

El ciclo de vida de los datos de la biodiversidad

Manuel Vargas
Instituto Nacional de Biodiversidad (INBio)
15 de abril de 2013

Resumen

Se presenta una descripción general de los datos sobre la biodiversidad y un esquema de las diferentes etapas del ciclo de vida de los mismos. Se exponen los aspectos generales del área de conocimiento conocida como Informática para la Biodiversidad y se ofrecen varios ejemplos de aplicaciones relacionadas, particularmente de aquellas que hacen uso de modelos de distribuciones de especies.

Contenidos

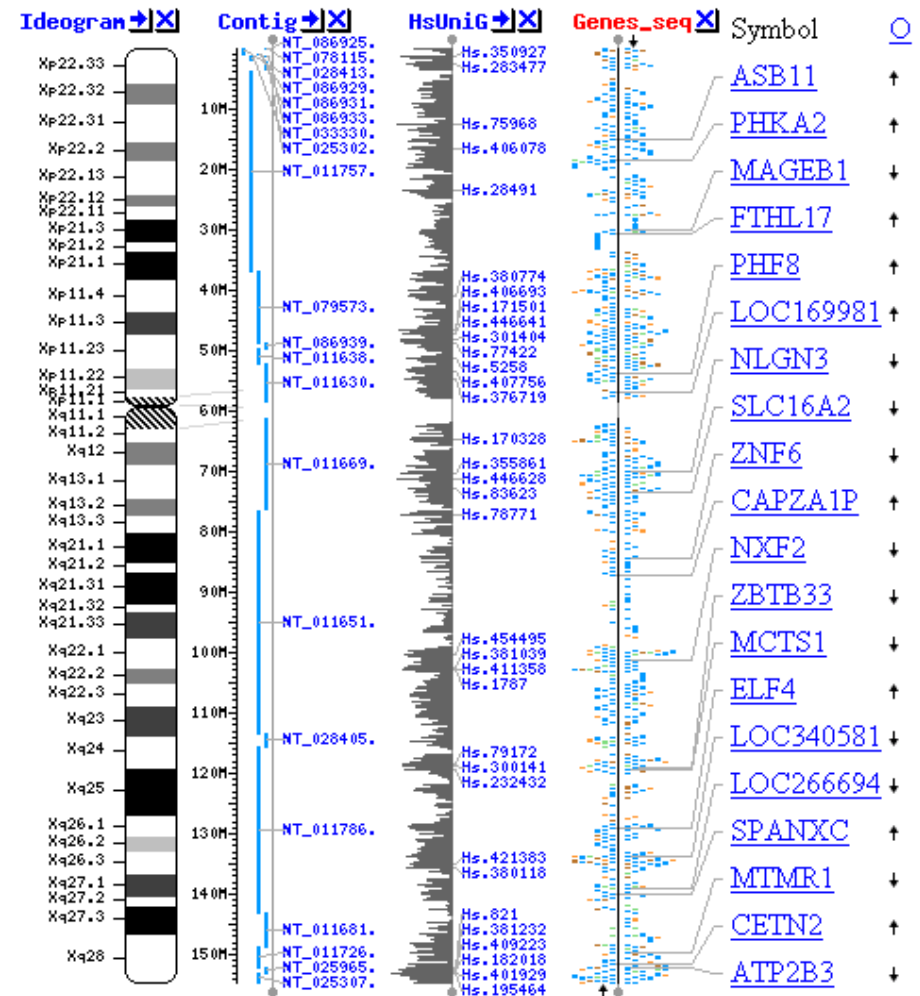
- La biodiversidad y sus datos.
 - Áreas de conocimiento relacionadas con la biodiversidad y con la computación.
 - Datos primarios.
 - Datos secundarios.
- La Informática de la Biodiversidad.
- El ciclo de vida de los datos de la biodiversidad.
- Ejemplos de aplicación de la Informática de la Biodiversidad.

La biodiversidad y sus datos

- Biodiversidad
 - *“Por “diversidad biológica” se entiende la variabilidad de organismos vivos de cualquier fuente, [...]; comprende la diversidad dentro de cada especie, entre las especies y de los ecosistemas” (Naciones Unidas, 1992).*
- Datos relacionados
 - **Moleculares:** Alineamiento de secuencias, genes, estructura de proteínas, ...
 - **Ecosistémicos:** Clima, vegetación, flujos energéticos, relaciones,...
 - **De especies y organismos:** nombres taxonómicos, coordenadas geográficas, nombres de lugares, fechas de recolección, rangos de distribución, ...

Nivel molecular – Bioinformática (*Bioinformatics*)

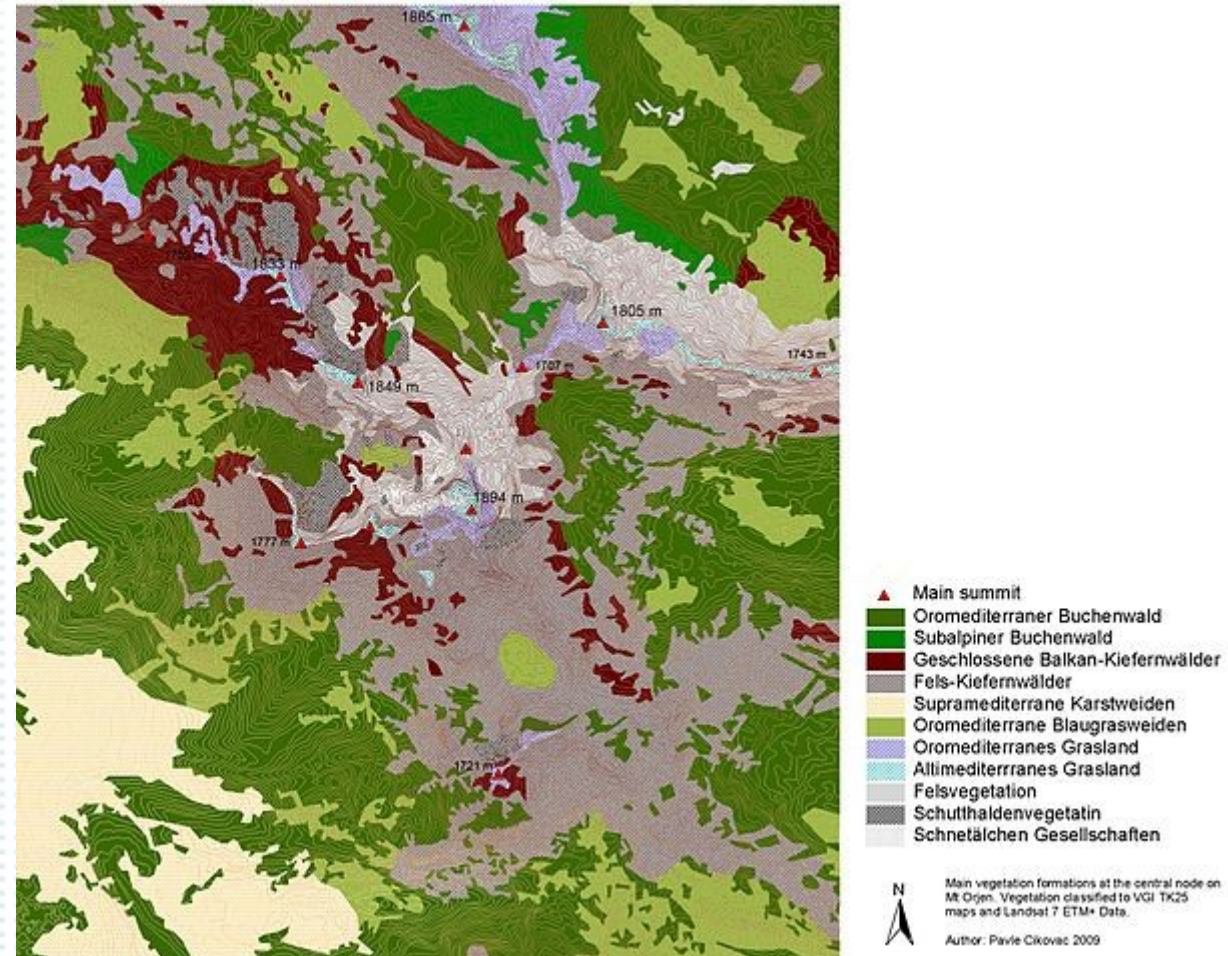
- Rama computacional de la biología molecular.
- Análisis de secuencias de proteínas almacenadas en bases de datos.
- Análisis de secuencias de ADN y ARN almacenadas en bases de datos.
- Interacciones genéticas.
- Árboles filogenéticos.
- Anotación de genomas.



Mapa del cromosoma humano X. La imagen es del National Center for Biotechnology Information (NCBI) y está disponible en Wikimedia (<http://www.wikimedia.org/>).

