o eventos que definen las interacciones entre un sistema y un actor (i.e. un usuario o un sistema externo) para lograr un propósito determinado (ej. importar un conjunto de datos, desplegar un gráfico, mostrar un mapa). Usualmente, se especifica a través de una descripción textual acompañada de un diagrama en un lenguaje gráfico (ej. UML).

Código fuente

Conjunto de instrucciones, escritas como líneas de texto en un lenguaje de programación (ej. Python, R, Java), correspondientes a los pasos o acciones que realiza un programa de software. El código fuente no es directamente ejecutable por una computadora o cualquier otro dispositivo programable (ej. teléfono celular), antes debe ser convertido a “lenguaje de máquina” por medio de, por ejemplo, un compilador.

Componente

En el contexto de la ingeniería de software, es una parte reemplazable y reutilizable de un sistema, y que se ajusta a un conjunto de interfaces.

Dato

Representación simbólica (numérica o textual) de un atributo o variable cualitativa o cuantitativa. Para convertirse en información, los datos necesita ser interpretados.

Hipertexto

Tipo de hiperenlace en el que se vinculan elementos de texto.

HyperText Markup Language (HTML, Lenguaje de Marcas de Hipertexto)

Lenguaje de marcado (i.e. que incorpora marcas o etiquetas con información adicional sobre la estructura o la presentación del texto) estandarizado por el *World Wide Web Consortium* (W3C - Consorcio WWW) para la elaboración de páginas web.

Información

Conjunto organizado de datos interpretados y, generalmente, procesados. Constituye un mensaje o una respuesta a alguna pregunta.

Informática de la Biodiversidad (IB)

Disciplina dedicada a la aplicación de técnicas informáticas a datos e información sobre biodiversidad con el fin de mejorar su captura, administración, presentación, descubrimiento, exploración y análisis.

Metadatos

Son datos (o información) que describen otros datos. En el caso de los datos de biodiversidad (ej. en un conjunto de registros de presencia de especies), los metadatos incluyen elementos tales como el autor del conjunto, las direcciones de contacto y su cobertura geográfica, temporal y taxonómica.

Modelo

Abstracción de un sistema (i.e. conjunto de características que lo diferencian de otros) que lo describe desde cierto punto de vista y en un determinado nivel de detalle. Puntos de vista son, por ejemplo, la vista de casos de uso o la vista de análisis de un sistema de información.

Paquete

En el contexto de UML y del Proceso Unificado, es un mecanismo de propósito general para organizar elementos en grupos.

Plataforma para la Gestión de Conocimiento e Información Nacional de Biodiversidad (PGCIB)

Plataforma web del Minae para la sistematización, documentación y publicación de la información sobre la biodiversidad de Costa Rica.

Proceso Unificado de Desarrollo de Software (Proceso unificado)

Metodología de desarrollo de software que se caracteriza por ser dirigida por casos de uso, centrada en la arquitectura e iterativa e incremental. Utiliza el UML como lenguaje visual para la diagramación de los artefactos contemplados por la metodología (ej. diagramas de casos de uso, diagramas de secuencia, diagramas de colaboración).

Prototipo

En el contexto de la ingeniería de software, un prototipo es una versión limitada de un producto de software, la cual no contiene toda la funcionalidad contemplada para el producto pero que permite a los usuarios explorarlo o probarlo en algunas situaciones reales, con el fin de proporcionar retroalimentación para mejorarlo, así como para advertir posibles fallas.

Registro de presencia de especie (*Species Occurrence Record*)

Registro en una base de datos (u otro medio digital) que detalla hechos acerca de la presencia de un organismo: su taxonomía, ubicación, así como la fecha en el que fue observado o recolectado, entre otros elementos de datos. Ejemplos típicos de registros de presencia de especies son los relacionados a los especímenes de una colección de historia natural o a las observaciones realizadas por ciudadanos naturalistas (ej. observadores de aves).

Servicio web

Componente de software al que se accede a través de la Web y que acepta consultas y retorna resultados en formatos estandarizados (ej. XML, JSON).

Sistema de información

Conjunto de tecnologías, procesos y personas orientados a la recolección, organización, almacenamiento y comunicación de datos e información, con un fin determinado. Los componentes principales de un sistema de información incluyen el *hardware* (i.e. componentes eléctricos, electrónicos y mecánicos); el *software* (i.e. programas de aplicación, sistema operativo y otros programas de soporte); los datos; la información; las redes de computadoras y dispositivos; y las personas que desarrollan, mantienen y utilizan el sistema.

Sistema Nacional de Información Ambiental (Sinia)

Plataforma oficial de coordinación y vinculación institucional y sectorial del Estado costarricense para facilitar la gestión y distribución del conocimiento de la información ambiental nacional.

Software libre

Software distribuido en términos que le otorgan a los usuarios la libertad de ejecutarlo, copiarlo, distribuirlo, estudiarlo, modificarlo y mejorarlo.

Subsistema

En el contexto de UML, es una agrupación de elementos (ej. paquetes, casos de uso) de los que algunos constituyen una especificación del comportamiento ofrecido por los otros elementos contenidos.

6. Referencias

Cavière, F., Figueira, R., Heughebaert, A., Lecoq, M., Martínez de la Riva, S. 2016. *Atlas of Living Australia: Key Technical Document (version 1.0.4, 6 July 2016)*. Publicado por el Secretariado de GBIF. Recuperado el 10 de noviembre de 2017 de <https://www.gbif.org/document/82847/ala-key-technical-documentation-english>.

Belbin, L., Williams, K.J., Towards a national bio-environmental data facility: experiences from the Atlas of Living Australia. *International Journal of Geographical Information Science* 2016 30(1), p. 108-125 <http://dx.doi.org/10.1080/13658816.2015.1077962>.

Jacobson, I., Booch, G. & Rumbaugh, J. (1999). *The Unified Software Development Process*. Reading: Addison Wesley.

7. Apéndices

Apéndice I: Miembros del equipo de análisis de requerimientos de los usuarios del sistema (Earus)

Nombre	Institución	Cargo
Federico Matarrita	ACG	Especialista en Informática, asesor del ACG
Ángela González Grau	Conagebio	Directora Ejecutiva
José Alfredo Hernández Ugalde	Conagebio	Área Técnica - Biólogo
Maribel Alvarez	Conagebio	Asesoría Legal
Melania Muñoz	Conagebio	Área Técnica - Bióloga
Lenin Corrales	Consultor independiente	Consultor independiente
Erick Mata Montero	ITCR	Profesor Asociado Escuela de Computación
Diana Montero	Micitt	Asesora, Dirección de Investigación y Desarrollo Tecnológico
María Celeste Brenes	Micitt	Dirección Investigación y Desarrollo Tecnológico
Armando Estrada	MNCR	Coordinador Técnico del Herbario Nacional Dpto. Historia Natural.
Esteban Quirós	MNCR	Jefe de la Unidad de Tecnología de Información

Nelson Zamora	MNCR	Investigador Asociado
Álvaro Aguilar Díaz	Sinac	Director Centro Nacional de Información Geoambiental
Ana María Monge Ortiz	Sinac	Encargada de Investigación
Eugenia Arguedas	Sinac	Coordinadora Pronamec
Gustavo Induni Alfaro	Sinac	Coordinador de Investigación Depto. de Conservación y Uso Sostenible Secretaría Ejecutiva
Mauricio Castillo	Sinac	Gerente Información y Regularización del Territorio
Manuel Vargas	Consultor	Consultor, especialista en Informática de la Biodiversidad
María Auxiliadora Mora	Consultora	Consultor, especialista en Informática de la Biodiversidad