

ESPECÍMENES E IMÁGENES (DIGITALES)

Francisco Pando - GBIF España

Colecciones Biológicas 3.0

Red Iberoamericana de Información sobre Biodiversidad.
Villa de Leyva. Colombia 2012



Guion

- Formatos
- Resolución
- Captura
- 3D alternativo
- Recomendaciones
- Publicación de imágenes

Formatos

JPEG (o JPG, para la extensión de archivo; Joint Photographic Experts Group)

- Compresión con pérdida, si web, no animación, no transparencia, muy compacto
- PNG (Portable Network Graphics)
 - Color verdadero, transparencia animación, no web
- BMP (Bitmap)
 - Contenedor, de Windows, sin compresión
- GIF (Graphics Interchange Format)
 - Color indexado, transparencias, animación , no web
- TIFF (TIF o, por extensión, Tagged Image File Format)
 - Contenedor, con o sin compresión, no web,
- <http://www.brighthub.com/multimedia/photography/articles/10036.aspx>

Resolución

¿Cuánto es suficiente o recomendable?

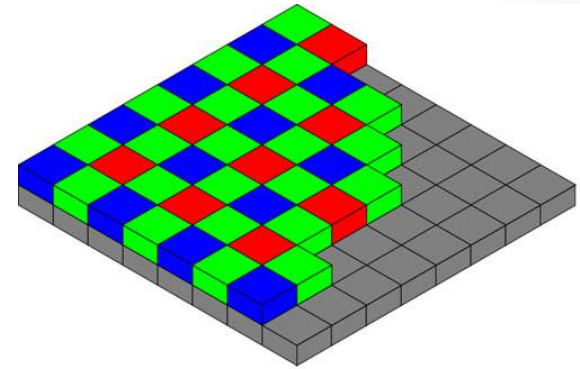
300 dpi

comprimido o no

... algunos dicen 600 dpi:

“Plant sheets are usually scanned at somewhere around 1000 DPI (600 DPI being now generally considered the absolute minimum requirement), which renders images in the hundred-megabyte range.” Ariño and Galicia (2005) in Christoph L. Häuser, Axel Steiner, Joachim Holstein & Malcolm J. Scoble. (eds) (2005) Digital Imaging of Biological Type Specimens: A Manual of Best Practice. ENBI, Stuttgart