Nivel ecosistémico – Ecoinformática (*Ecoinformatics*)

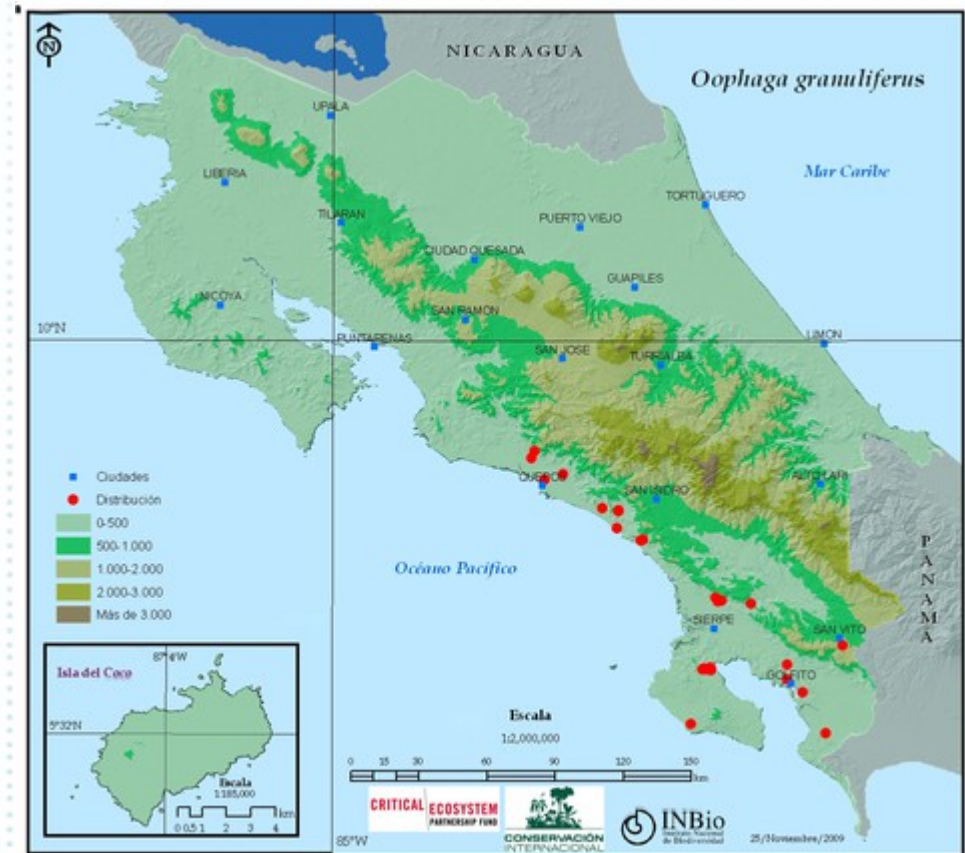
- Integración de la ecología, las ciencias ambientales y las ciencias de la computación.
- Mapeo de ecosistemas.
- Clasificación de la vegetación.
- Dinámica de poblaciones.
- Procesamiento de datos de sensores remotos ambientales.



Clasificación de la vegetación basada en información del satélite Landsat7.
La imagen es de Pavle Cikovic y está disponible en Wikimedia (<http://www.wikimedia.org/>).

Nivel de especies – Informática de la biodiversidad (*Biodiversity Informatics*)

- Aplicación de tecnologías computacionales al manejo de datos sobre biodiversidad del nivel de especies y organismos.
- Digitalización de datos de colecciones de historia natural.
- Georreferenciación de localidades de recolección.
- Integración de datos de museos y herbarios en portales en el web.
- Modelos de distribución de especies.

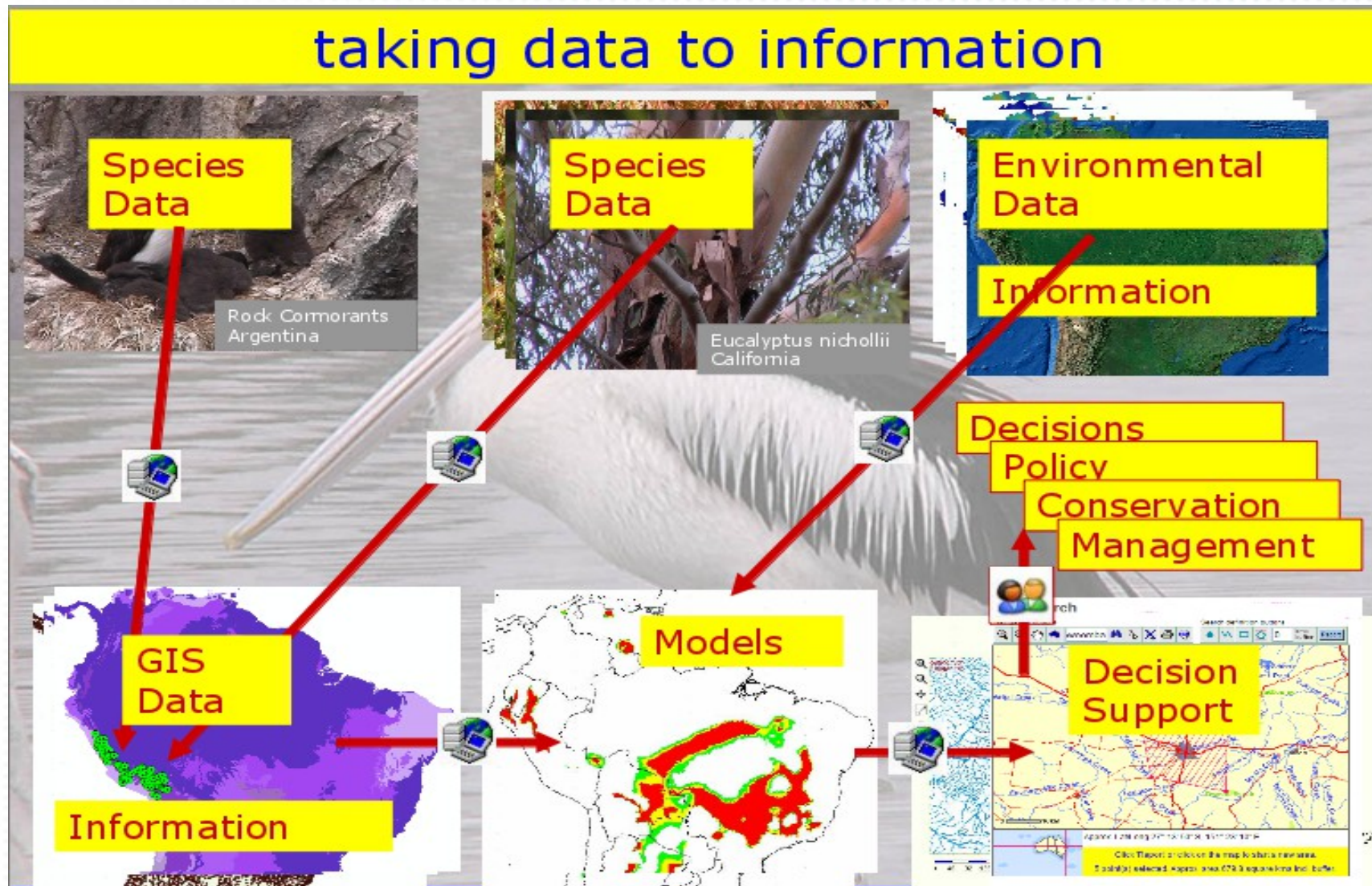


Mapa de distribución de la rana venenosa *Oophaga granuliferus* en Costa Rica. La imagen está disponible en el sitio web sobre especies de Costa Rica del Instituto Nacional de Biodiversidad (<http://darnis.inbio.ac.cr/>).

La Informática de la Biodiversidad (IB)

- La primera mención a la disciplina se dio en 1998 (Schalk, 1998).
- Ha sido impulsada por factores como:
 - La necesidad de encontrar respuestas urgentes a la problemática de la biodiversidad (e.g. impacto del cambio climático, efecto de las especies invasoras).
 - La disponibilidad a gran escala de datos primarios de biodiversidad (i.e. especímenes y observaciones de organismos) en formato digital.
 - La disponibilidad a gran escala de información ambiental (e.g. sensores remotos) en formatos apropiados para manejarse con Sistemas de Información Geográfica (SIG).
 - El desarrollo de herramientas de software apropiadas para almacenar, recuperar, analizar y visualizar esta información (e.g. PostgreSQL/PostGIS, SIG, R).

La Informática de la Biodiversidad



La imagen es de Arthur D. Chapman.

