



Biodiversity information

Francisco Pando



Spanish GBIF Node

CEPDEC Pilot Project in Tanzania
Training on Node Management

The Tanzania Commission for Science and
Technology (COSTECH)




Dar es Salaam

Tanzania

25-29 February 2008



Biodiversity levels, a caveat

- **Genetic variability:** refers to the genetic differences that occur within a particular species that can be passed along to offspring.

- **Species diversity:** refers to the variety of species that occur within a particular area. Collectively, all of the individuals of a particular species in a particular area form a **population**.

- **Community diversity:** refers to the associations of species within an area. These associations, also called biological communities, are the living components of **ecosystems**.

- **Landscape/regional diversity:** refers to the variety of ecosystems and communities that can be found within the landscape.

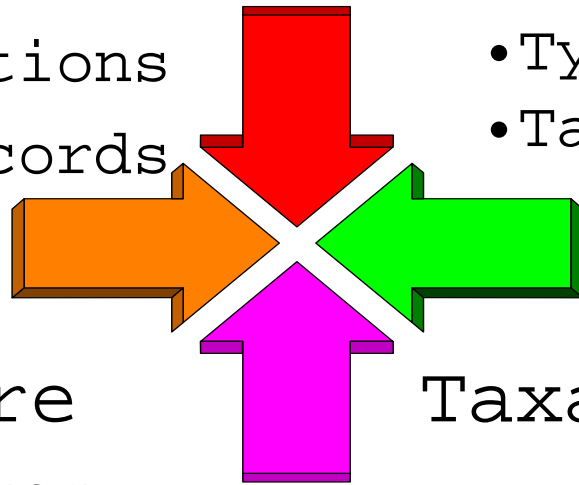
The nature of biodiversity information

Primary data

- Specimens
- Observations
- Lit. records

Names

- Accepted & synonyms
- Type information
- Taxonomic schemas



Literature

- References
- TL2 & BPH
- Key-words

Taxa

- Descriptions, identification keys, conservation, uses, distribution, habitat, etc.

Adapted from:

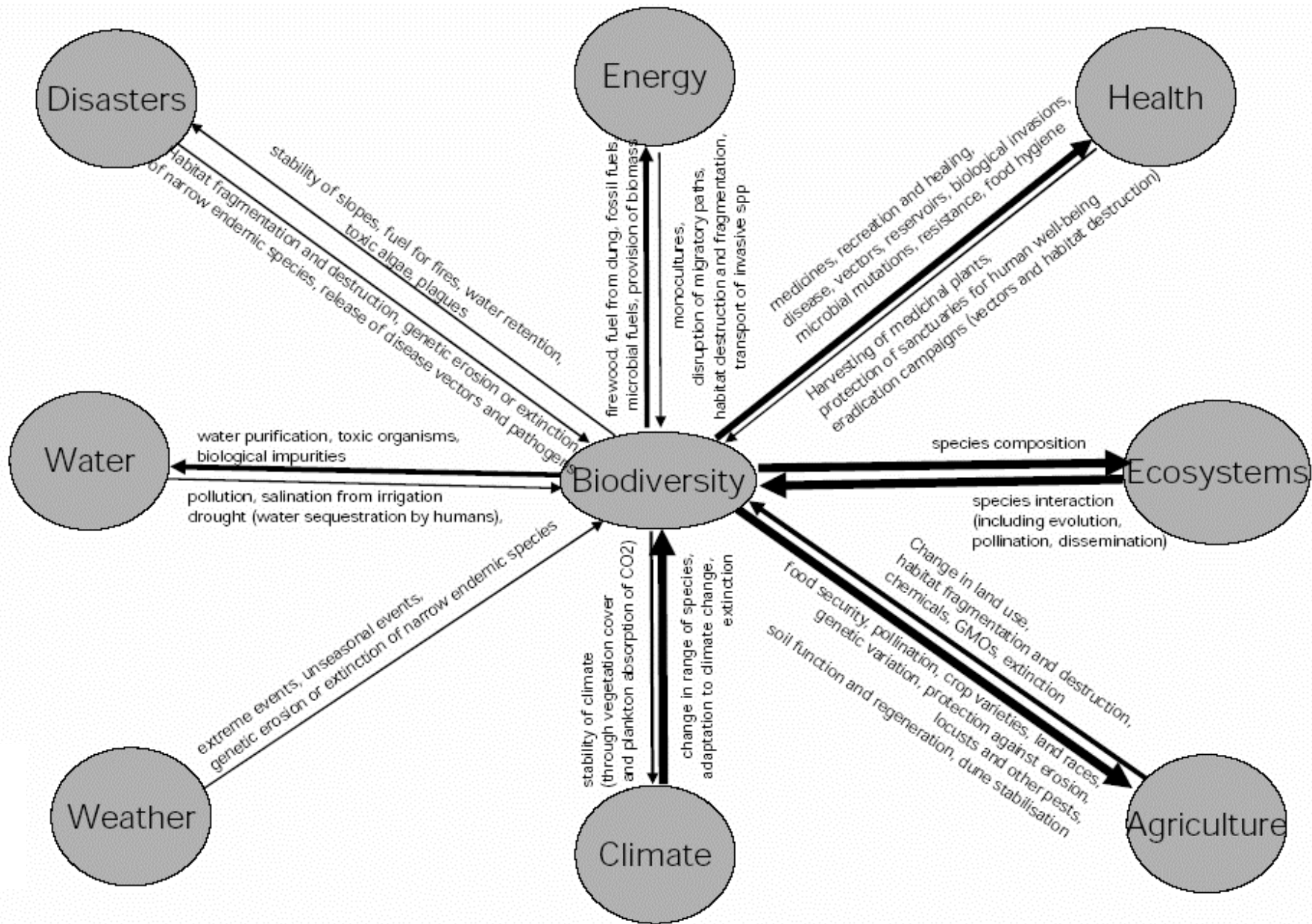
Leenhouts, Regnum Veg. 58. 1968.