Pero atención a la interpolación



Captura

- Scanner
 - Invertido
 - Cenital
 - 1400- 300 DPI

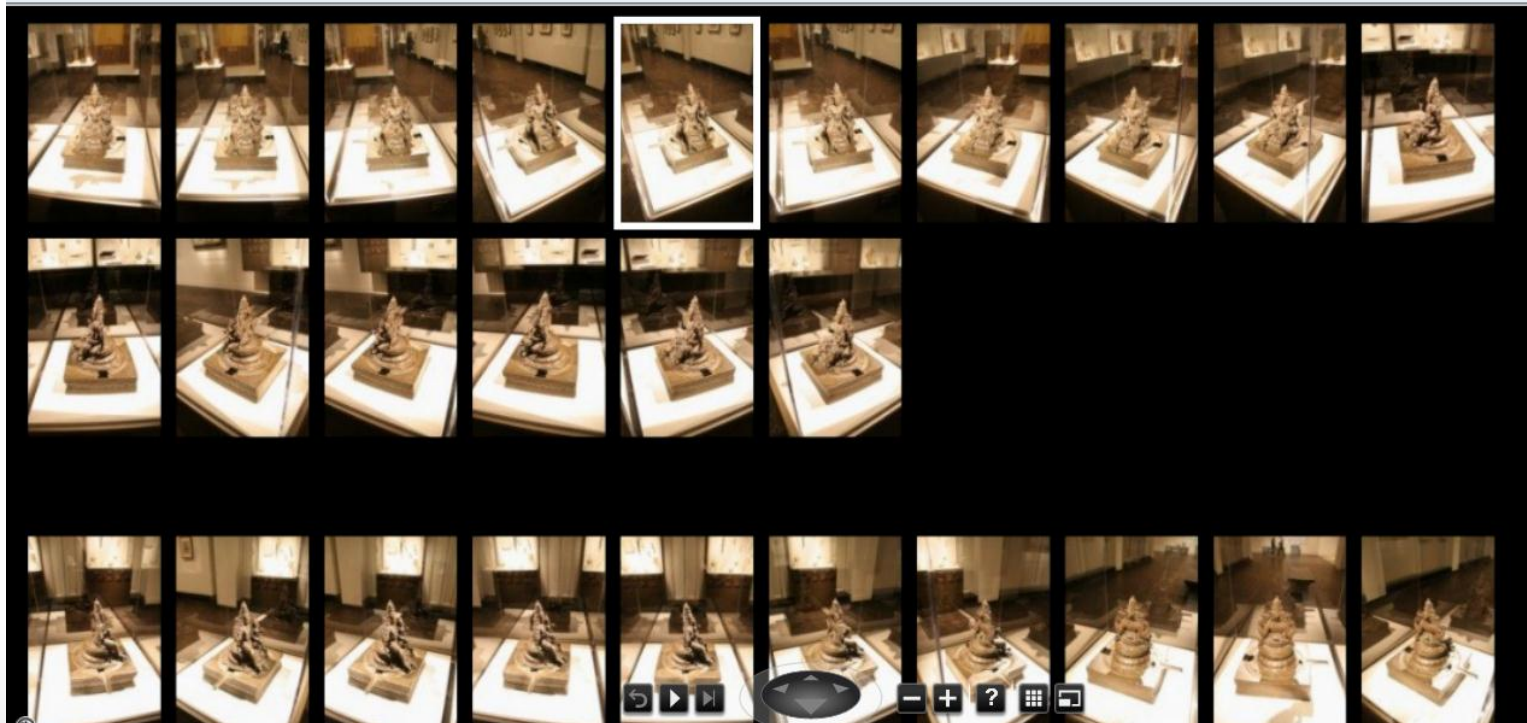


- Cámara
 - c. 300 DPI



Photosynth (Alternativa a 3D)

- <http://photosynth.net/view.aspx?cid=7a1eafc3-07ce-415f-8a9d-67c42d8a90c5>
- <http://photosynth.net/view.aspx?cid=4bb0ea1d-fd0f-4eb9-b347-21101222b4a8>



Para saber más



http://www.gbif.es/imagenes_in.php

http://sciweb.nybg.org/science2/herbarium_imaging/

THE NEW YORK BOTANICAL GARDEN *Vascular Plant Type Specimens Imaging Project*

(Page last updated July 29, 2002)

<http://mcz-28168.oeb.harvard.edu/etypes/index.htm>

The New York Botanical Garden has completed the project to photograph its approximately 89,000 vascular plant type specimens.

This project is part of the E-Type Initiative, a project to digitize information about its type specimens. Over 82,000 specimens are now available in the [Virtual Herbarium of the New York Botanical Garden](#).

The digital images are available to replace the actual specimens in the herbarium. The [Plant Type Catalog](#) provides information about individual specimens through the Internet. The first 10,000 specimens selected to image first in the herbarium.

This project has been funded by the [Biomedical Photo](#)



Bioinformatics | MCZ Type Database | Caribbean Insects

The E-Type Initiative @ **Harvard Entomology**

[Imaging Considerations](#) | [Imaging Protocols](#) | [Databasing](#) | [Downloads](#) | [Starter Kit](#) | [General Information](#)



A first phase of the ALL Species 25 year project, the [E-Type Initiative](#) was established to develop Web-accessible electronic cataloging and imaging of primary type specimens that are available for use by taxonomists and others in the research community. The goal is to image and web-enable 50 percent of primary type specimens in five years. To reach this level most effectively, focus has initially been concentrated at about 100 primary type specimens, focusing on the 100 of all specimens.