Datos primarios

- Son textos digitales o registros de datos multimediales que detallan hechos acerca de la instancia u ocurrencia de un organismo, i.e. sobre el **qué, dónde, cuándo, cómo y quién** de cada ocurrencia (GBIF, 2010).
- Ejemplos
 - Etiquetas de especímenes en herbarios y museos.
 - Registros de observaciones de aves.
 - Fotografías de “cámaras trampa” (*camera trap*).



Fotografía de un tigre de Sumatra tomada con una “cámara trampa”. La imagen está disponible en Wikimedia (<http://www.wikimedia.org/>).

Datos primarios



Espécimen de una planta de la familia *Apocynaceae* y su correspondiente etiqueta de herbario. La imagen es del Instituto Nacional de Biodiversidad de Costa Rica (INBio).

Fuentes de datos primarios

- Colecciones de historia natural en museos y herbarios.
- Registros de observaciones de vida silvestre (e.g. observadores de aves).
- Portales de biodiversidad en Internet.



Arriba, colección de *Lepidoptera* del Museo de Historia Natural de Londres, imagen del periódico *The Guardian* (<http://www.guardian.co.uk/>). Abajo, observadores de aves en el Parque Central de Nueva York, imagen disponible en Wikimedia (<http://www.wikimedia.org/>).

Datos secundarios

- Son síntesis basadas en datos primarios (Peterson et al., 2002).
- Ejemplos:
 - Mapas de rangos de distribución.
 - Listas de autoridad taxonómica y *checklists*.
 - Descripciones de la historia natural de las especies.

Datos secundarios



Fuente: INBio
© Derechos reservados

Mapa del rango de distribución de la lapa roja (*Ara macao*) en Costa Rica, confeccionado en el Instituto Nacional de Biodiversidad de Costa Rica (INBio).

¿Cómo se generan estos datos?

El proceso comprende varias etapas (Chapman, 1995):

- Recolección.
- Digitalización.
- Edición, validación y curación.
- Presentación.

Etapa de recolección



Las imágenes son del Instituto Nacional de Biodiversidad de Costa Rica (INBio).

- Se realiza en el campo.
- Se recolectan datos geográficos, taxonómicos, ambientales y morfológicos.
- Los datos que se omiten o se registran incorrectamente en esta etapa son difícilmente recuperables en las etapas posteriores.

Etapa de digitalización



Las imágenes son del Instituto Nacional de Biodiversidad de Costa Rica (INBio).



- Suele ser un “cuello de botella” en muchas instituciones.
- Pueden producirse errores (ortográficos, geográficos, taxonómicos, ...) que deben ser corregidos luego.

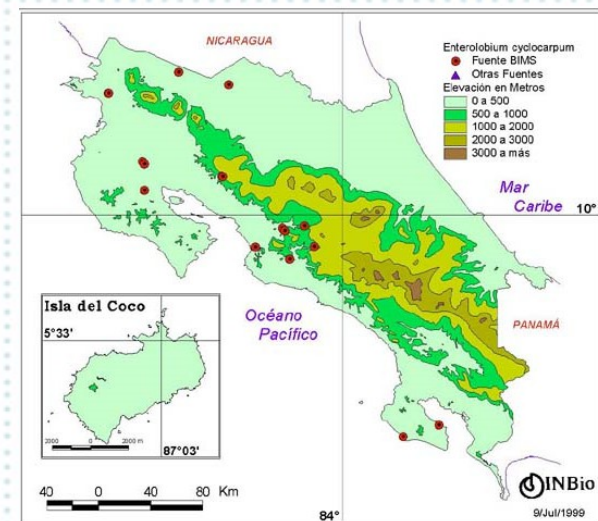
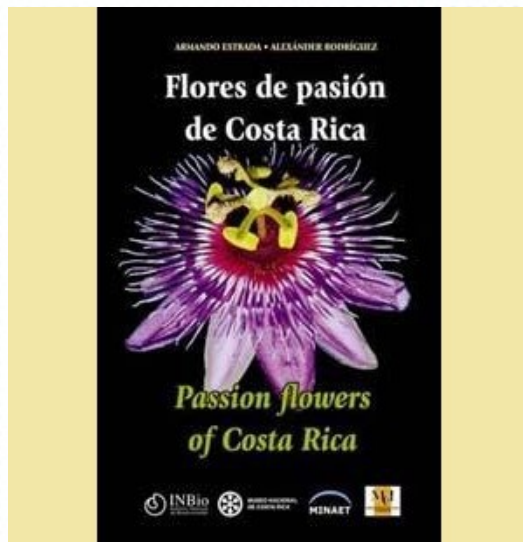
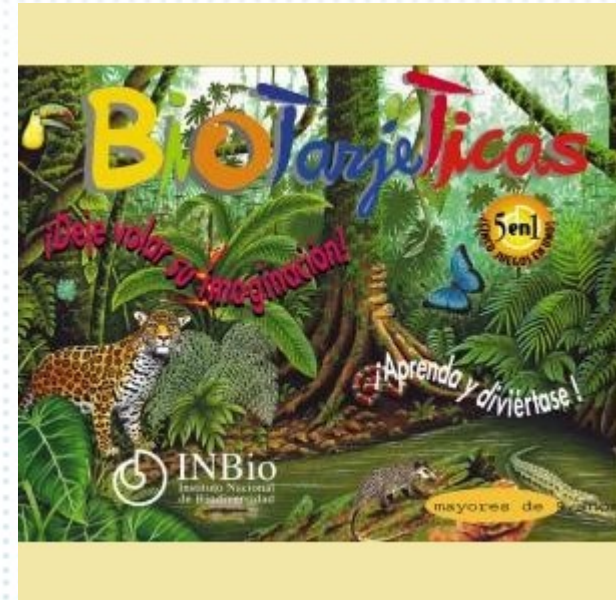
Etapas de edición, validación y curación



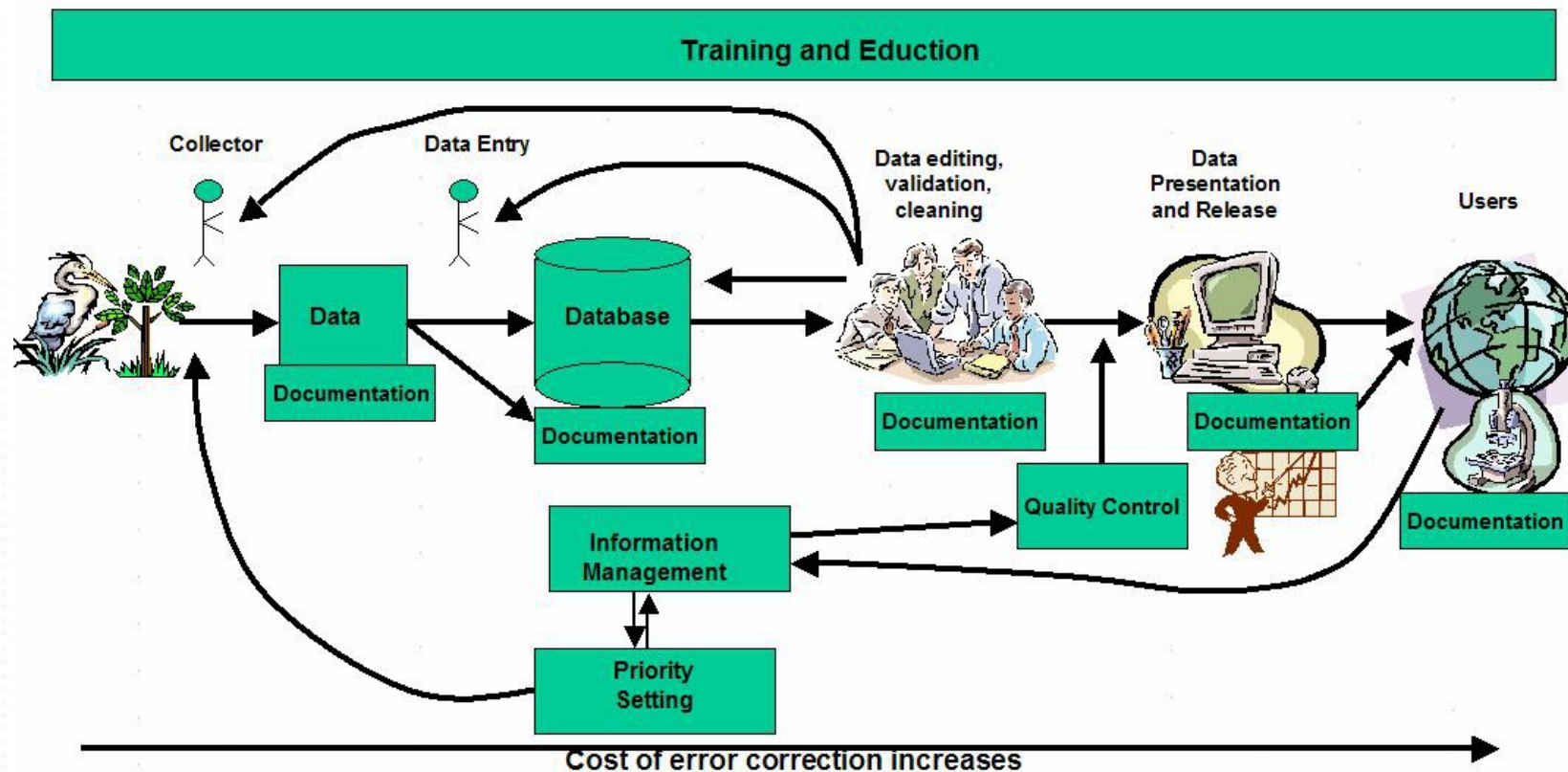
Las imágenes son del Instituto Nacional de Biodiversidad de Costa Rica (INBio).