Accessibility of biodiversity information

Biodiversity information
was/is not easily
accessible:

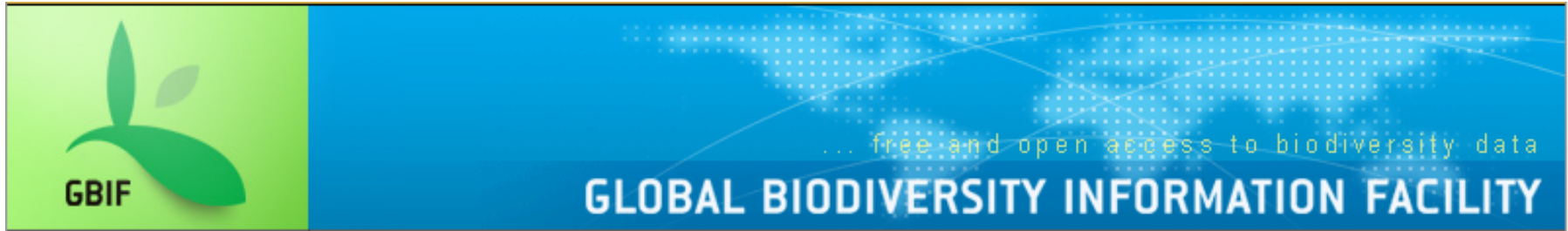
- It is not in digital form
- It is scattered

This is a problem, not just
for scientists, but for
society



Biodiversity linkages. Developed by Martin Sharman, EC

GBIF



Megascience project aimed at

- Making the world's **biodiversity data** freely and universally available via the Internet
- Sharing primary scientific biodiversity data to **benefit science and society**

GBIF's programmes

GBIF has a number of work programmes to help develop a global biodiversity Network:

- Content
 - Digitisation of Natural History Collections (DIGIT)
 - Electronic Catalogue of Names of Known Organisms (ECAT)
 - [Species banks, Biodiversity Digital Library]
- Informatics
 - Data Access and Database Interoperability (DADI)
- Participation
 - Outreach and Capacity Building (OCB)
 - Nodes
 - Training

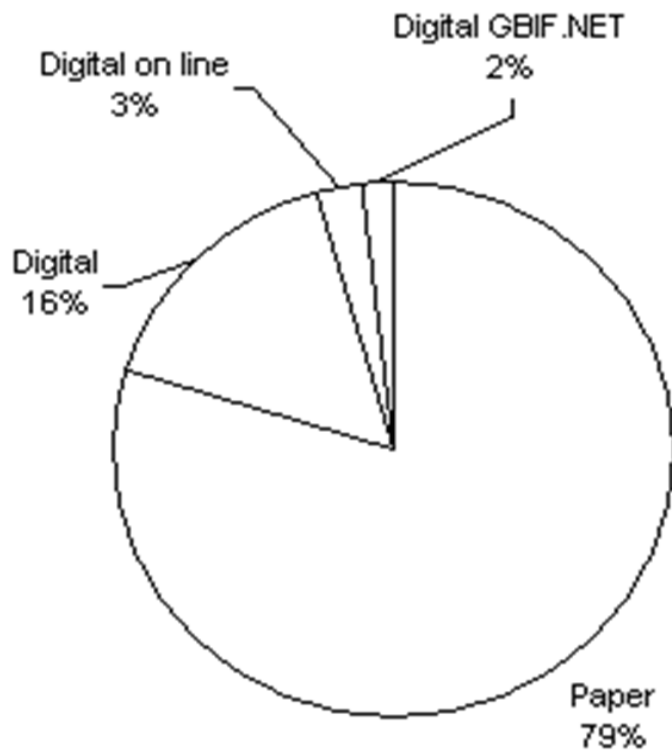
Names Specimens
Literature  SppBank

Accessibility of biodiversity information

- Where is the information?
 - In a myriad of places
 - ⇒ Access hampered
 - ⇒ Limited used (outside the scientific community)
 - ⇒ Substitutes sought
- Picture of the past?

Accessibility of biodiversity information

- Primary data: collections



3-1.5 * 10⁹ specimens
3000-6000 institutions

...and this is just GBIF members!

600.000.000
specimens

Access to information: types of information

Primary data: collections

- taxonomic
- people
- habitat / ecological
- Spatial / geographic
- historical / phenological
- molecular

HORTUS REGIUS MATRITENSIS (MA-Fungi)

Martellia sp.

ESP. CIUDAD REAL: Viso del Marqués, Arroyo de la Poveda, bajo Halimium sp., 23-XI-1993, T. Pérez Jarauta, det. F.D. Calonge

MA-FUNGI 32269



Zelleromyces piennensis Horno-Arroyo, Pérez & Calonge

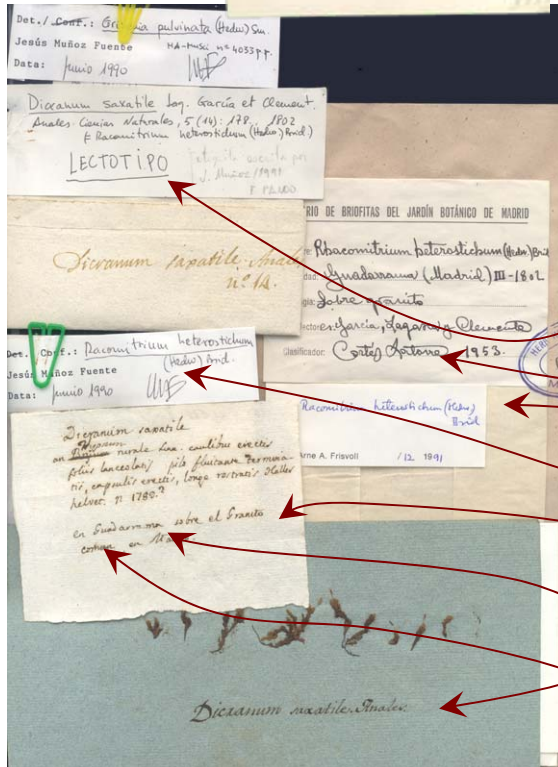
Rev. Jop & Vidal

Dat. 25-5-2000

GenBank Acc. No. AF215649
MA-Fungi 32269

Access to information: not digitalised vs. digitalised

- one (card) index
⇒ n-1 difficult tasks
- one database ⇒ multiple indexing ⇒ multiple uses