Almacenar metadatos de las imágenes

¿Que?

- Datos técnicos
- Datos de color
- Metadatos de gestión/manejo (incluido IPR)
- Metadatos de contenido

Almacenar metadatos de las imágenes ¿Como?... estándares

- EXIF (usado por los fabricantes de cámaras digitales)
- IPTC (Desarrollado por el International Press Telecommunications Council)
- XMP Extensible Metadata Platform, Adobe
- Z39.87 (NISO std.) U.S.A.

Algunas recomendaciones habituales

Almacenar metadatos de las imágenes ¿Donde?

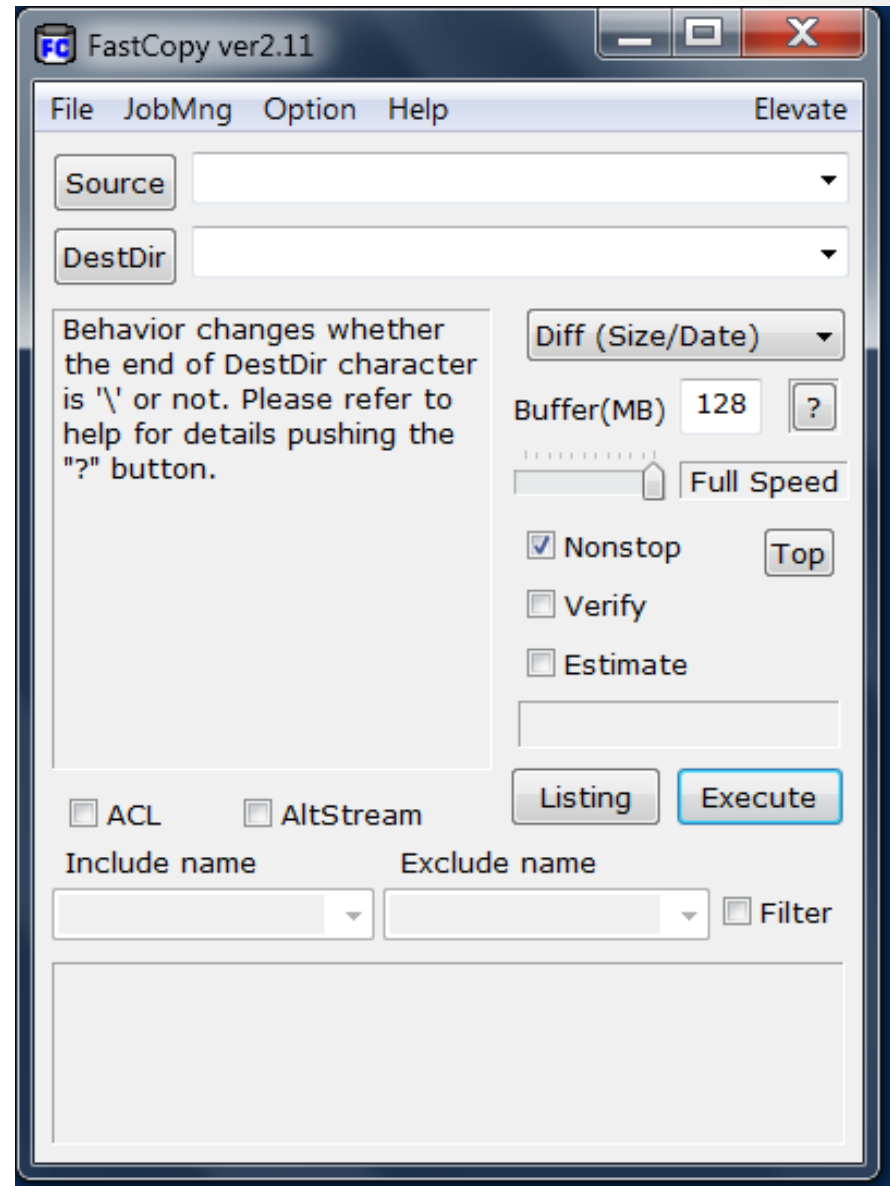
- a) Via database
- b) dentro de la imagen (EXIF, IPTC)
- No en nombre del archivo, no en el nombre de la carpeta