- Reidentificación.
- Georreferenciación.
- Detección y corrección de errores de etapas anteriores.

Etapa de presentación



Ciclo de vida de los datos de biodiversidad



La imagen es de Arthur D. Chapman (1995).

¿Cómo estructurar los datos generados a través de todo este ciclo de vida?

Hay varios estándares globalmente aceptados para los dos tipos de datos principales:

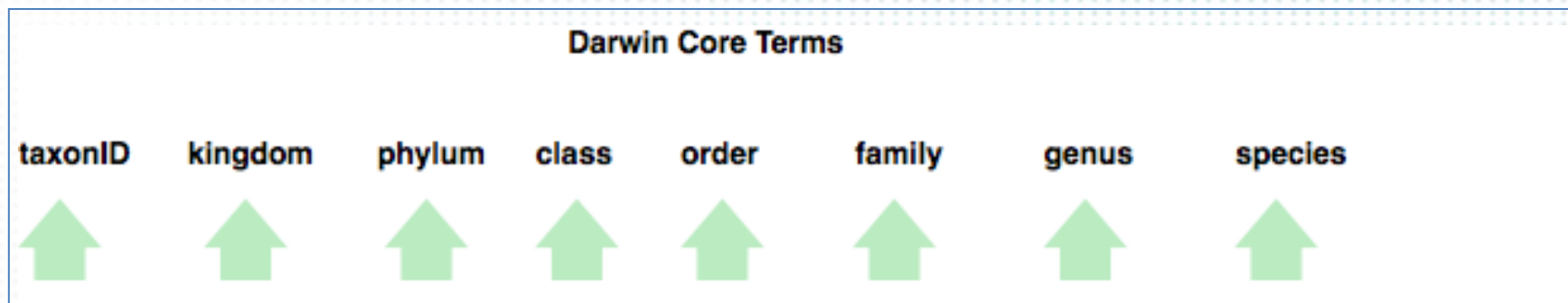
- Para ocurrencias de organismos (i.e. especímenes y observaciones):
 - El *Access to Biological Collections Data* (ABCD, <http://www.tdwg.org/activities/abcd/>).
 - El *Darwin Core* (DwC, <http://www.tdwg.org/activities/darwincore/>).
- Para descripciones de especies:
 - El *Plinian Core* (PLIC, <http://www.pliniancore.org/>).
 - El *Species Profile Model* (SPM, <http://wiki.tdwg.org/twiki/bin/view/SPM/WebHome>).

El Darwin Core

- Es un vocabulario o lista de más de 150 términos (i.e. conceptos o elementos) relacionados con la información contenida en colecciones biológicas. Los términos están agrupados en varias categorías:
 - **Record-level Terms:** *language, accessRights, ...*
 - **Occurrence:** *catalogNumber, sex, lifeStage, ...*
 - **Event:** *samplingProtocol, eventDate, habitat, ...*
 - **Location:** *country, stateProvince, locality, decimalLongitude, decimalLatitude, ...*
 - **GeologicalContext:** *earliestAgeOrLowestStage, latestAgeOrHighestStage, ...*
 - **Identification:** *identifiedBy, dateIdentified, identificationQualifier, ...*
 - **Taxon:** *scientificName, kingdom, phylum, ..., genus, specificEpithet, taxonRank, ...*
- El Dwc funciona como una extensión del Dublin Core para aplicaciones de IB.
- Se pretende que funcione como una referencia estable y estandarizada para compartir datos de biodiversidad.
- Es un estándar de TDWG y es usado y soportado por las principales iniciativas de IB: GBIF, IABIN, ALA, VertNet, MaNIS, ORNIS, Fishnet 2,...

El Darwin Core

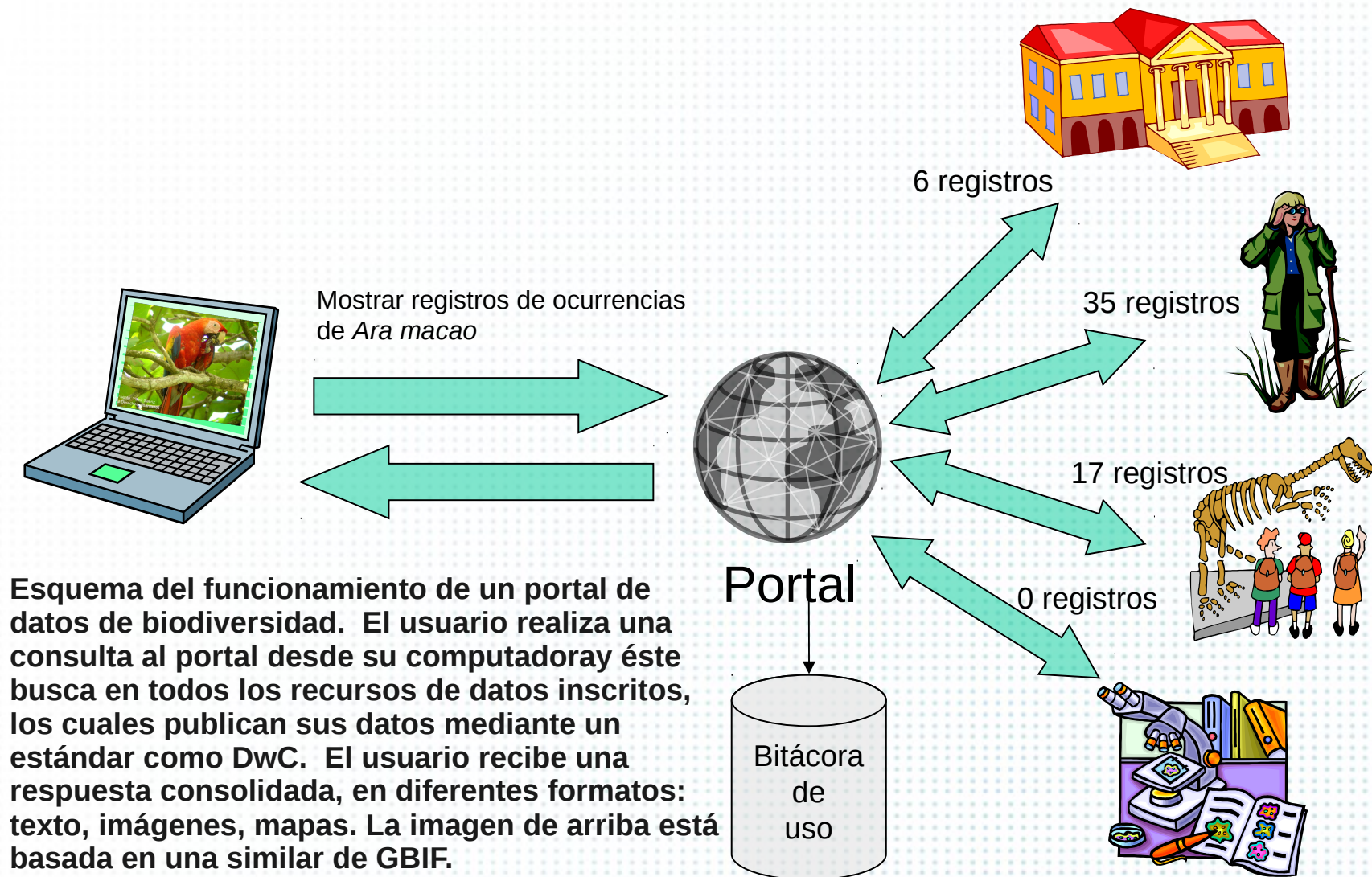
- Las columnas de una base de datos se mapean a los términos del DwC y así pueden compartirse y ser consultadas simultáneamente con otros recursos de información sobre biodiversidad.
- La base de datos de cada museo o herbario puede estar en cualquier formato: Excel, Access, FileMaker, BRAHMS, Specify, etc.