Herbario - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media History

Address [/asp/dos.asp?conque=MA-Musci&fnhe=4033&fadic=1](#) Go Google

RJB > Colecciones > Herbario > Criptogamia > Bases de datos > Búsqueda

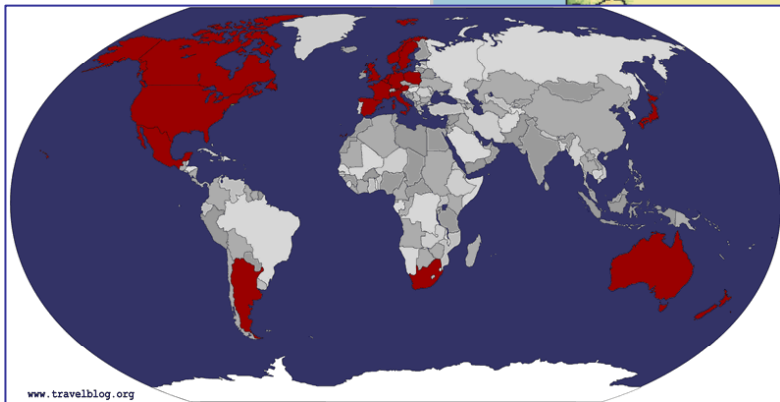
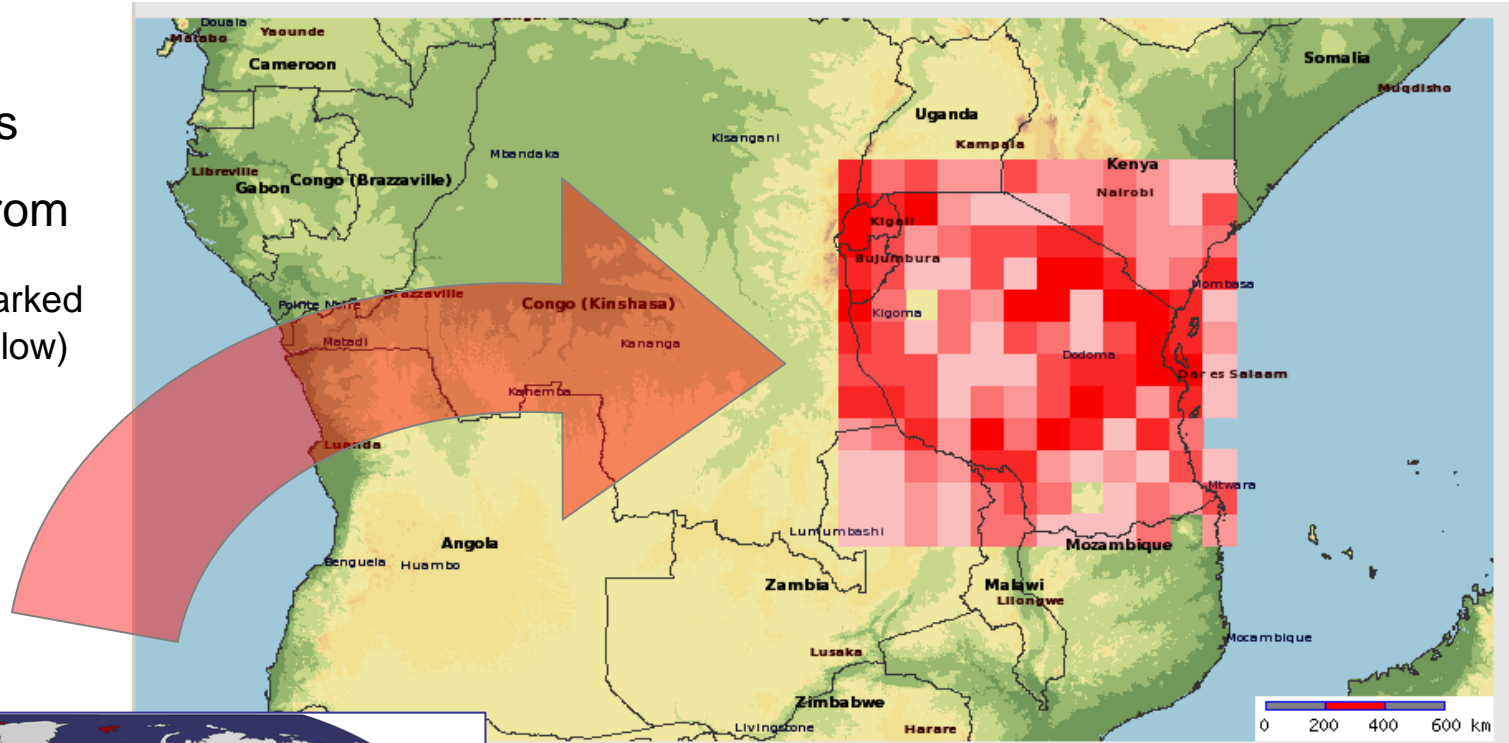
Resultados

Dicranum saxatile	Typus (L)	1
Racomitrium heterostichum (Hedw.) Brid.		2
Det. Cortés Latorre (/1953)		
Racomitrium heterostichum (Hedw.) Brid.		3
Det. J. Muñoz Fuente (05/1990)		
Dicranum saxatile Lagasca, García & Clemente	Typus (L)	4
Det. J. Muñoz (/1991)		
Racomitrium heterostichum (Hedw.) Brid		5
Det. Arne A. Frisvoll (12/1991)		

ESP. M. Guadarrama
sobre granito
García, Lagasca & Clemente
III-1802
MA-Musci 4033
Nombre aceptado en el Herbario : Racomitrium heterostichum (Hedw.) Brid

Access to information: scattered sources

118,809 records
155 data sets from
20 countries (marked
in red in the map below)



Map scale: 1:15749114
Click x,y: 26.325000, -6.850000
Map Extent: 3.825000 -18.100000 48.825000 4.400000

Mode

Browse
 Query

Search radius: (kilometers) or
select the bounding box of a country:

Range of records per colour		
	1 -	26
	28 -	75
	87 -	208
	223 -	509
	521 -	1385
	1622 -	5051

Access to scattered sources: a workable solution



Unified access, distributed information:

The GBIF Network
data.gbif.org

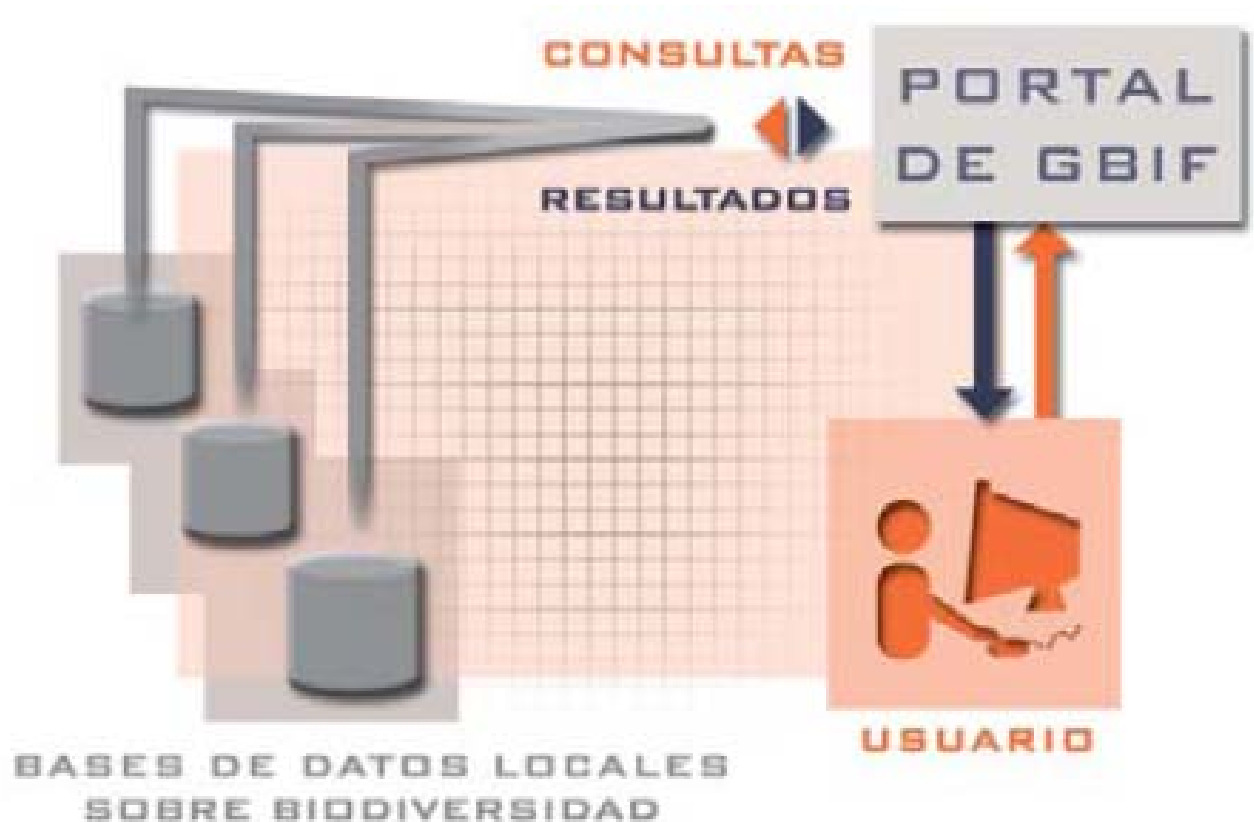
Precursors:

REMIB - CONABIO
(México)

http://www.conabio.gob.mx/remib/doctos/remib_esp.html

TSA -
Univ. Kansas

<http://speciesanalyst.net/index.html>



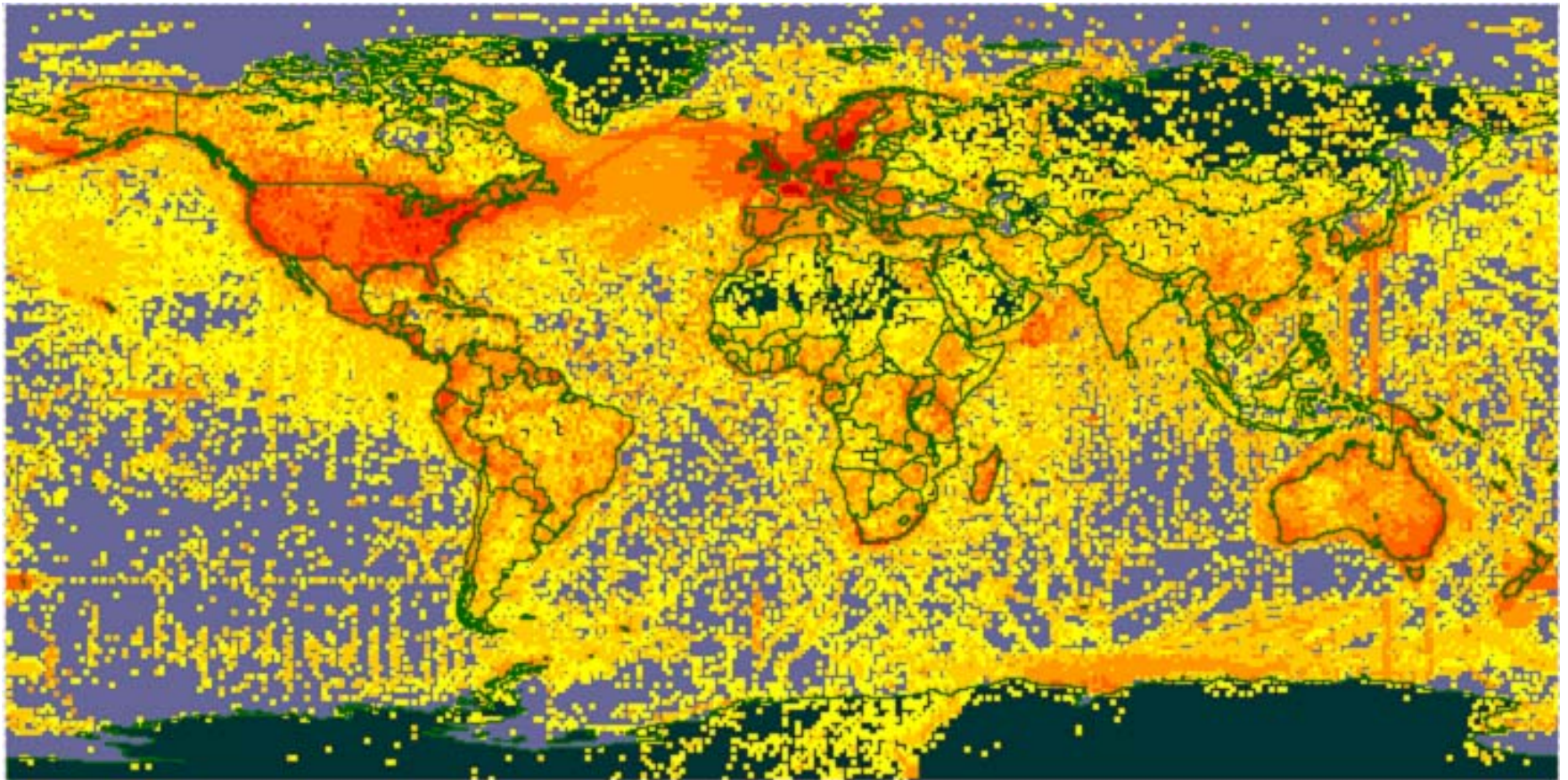
Unified access, distributed information: The GBIF network

<http://data.gbif.org>

235 providers (servers)

1023 databases

142,048,114 records



Unified access, distributed information: common profile + standards

- Common profile:
Each particular database structure is translated into a "profile", a table with common field list that can be accessed in a uniform manner

Standards:

www.tdwg.org

Attr#	Name	Description
1	ScientificName	Genus [+ "" + species [+ "" + subspecies]]. This
2	Kingdom	The kingdom to which the organism belongs.
3	Phylum	The phylum (or division) to which the organism
4	Class	The class name of the organism.
5	Order	The order name of the organism.
6	Family	The family name of the organism.
7	Genus	The genus name of the organism.
8	Species	The species name of the organism.
9	Subspecies	The subspecies name of the organism.
10	InstitutionCode	A unique identifier for you institution.
11	CollectionCode	Unique identifier for the collection within the ins
12	CatalogNumber	Unique identifier for the specimen record within
13	Collector	The name of the collector or collectors that were observation) from the field.
14	Year	The year (four digit) in which the specimen was
15	Month	The month of the year (1..12) in which the speci
16	Day	The day of month that the specimen was collec
17	Country	The country or major political unit (ocean) from
18	State/Province	The state, province or region (i.e. next political r

Access to information...