Algunas recomendaciones habituales

Copiar ficheros

- Comprobar la integridad de la copia
- Hash (Funciones resumen) MD5, SHA1

<http://ipmsg.org/tools/fastcopy.html.en>



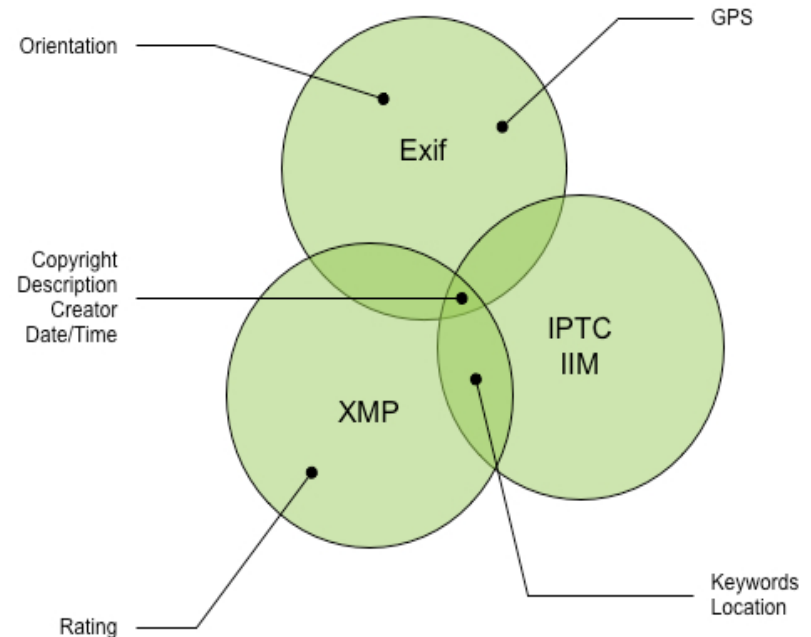
PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

- Metadatos
- Formatos de imágenes de alta resolución: TIFF Piramidal y JPEG 2000
- Protocolo IIP
- Qué es un servidor de imágenes. Soluciones en el mercado
- Visualizadores web de imágenes de alta resolución
- Modelos de trabajo propuestos
- Proyectos y bibliotecas digitales con información sobre biodiversidad en Internet
- Servicio de Alojamiento y Publicación de imágenes de GBIF.es:
 - -

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

Metadatos

- **Metadatos:**
 - Organización & Búsqueda
 - Qué + Dónde + Cuándo + Cómo + Quién
 - embebidos vs externos
- Distintas especificaciones (EXIF, IFDs, Adobe XMP, PSIR, IPTC-IIM)
 - Reglas cómo almacenar + ordenar + codificar
 - Cadenas de caracteres, números, arrays
 - Leídos y modificados vs leídos
 - Solapamiento información
 - Compatibilidad
- Iniciativas de trabajo. “Metadata Working Group”.
 - 2006 by Adobe Systems, Apple, Canon, Microsoft and Nokia. Sony se unió en in 2008.
 - <http://www.metadataworkinggroup.org/>
 - interoperabilidad
 - disponibilidad