SPEC_ID	KING	PHY	CLASS	ORD	FAM	GEN	SP	Wingspan
1	Animalia	Arthropoda	Insecta	Odonata	Synlestidae	<i>Chlorolestes</i>	<i>apricans</i>	2.5
2	Animalia	Arthropoda	Insecta	Odonata	Synlestidae	<i>Chlorolestes</i>	<i>conspicuus</i>	3.2
3	Animalia	Arthropoda	Insecta	Odonata	Synlestidae	<i>Chlorolestes</i>	<i>draconicus</i>	1.9
4	Animalia	Arthropoda	Insecta	Odonata	Synlestidae	<i>Chlorolestes</i>	<i>elegans</i>	11
5	Animalia	Arthropoda	Insecta	Odonata	Synlestidae	<i>Chlorolestes</i>	<i>fasciatus</i>	14.3
6	Animalia	Arthropoda	Insecta	Odonata	Synlestidae	<i>Chlorolestes</i>	<i>tessellatus</i>	9.5

Mapeo de las columnas de una base de datos particular con los términos del DwC. La imagen es de GBIF.

Libre acceso a datos de biodiversidad compartidos mediante estándares



¿Cuáles datos son importantes?

- Darwin Core tiene más de 150 campos de datos (¡ABCD tiene más de 1200!).
 - En el caso de la última versión del DwC, solo dos son obligatorios: *catalogNumber* y *basisOfRecord*.
- La importancia de cada campo está fuertemente relacionada con el uso que se le vaya a dar a los datos.
- En el caso de los modelos de distribución, tres campos son indispensables:
 - *scientificName*.
 - *decimalLongitude*.
 - *decimalLatitude*.

¿Cuáles datos son importantes?

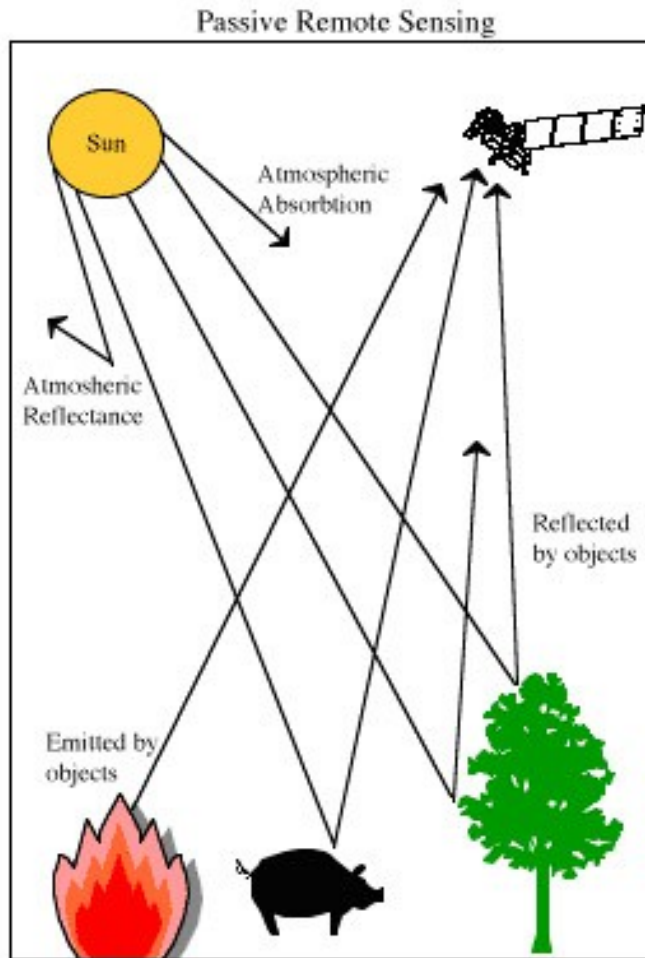
- Datos ambientales
 - Datos interpolados de estaciones meteorológicas:
 - WorldClim (<http://www.worldclim.org/>): 19 variables climáticas (temperatura, precipitación, altitud, ...).
 - Datos de sensores remotos:
 - Reverb (<http://reverb.echo.nasa.gov/reverb/>): datos de instrumentos de la NASA (Landsat, MODIS, LIDAR, ...).
 - Datos de estudios de campo:
 - Ecomapas (<http://www.inbio.ac.cr/ecomapas/introduccion.htm>).

Teledetección pasiva y teledetección activa

Pasiva:

Radiación
EM natural

e.g. rayos
infrarojos,
fotografías
aéreas

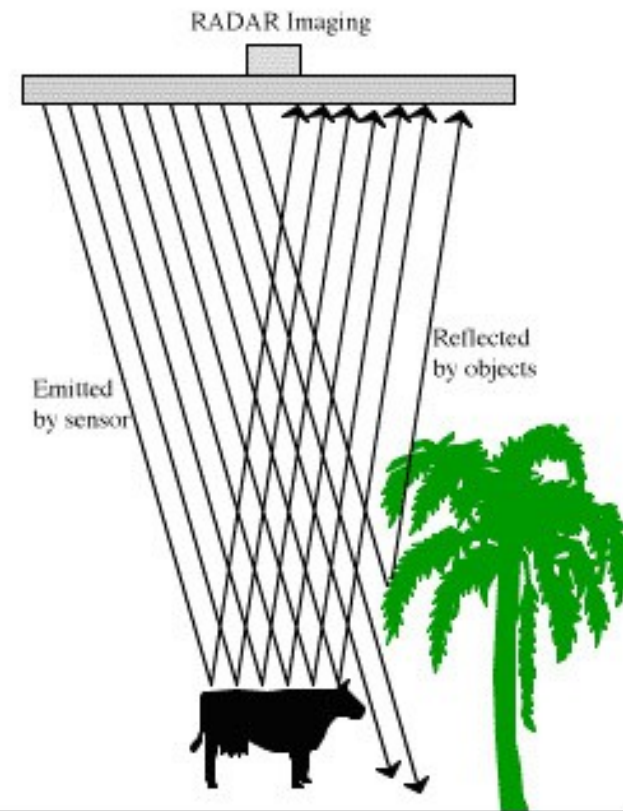


Active Remote Sensing

Activa:

Radiación
EM emitida
por sensores

e.g. RADAR,
LIDAR

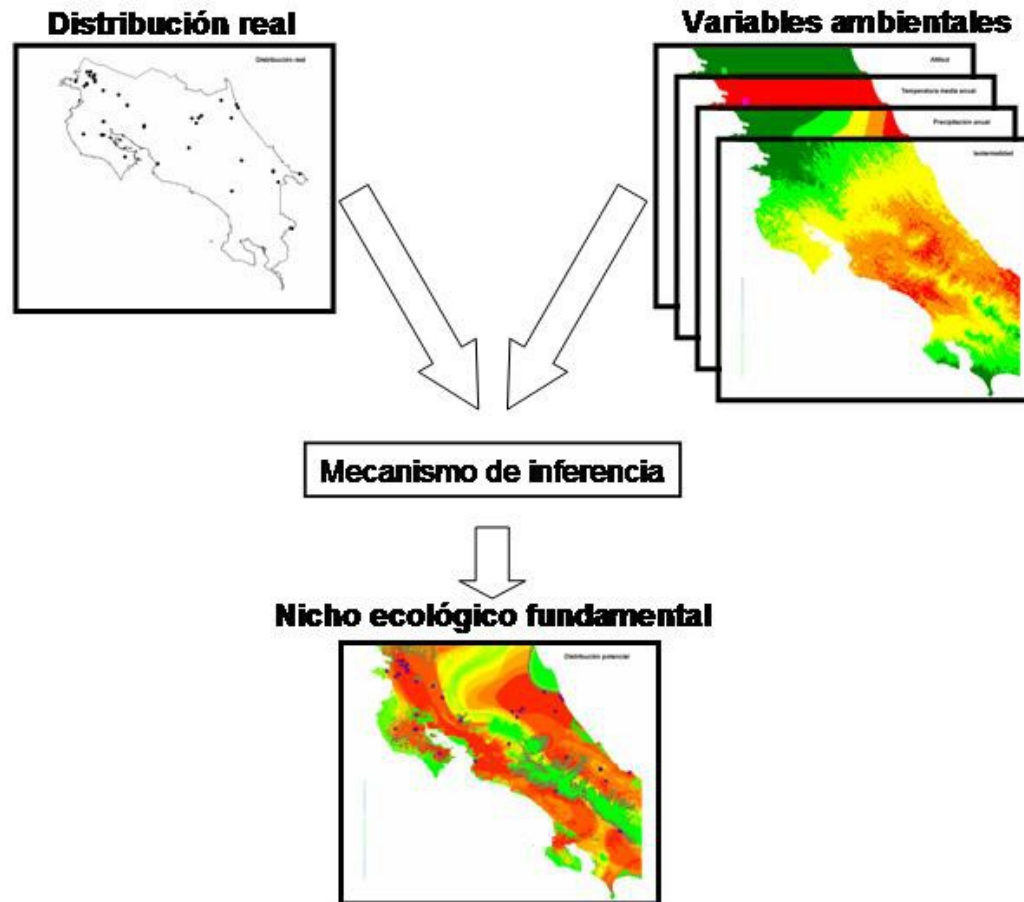


Clasificación de los productos de datos sobre tierra de MODIS

En la actualidad, MODIS ofrece 67 productos de datos para la tierra que pueden clasificarse de la siguiente forma:

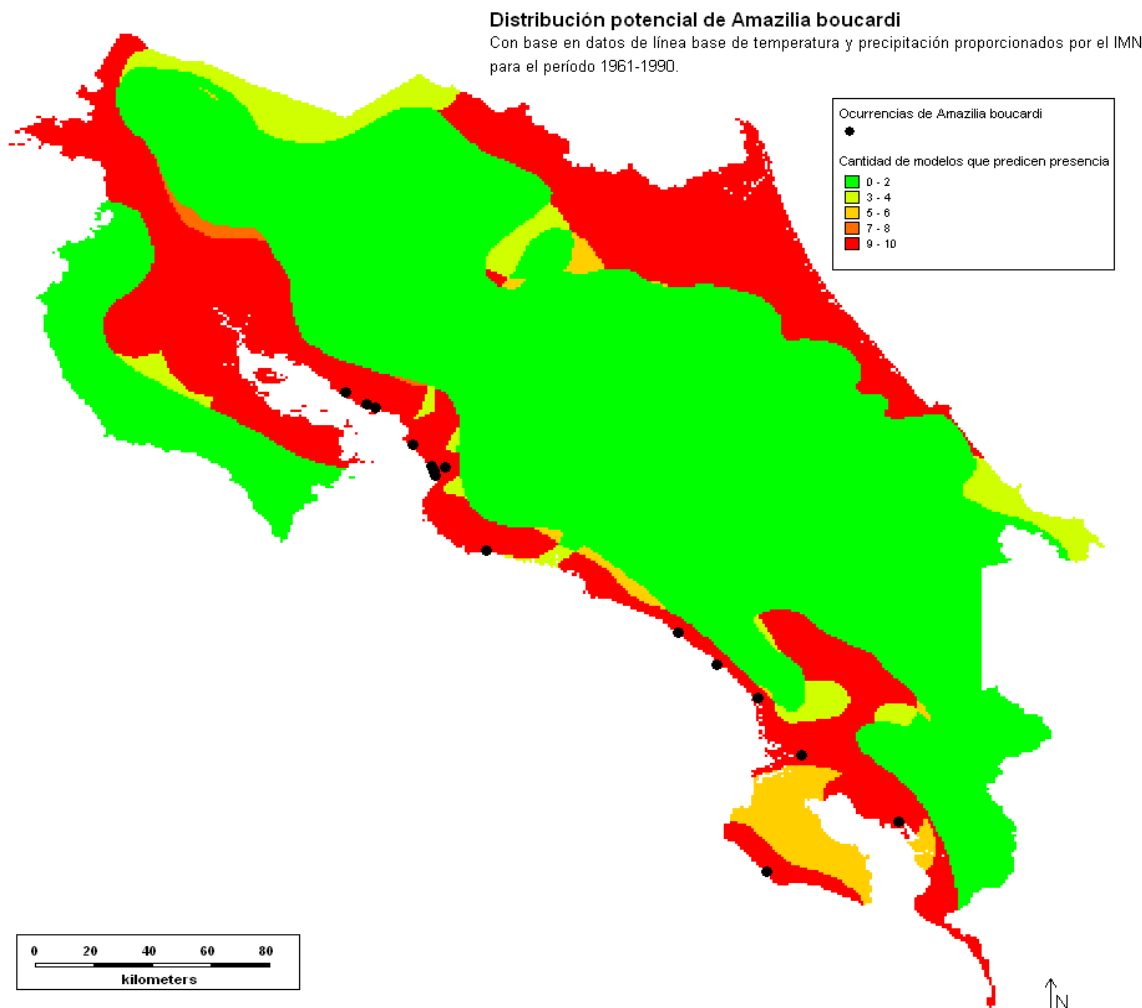
- Variables relacionadas con el presupuesto energético:
 - Reflexión de la superficie de la tierra.
 - Temperatura de la tierra y emisividad.
 - BRDF (*Bidirectional Reflectance Distribution Function*) y albedo (i.e. coeficiente de reflexión).
- Variables relacionadas con los ecosistemas:
 - Índices de vegetación.
 - Índice de superficie foliar y FPAR (*Fraction of Absorbed Photosynthetically Active Radiation*).
 - Productividad Neta Primaria.
- Características de la cobertura de la tierra
 - Anomalías térmicas e incendios.
 - Cobertura de la tierra.
 - Campos continuos de vegetación.

Modelos de Distribución de Especies



Para construir un modelo de distribución se necesitan dos conjuntos de datos: los lugares (con coordenadas) en donde la especie ha sido vista o recolectada y las capas de variables ambientales que influyen en la distribución de la especie (expresados como mapas *raster*). Un mecanismo de inferencia (e.g. entropía máxima) toma ambos conjuntos y construye un modelo del nicho ecológico fundamental de la especie. La imagen fue confeccionada por Manuel Vargas, del Instituto Nacional de Biodiversidad de Costa Rica (INBio).

Impacto del cambio climático



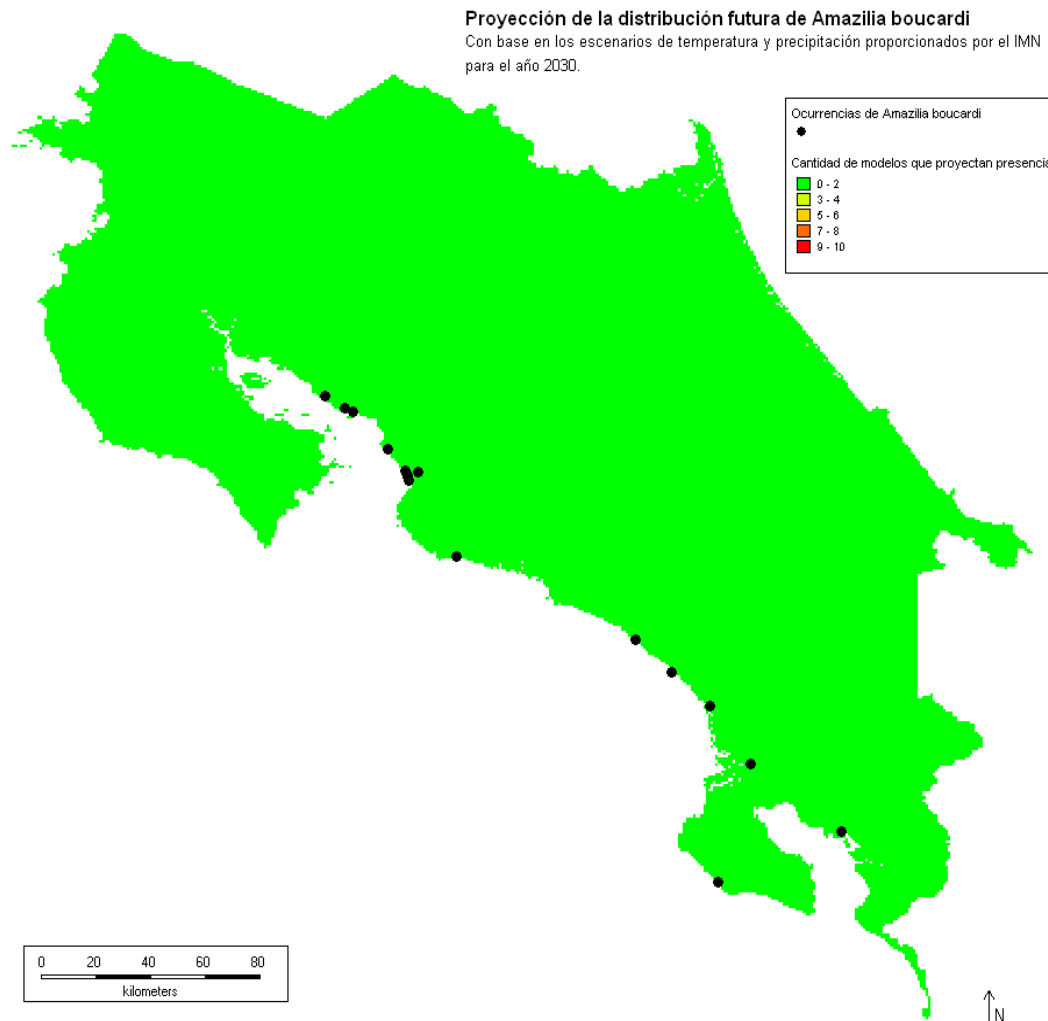
Amazilia boucardii



Distribución potencial actual de *A. Boucardii* basada en datos climáticos de 1961-1990.