Data integration & interoperability: Standards

<http://www.tdwg.org/standards.html>



Biodiversity Information Standards
TDWG

username LOGIN

You are here: [Home](#) > Standards

TDWG Current (2005) Standards (*)

Title	Activity	Category	Status
	(Task or Interest Group)		
Access to Biological Collection Data - version 2.06	Access to Biological Collections Data	Technical Specification	Current (2005)
Structured Descriptive Data	Structured Descriptive Data Interest Group	Technical Specification	Current (2005)
Taxonomic Concept Transfer Schema	Taxonomic Names and Concepts	Technical Specification	Current (2005)

Home
About TDWG
Standards
| [Status and Categories](#)
Activities
Membership
2008 Conference

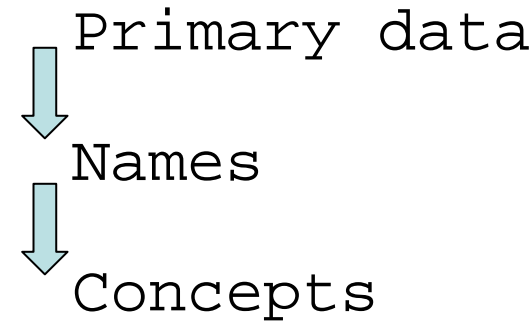
Biodiversity Information Projects of the World

Biodiversity Information Networks Database

<http://www.tdwg.org/standards.html>

Title	Activity (Task or Interest Group)	Category	Status	Date Published	
HISPID3 - Herbarium Information Standards and Protocols for Interchange of Data	Observation and Specimen Records	Technical Specification	Prior	01-Oct-1996	Download
Economic Botany Data Collection Standard	Economic Botany Interest Group	Best Current Practice	Prior	01-Oct-1995	Not available for download
Plant Occurrence and Status Scheme		Data Standard	Prior	01-Oct-1995	Download
Plant Names in Botanical Databases		Best Current Practice	Prior	01-Oct-1995	Download
Authors of Plant Names		Data Standard	Prior	01-Oct-1992	Not available for download
World Geographical Scheme for Recording Plant Distributions		Data Standard	Prior	01-Oct-1992	Download
XDF - A Language for the Definition and Exchange of Biological Data Sets		Technical Specification	Prior	01-Oct-1991	Not available for download
Botanico-periodicum-huntianum/supplementum		Data Standard	Prior	01-Oct-1991	Not available for download
Index Herbariorum. Part I: The Herbaria of the World		Data Standard	Prior	01-Oct-1990	Not available for download
International Transfer Format for Botanic Garden Plant Records		Technical Specification	Prior	01-Oct-1987	Download
Floristic Regions of the World		Data Standard	Prior	01-Oct-1986	Not available for download
Users Guide to the DELTA System		Technical Specification	Prior	01-Oct-1986	Download
Taxonomic Literature, ed. 2 and its Supplements		Data Standard	Prior	01-Oct-1976	Not available for download
Botanico-periodicum-huntianum		Data Standard	Prior	01-Oct-1970	Not available for download

Data integration & interoperability: taxonomic concepts



- ✓ Access to information: one collection /source
- ✓ Access to information: multiple collections
- ✗ Biodiversity information users must confront an impediment: one species may be addressed under different names; a name may refer to different species or concepts of a species.

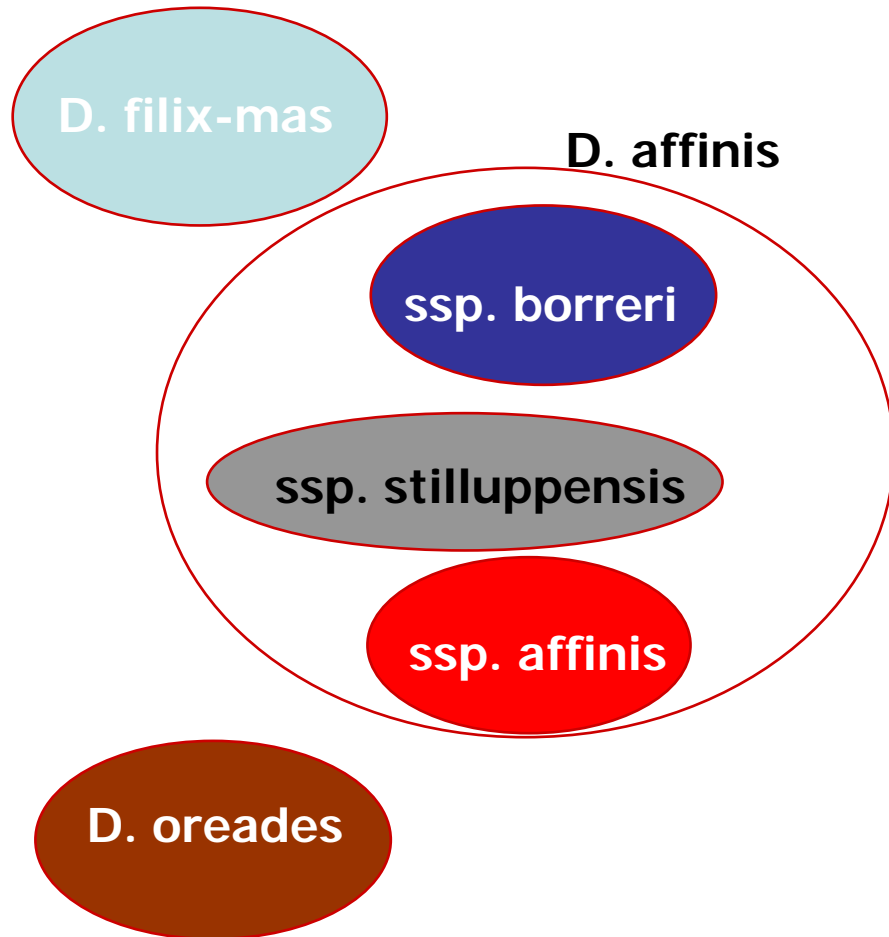
Information, as it is currently presented only make sense to the specialist