Fuente img: "Guidelines For Handling Image Metadata" - Metadata Working Group

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

Metadatos

- EXIF
 - En TIFF, JPEG y PSD files
 - No soportado por PNG o GIF y
 - Estandarizado en XML para JPEG 2000 según ISO TC42 (12234-1.4.)
 - Date & time, camera settings (camara model, orientation, aperture...), thumbnail
 - Usado por cámaras fotográficas.
 - version 2.3 Abril 2010
- IPTC-IIM (*Information Interchange Model*)
 - En JPEG, TIFF
 - No soportado por GIF, PNG
 - <http://www.iptc.org>
 - “IPTC headers” ampliamente adoptado por software para edición
- XMP (*Extensible Metadata Platform*)
 - <http://www.adobe.com/products/xmp>
 - XML y RDF
 - Información de procesos por los q la imagen pasa. Dispositivos y software
 - Dublin core +Elementos extensibles + propios
 - En PDF, JPEG, JPEG 2000, GIF, PNG, HTML, TIFF, Adobe Illustrator, PSD, MP3, MP4,...
 - “IPTC Core Schema for XMP”

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

Metadatos

Metadatos biodiversidad

TDWG. Image Interest Group-IIG

<http://www.tdwg.org/activities/img/charter/>
<http://wiki.tdwg.org/Image>

Estándar MRTG (Multimedia Resource Task Group):

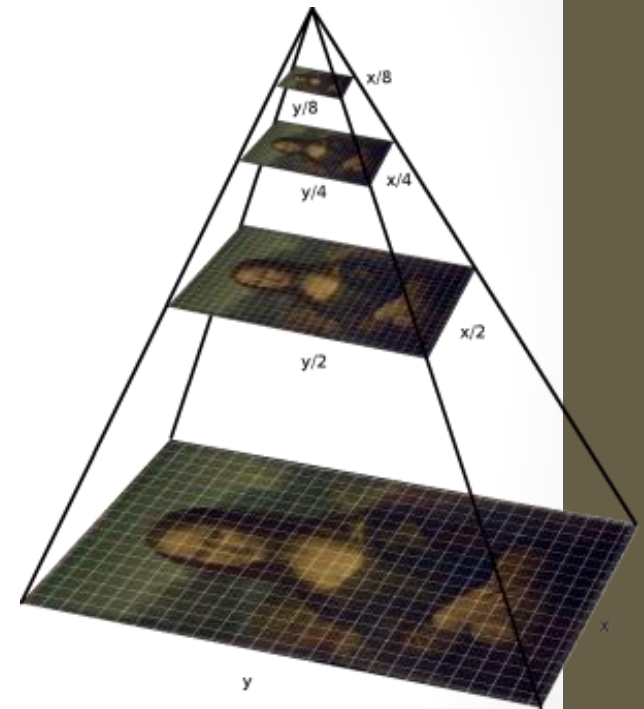
- Estado: Draft
- http://www.keytonature.eu/wiki/MRTG_v1.0 (AudubonCore v1.0)
- Vocabularios para describir metadatos asociados a multimedia en fuentes de datos biodiversidad
- Dublin Core(DC), Darwin Core (DwC), XMP, IPTC, Natural Collections Schema (NCD), etc.

Biodiversity
Information
Standards
T D W G

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

TIFF piramidal

- Formato multi-resolución, datos para múltiples resoluciones y lectura de porciones en distintos niveles de resolución a gran velocidad
- *Almacenamiento en secciones rectangulares.*
(múltiplo de 16 pixeles; mismo tamaño (ancho y alto) todas las secciones)
- Comprimidas individualmente (JPEG,LZW,PAckBits, Huffman)
- Para imágenes de alta resolución
- Especificaciones IIP limitan el tamaño de las secciones a 64×64 pixeles



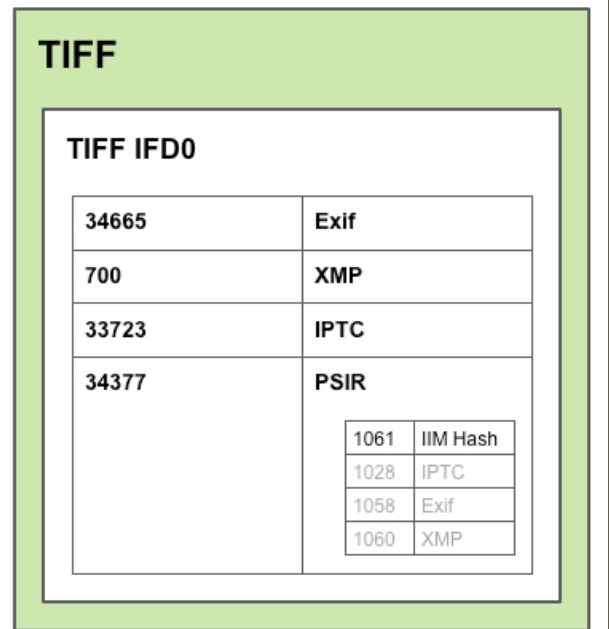
Fuente de la imagen:

<http://iipimage.sourceforge.net/documentation/images/>

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

TIFF piramidal

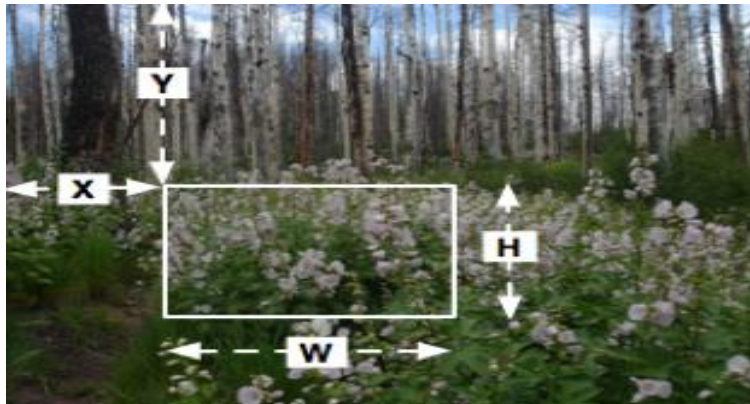
- Espacios de color de CMYK y CIELAB
- Perfiles ICC embebidos, transparencia, capas de Photoshop
- 16-bit por canal de datos
- Procesadores:
 - VIPS image processing suite <http://vips.sourceforge.net>
 - ImageMagick <http://www.imagemagick.net/script/index.php>
 - Implementaciones: JMagick, IMagick, PythonMagick
 - Desde la línea de comandos.
 - Automatización
- Metadatos: EXIFF, IPTC y XMP



PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

JPEG 2000

- Segunda versión del estandar JPEG (compresión y formato). ISO/IEC 29 WG01
- <http://www.jpeg.org/jpeg2000/index.html>
- Algoritmo DWT (wavelet technology)
- Compresión
 - 20%+
 - menor degradación
 - lossless vs lossy
- Multiresolución.
 - Code/Decode por regiones: visualización y transmisión a diferentes calidades & navegación por la imagen
 - Rápida accesibilidad. Transmisión progresiva en:
 - Precisión (pixel)
 - Resolución



Fuente img: <http://www.dlib.org/dlib/september08/chute/09chute.html>

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

Internet Imaging Protocol (IIP)

- Digital Imaging Group (DIG)
- Cliente-Servidor
- Sobre HTTP (imágenes y metadatos)
- Trabaja sobre “multiresolution tiled images” (TIFF piramidal, FlashPix, JPEG 2000)
- Descarga por cliente de porción de imagen (a determinada resolución)
- Servidas y comprimidas dinámicamente en JPEG
- Actualmente version 1.05.
- Especificaciones: <http://iipimage.sourceforge.net/IIPv105.pdf>
<http://iipimage.sourceforge.net/documentation/protocol/>

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

Qué es un Servidor de imágenes

- **Software especializado en servir imágenes. (creadas por demanda basadas en una petición específica)**
- **Arquitectura cliente-servidor (API's públicas)**
- **Almacenamiento (formato base) y generación dinámica de x formatos (tamaños/resolución).**
- **Herramientas para navegación en tiempo real por imágenes de tamaño mega y gigapixel.**
- **Optimización tráfico cliente – servidor (ancho de banda) y eficiencia en gestión de la memoria del servidor**
- **Procesos extra: gestión de metadatos, ajustes de color, usuarios y grupos para control de acceso etc**
- **Ajuste dinámico:**
 - **tamaño de imagen - resolución navegador.**
 - **formatos - tipo/versión del navegador**
- **Formatos base: TIFF (tiled) piramidal, FlashPix/OpenPix, JPEG2000**