El mapa fue confeccionado en el Instituto Nacional de Biodiversidad de Costa Rica (INBio).

Impacto del cambio climático



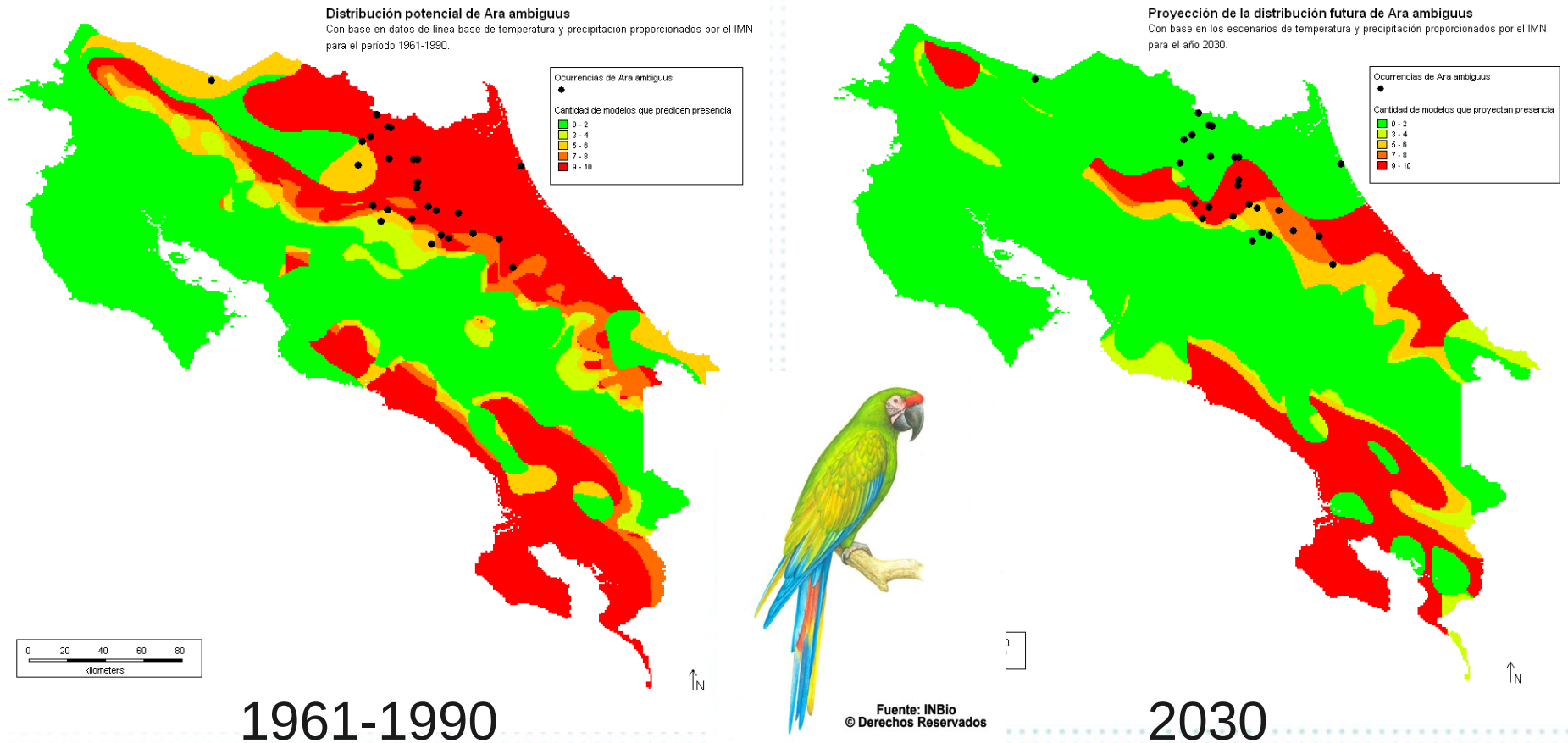
Amazilia boucardii -100%



Distribución potencial
de *A. Boucardii* en 2030
basada en el escenario
climático A2 del IPCC.

El mapa fue confeccionado en el Instituto Nacional de Biodiversidad de Costa Rica (INBio).

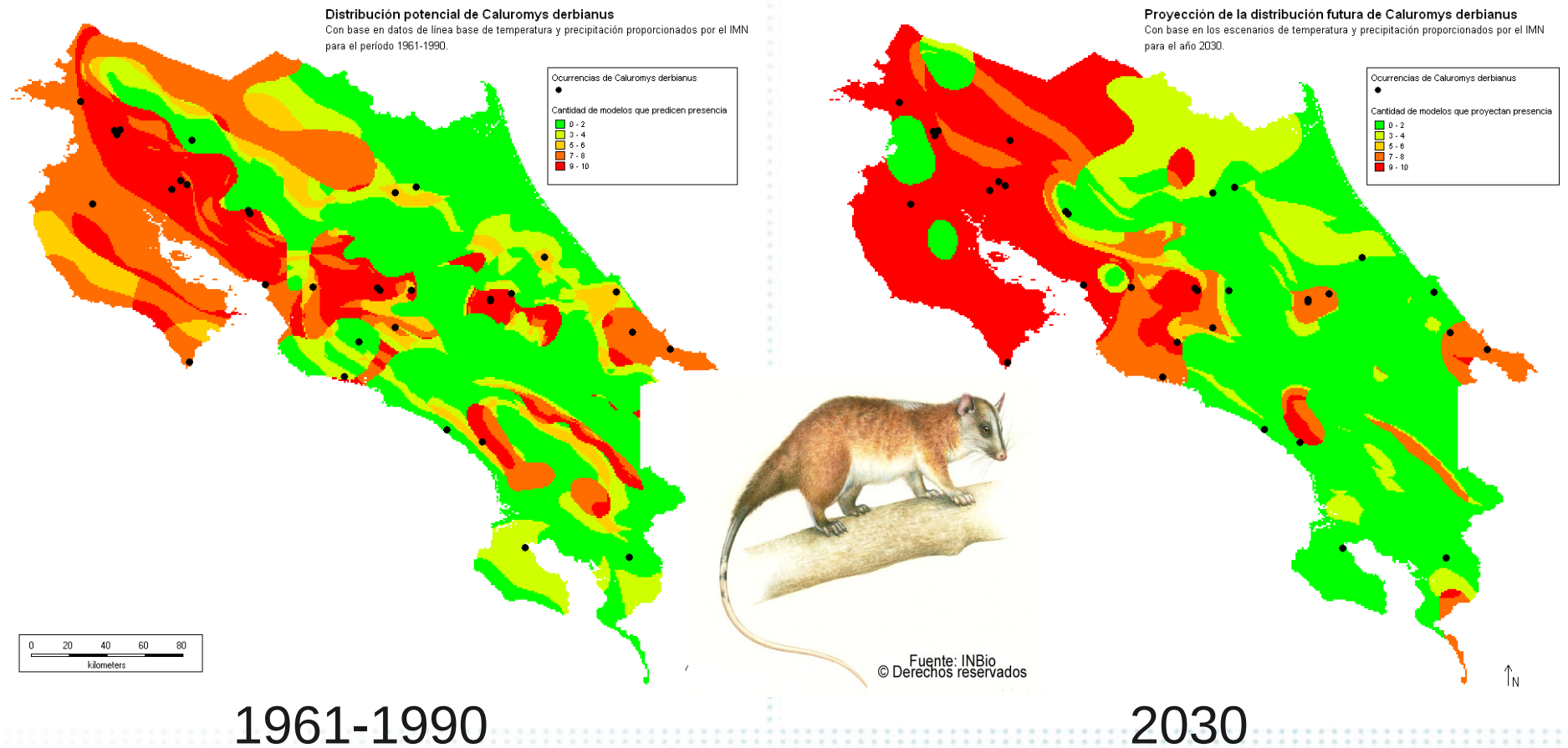
Impacto del cambio climático



Ara ambiguus -41,07%

Los mapas fueron confeccionados en el Instituto Nacional de Biodiversidad de Costa Rica (INBio).

Impacto del cambio climático



Caluromys derbianus +84,36%

Los mapas fueron confeccionados en el Instituto Nacional de Biodiversidad de Costa Rica (INBio).