An example: the male fern



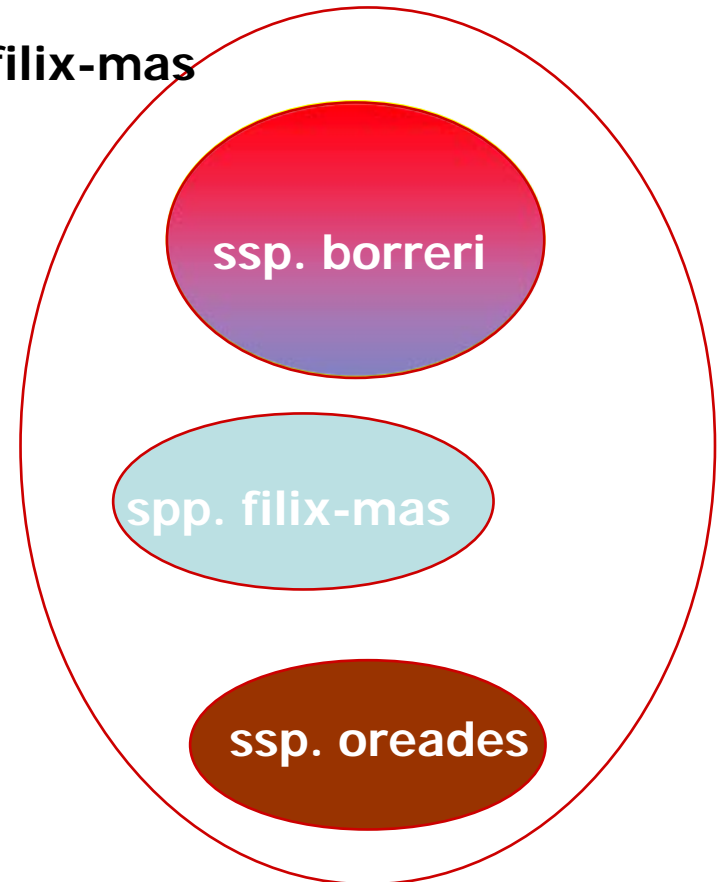
Example: names & concepts

Fl. iberica

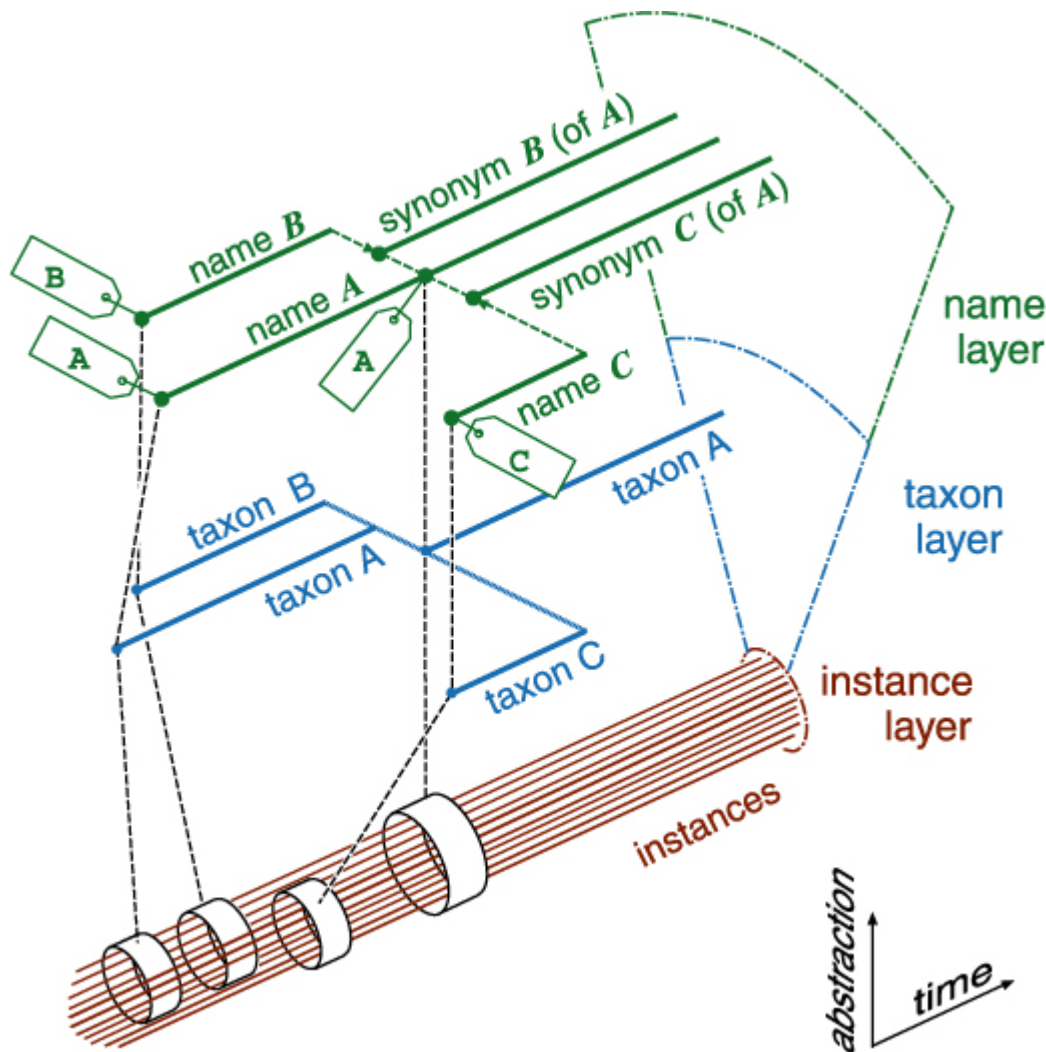


Fl. Paisos Catalans

D. filix-mas



Names & concepts: abstraction & generalization



Big challenge for real integration of biodiversity information.

This level of integration is needed to unlock the information and make it available to society

By Nozomi Ytow & al.

Access to information & integration. How big is the problem?