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET
Servidores de imágenes: Software

IIPImage



<http://iipimage.sourceforge.net/> Open Source

Servidor: C++

Varios clientes: AJAX, flash, JAVA applet

Imágenes 8/16 bit

Imágenes alta resolución (mega, gigapixel size), multispectrales, objetos 3D

Formato generado: JPEG

Soporte JPEG2000 (Kakadu JPEG2000 library) desde enero 2010

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET
Servidores de imágenes: Software

DJATOKA

Open source. <http://sourceforge.net/projects/djatoka/>

Java

Version 1.1 2009

Formato base: JPEG 2000

Kakadu JPEG 2000 libreria (Extracción dinámica en múltiples resoluciones de regiones de archivos jpeg 2000)

Formatos de entrada/salida: BMP, GIF, JPG, PNG, PNM, TIF, JPEG 2000

Service framework, basado en “OCLC OpenURL Resolver”

Caché configurable

Linux/UNIX/MacOS-X/Windows

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET
Servidores de imágenes: Software

<http://www.yawah.com/> (Requiere licencia) (eRez)



ADOBE® SCENE7®

<http://billeder.silkeborgkommune.dk/erez4/html/index.html>

Servidor: Java

Cliente: Flash, AJAX, Javascript

Gestión de metadatos asociados (IPTC y EXIFF). Interfaz de búsqueda basada en metadatos

Formato base TIFF piramidal, FlashPix o PSD

Carga de imágenes (upload) en diferentes formatos: JPEG, EPS, GIF PDF, AI, SVG y CAD y otros implementando plug-ins. También .zip

Formatos generados: JPEG, GIF, PNG, Flash, PDF y TIFF a través de http requests

Gestión de directorios, usuarios y roles

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET
Servidores de imágenes: Software

Otros :

- iSeeMedia: <http://www.iseemedia.com/>
- MediaRich Server:
<http://www.equilibrium.com/Internet/Equil/Products/MediaRich/Product+Tour/index.htm>
- Adobe Scene 7: <http://www.scene7.com/>
(antes TrueSpectra, eREZ)
- FSI-Server: <http://www.fsi-viewer.com/>



PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

Visualizadores web de imágenes alta resolución

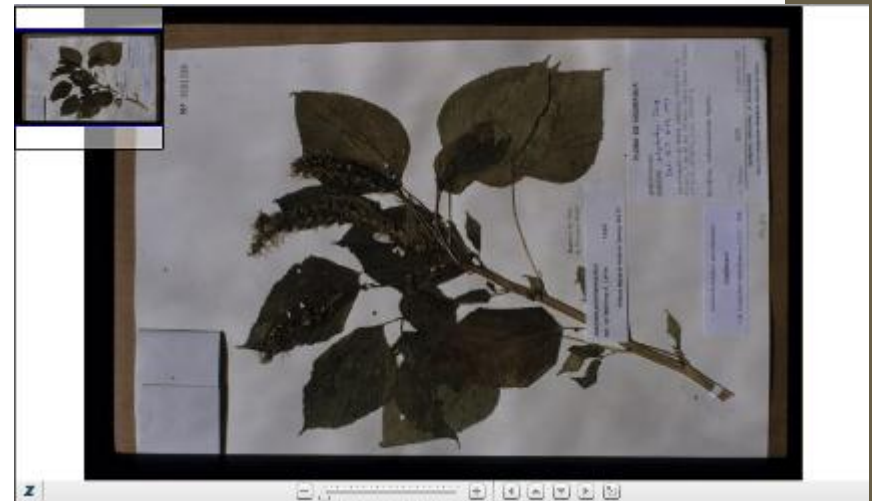
Software para navegación web con imágenes digitales de alta resolución:

Zoom, desplazamiento, rotación, plug-ins (medidas, pase de diapositivas,etc)

Integración a nivel cliente.