Recomendaciones para la toma de decisiones

Identificación de corredores biológicos

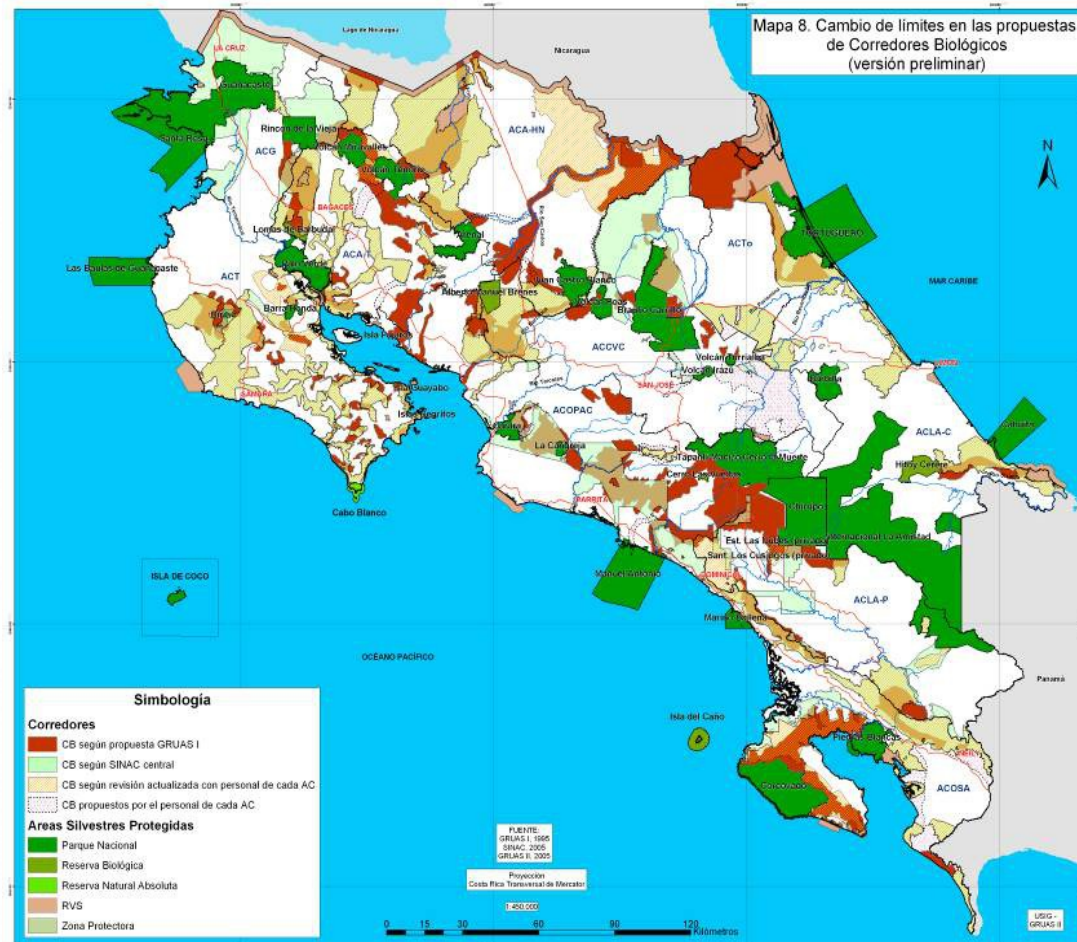
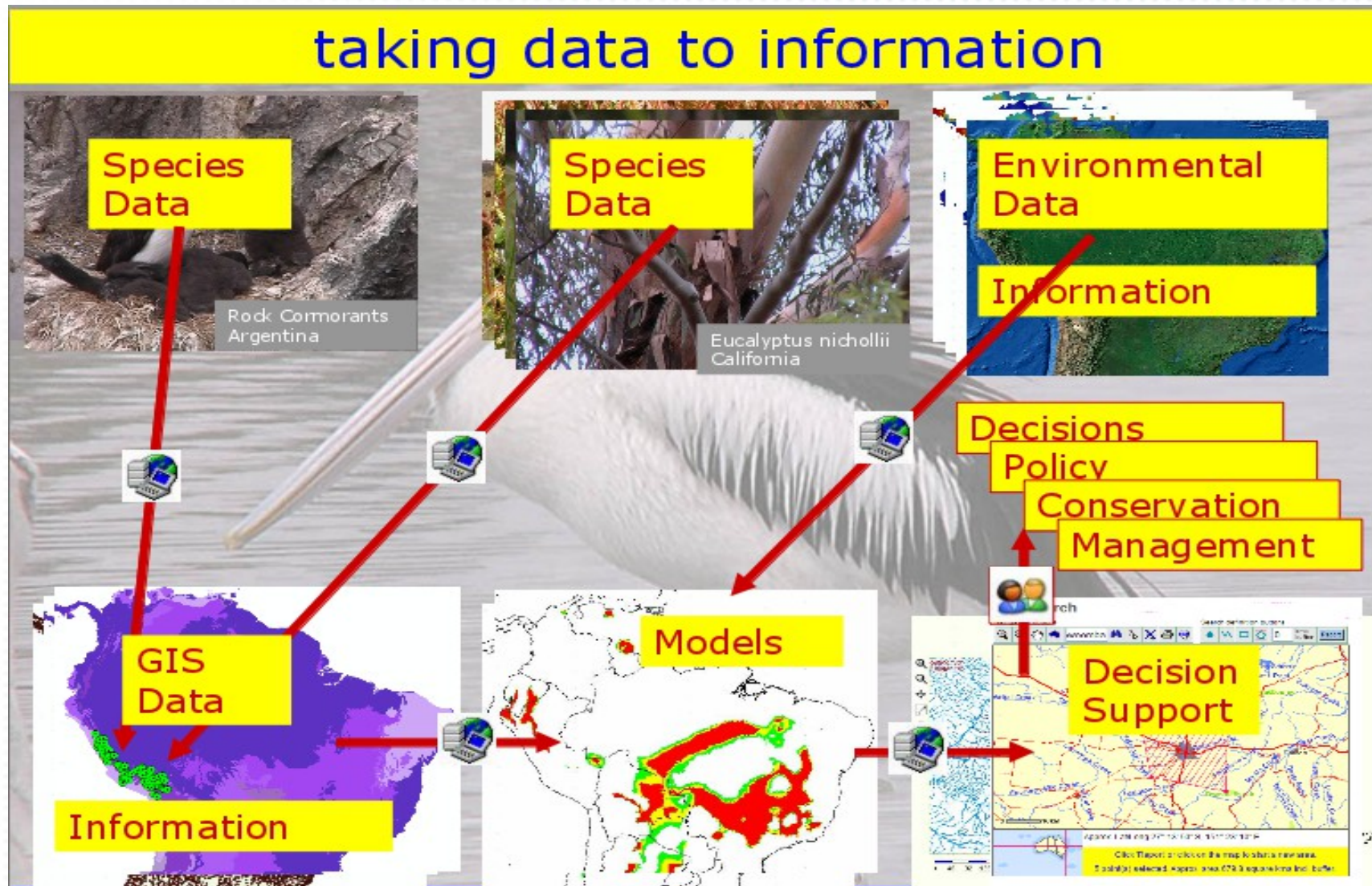


Imagen del proyecto GRUAS II: Propuesta de ordenamiento territorial para la conservación de la biodiversidad en Costa Rica.

La Informática de la Biodiversidad



La imagen es de Arthur D. Chapman.

Bibliografía citada

- Chapman, A. D. (2005). *Principles of Data Quality*, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.
- GBIF (2010). Best practice guide for 'Data Discovery and Publishing Strategy and Action Plans' versión 1.0. Escrito por Chavan, V. S., Sood, R. K., and A. H. Arino. 2010. Copenhagen: *Global Biodiversity Information Facility*, 29 pp. ISBN: 87-92020-12-7. Disponible en <http://www.gbif.org>.
- Naciones Unidas (1992). Convenio sobre la diversidad biológica. (C. O. B. Diversity, Ed.) *Programa de Naciones Unidas para el Medio Ambiente PNUMA*, 1760(6005), 226-253. UICN. Disponible en <http://www.cbd.int/convention/convention.shtml>.
- Peterson, A. T., Stockwell D. R. B., & Kluza, D.A. (2002). Distributional prediction based on ecological niche modeling of primary occurrence data. En Scott, J. M., Heglund, P. J., & Morrison, M. L. [Eds.], *Predicting species occurrences: issues of scale and accuracy*. Island Press, Washington, DC.
- Schalk, P. (1998). Management of marine natural resources through by biodiversity informatics. *Marine Policy*, 22(3), 269-280. Disponible en <http://linkinghub.elsevier.com/retrieve/pii/S0308597X9800013X>.