- Koperski & al. 2000. *Referenzliste der Moose Deutschlands*: 45% of treated taxa are unstable
- Current alternative classifications
- Implicit information :
Identifications in collections do not (usually) indicate the taxonomic treatment followed



Accessibility of biodiversity information: "names"

Levels of information:

- Name indexes: provide list of available scientific names
- Taxonomic views of names: provide names arranged according to a specific taxonomic treatment. These can be regional or global (then called GSDs, "global species datasets" in Species 2000 terminology)
- Trans-taxonomic mapping tools, services or databases.

Accessibility of biodiversity information: Name indexes

Animals	<i>Index to Organism Names</i> ION
Mammals	<i>Mammal Species of the World</i> MSW
Birds	AviBase
Fishes	FishBase
Amphibians	<i>Amphibian Species of the World</i> ASW
Vascular plants	<i>International Plant Names Index</i> IPNI , VAST
Mosses	Most ; also in <i>Index to Organism Names</i> ION
Fungi & lichens	CABI also in <i>Index to Organism Names</i> ION
Algae	AlgaeBase , Index Nominum Algarum also in ION
Bacteria	<i>List of Bacterias with Standing in Nomenclature</i> LBSN
Viruses	<i>The Universal Virus Database of the International Committee on Taxonomy of Viruses</i> ICTVdB

Accessibility of biodiversity information: Names, taxonomic views

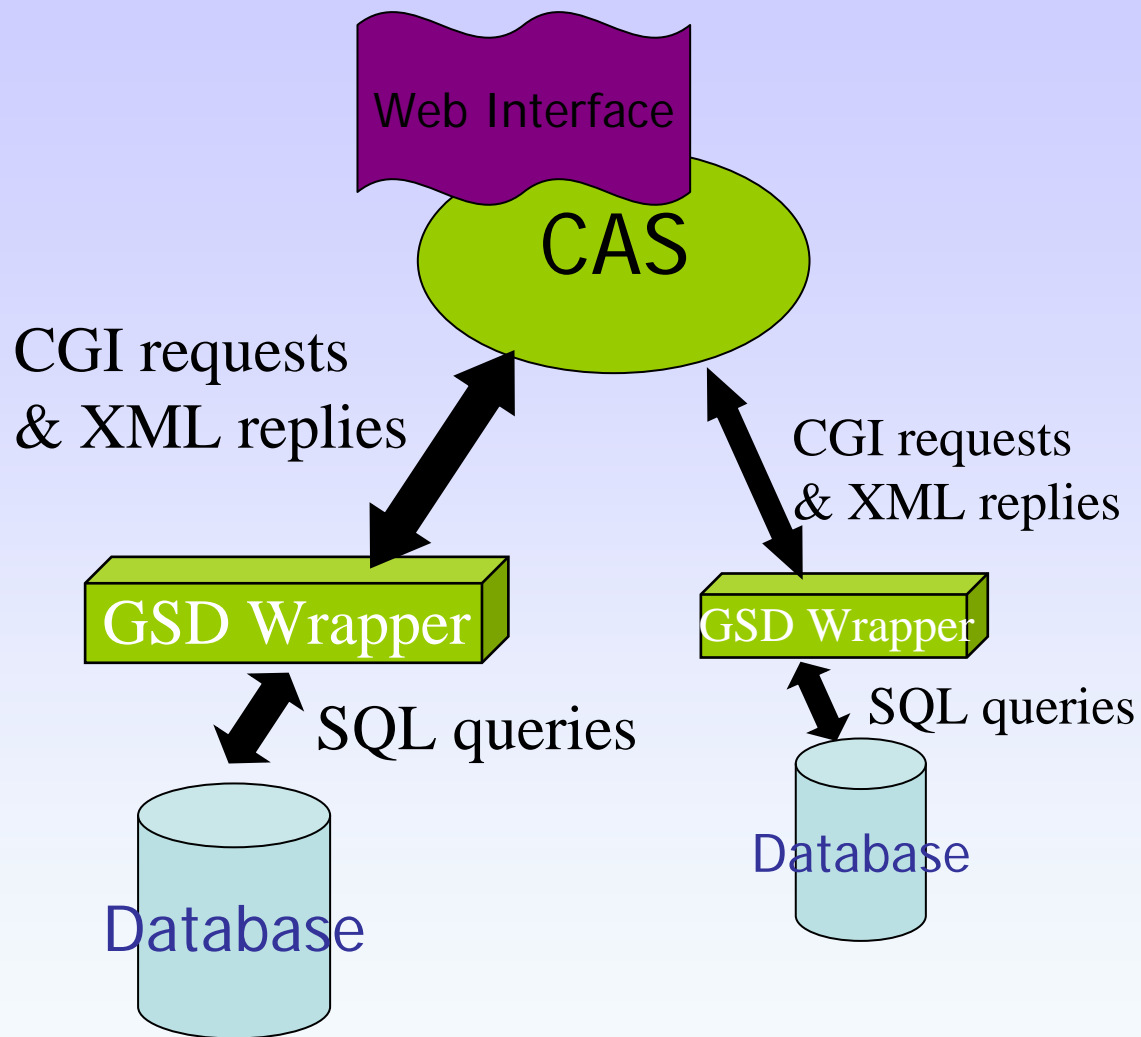
Taxonomic views of names provide names arranged according to a specific taxonomic treatment:

- Species 2000: <http://www.sp2000.org>
- ITIS: <http://www.itis.usda.gov>
- UBIO: <http://www.ubio.org>

Many others at regional level, e.g.:

- Index Synonymique de la Flore de France:
<http://www.dijon.inra.fr/flore-france/>
- Anthos (flora of Spain):
<http://www.programanthos.org/>

Species 2000



- Users Access SPICE via Web Interface
- CAS fetches data from databases via wrappers
- CDM: common profile + operations

<http://sp2000europa.org/meetings/2004/eurocat2/presentations/Qinglai/Linking%20Up%20with%20SPICE.ppt>

<http://spice.sp2000europa.org/SPICE/> (User = password = guest)

Accessibility of biodiversity information: Trans-taxonomic mapping tools

Concepts still under development:

TDWG: "Taxonomic Concept Transfer Schema" subgroup

[http://www.soc.napier.ac.uk/tdwg/index.php?page
name=HomePage](http://www.soc.napier.ac.uk/tdwg/index.php?page=name=HomePage)

Some experimental works:

- The MoreTaxa project

<http://www.bgbm.org/BioDivInf/Projects/MoreTax/>

- ILDIS Legumes of Northern Eurasia CD-Rom

<http://www.ildis.org/cd/>

Accessibility of biodiversity information

Literature

- As a way to enable access the scientific knowledge.
- "digital libraries" are becoming a common thing:
 - One of the first and biggest:
Bibliothèque numérique de la Bibliothèque nationale de France: <http://gallica.bnf.fr/>
 - Developments in the area:
project to join current initiatives by big institutions: The "Biodiversity Heritage Library" project (www.bhl.org)
- The challenge is open access and, again integration