Zoomify <http://www.zoomify.com/>

- Sin servidor de imágenes ni TIFF piramidal
- Repositorios de imágenes preprocesadas
- web sites, CDs, or DVDs interactivo
- Multiplataforma: Windows, MacOS, Linux
- Express/Design/Flash/Enterprise
- Zoomify Express: Versión gratuita:
 - Procesador de imágenes + Visor (flash)
- .Para imágenes JPEG (JPEG2000 y TIFF¿implementado?)



PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET
Modelos de trabajo propuestos

Modelo I.

Base de datos (datos biodiversidad: taxonómicos, nomenclaturales, especímenes/observaciones, especies, etc + (metadatos))

+

Repositorio de imágenes (procesadas y en n formatos y resoluciones)

+

Sitio web (Interacción con bases de datos + Visor para imágenes+ Metadatos e imágenes)

Más simple, más espacio en disco, p.e.; con Zoomify

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

Modelos de trabajo propuestos

Modelo II.

Base de datos (datos biodiversidad:taxonómicos, nomenclaturales, especímenes/observaciones, especies etc + (metadatos))

+

Servidor de imágenes (imágenes, una única imagen formato base)

+

Sitio web (Interacción con base de datos + Cliente- servidor imágenes+Metadatos)

Más complejo de montar, un solo fichero por imagen

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

Proyectos y bibliotecas digitales con información sobre biodiversidad en Internet

- Sin servidor de imágenes. Zoomify:

Flora brasiliensis:

<http://florabrasiliensis.cria.org.br/>

- Con servidor de imágenes:

The Biodiversity Heritage Library

<http://www.biodiversitylibrary.org/>
(Djatoka + IIPMooViewer)

Herbario virtual de Santiago de Compostela – SANT

<http://www.gbif.es/Imagenes.php>
(eREZ + FSI viewer)



Fuente img:<http://www.biodiversitylibrary.org/>

PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET

Ejemplo de uso

"Proyecto de digitalización de los dibujos de la Real Expedición Botánica del Nuevo Reino de Granada (1783-1816), dirigida por José Celestino Mutis"

Visualización de imágenes de alta resolución:

Imágenes en la web del Proyecto:

<http://www.rjb.csic.es/icones/mutis>

Imágenes en el Servicio de Alojamiento y Publicación de imágenes de GBIF.es:

<http://www.gbif.es/dwclmg.php?CollectionCode=Arch-Div-III&CatalogNumber=A00088->



PUBLICACIÓN DE IMÁGENES DIGITALES DE BIODIVERSIDAD EN INTERNET
Servicio de Alojamiento y publicación de imágenes de GBIF.ES

- La Unidad de Coordinación de GBIF.es ofrece a los proveedores de datos de la red de GBIF la posibilidad de alojar y/o publicar imágenes de alta resolución en Internet.
- Condiciones de uso
 - Ser proveedor de datos de GBIF
 - Sincronizar la actualización de datos e imágenes (imgs siempre asociadas a datos publicados)
- Condiciones técnicas:
 - Formato base para las imágenes TIFF piramidal
 - Datos necesarios (DarwinCore + Datos imágenes + Metadatos EXIFF e IPTC)
 - <http://www.gbif.es/Alojamiento/ImagenesIndex.php>
 - <http://www.gbif.es/imagenes.php>

A vuestra disposición

Francisco Pando

Responsable

GBIF España, U. de Coordinación

Real Jardín Botánico-CSIC

Pza. de Murillo, 2

28014 Madrid, España

pando@gbif.es

Tel.: + 34 91 420 3017

Fax: + 34 91 429 2405

www.gbif.es

www.gbif.org