Accessibility of biodiversity information

Literature

– Example:

The screenshot shows a web browser window with the address bar displaying <http://www.rbghkew.org.uk/epic/>. The browser's taskbar includes icons for 'Customize Links', 'Free Hotmail', 'Windows Marketplace', 'Windows Media', and 'Windows'. The website header features the 'Royal Botanic Gardens, Kew' logo and navigation links for 'SCIENCE & HORTICULTURE', 'COLLECTIONS', 'CONSERVATION & WILDLIFE', 'EDUCATION', and 'DATA & PUBLICATIONS'. The main heading is 'ePIC electronic plant information centre', with sub-links for 'ePIC Home', 'News', and 'Search'. A green sidebar on the left contains a date announcement: '6th December 2004. Monocot Checklist added! See the [News](#) page for more information.' Below this is a list of navigation links: 'ePIC Home', 'Terms of Use', 'News', 'Search', 'F.A.Q.', 'Report an Error', 'About ePIC', 'Contact Us', 'Cite Us', 'Link to us', 'Data Sources', and 'Partners'. The main content area has a purple heading 'Welcome to the electronic Plant Information Centre (ePIC)' followed by a paragraph explaining the project's goal to digitize Kew's information. It includes a link 'here' and a 'Search ePIC' button at the bottom right.

6th December 2004.
Monocot Checklist added! See the [News](#) page for more information.

ePIC Home
Terms of Use
News
Search
F.A.Q.
Report an Error
About ePIC
Contact Us
Cite Us
Link to us
Data Sources
Partners

Welcome to the electronic Plant Information Centre (ePIC)

ePIC is a major project to bring together all of Kew's digitised information about plants and make it easier to search. You can use it to pinpoint information of interest in our varied collections, bibliographies, nomenclators and checklists, publications and taxonomic works, as well as links to information resources provided by external organisations. Where further information from Kew is available online, we will direct you to it.

You can learn more about ePIC [here](#).

Use the menu bar at the bottom of the page to find out more, or start searching our data right away...

[Search ePIC](#)

Accessibility of biodiversity information

Species level information

(Species bank):

information on taxa

specifically:

- not associate to specimens
- independent of taxonomic schemas

Accessibility of biodiversity information

Species level information
(Species bank):

many initiatives

- Survey commissioned by GBIF
gathered 298 resources. Report &
database available at:

http://circa.gbif.net/Public/irc/gbif/pr/library?l=/speciesbank_workshop/database_speciesbanks&vm=detailed&sb=Title

Accessibility of biodiversity information

Species level information (Species
bank) :

examples :

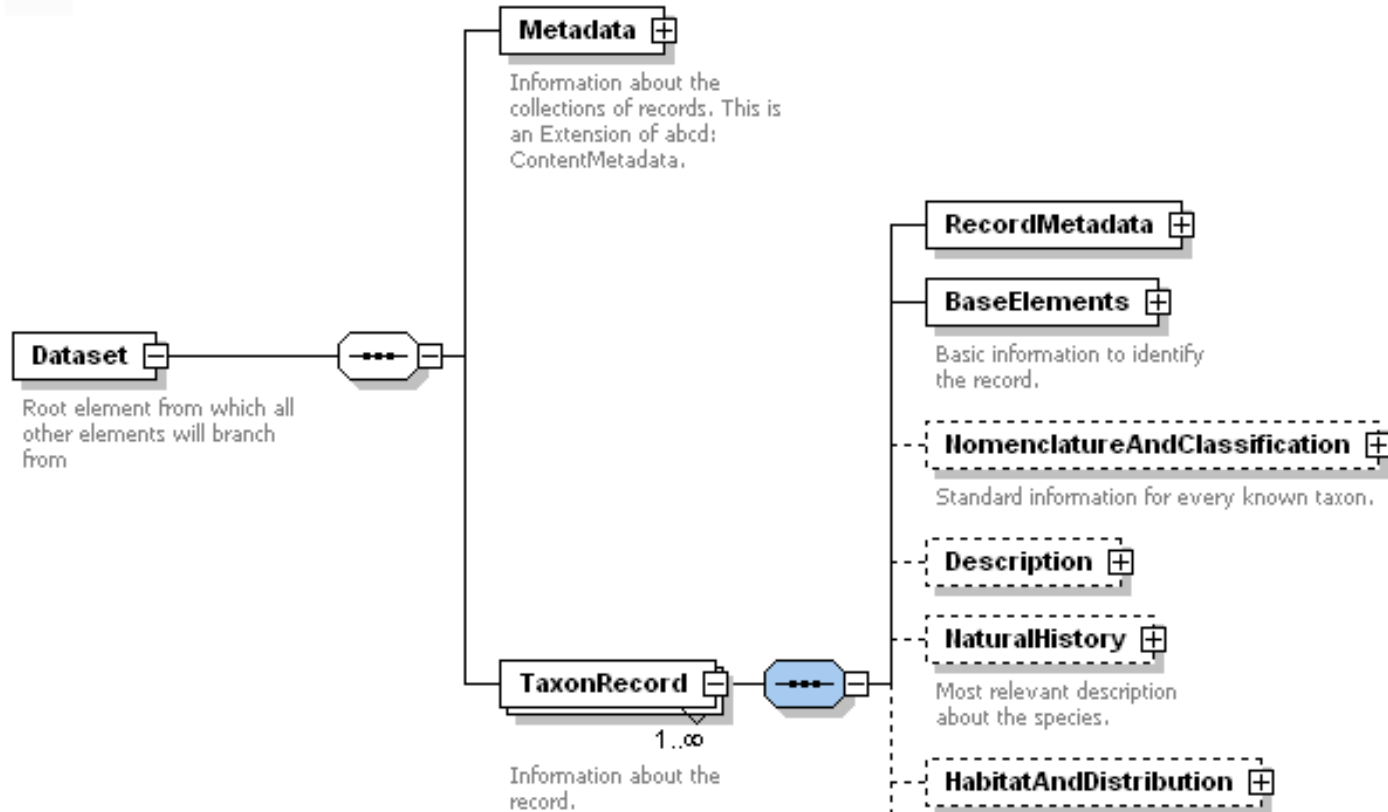
INBIO's UBIs <http://darnis.inbio.ac.cr/ubis/>

Fishbase <http://www.fishbase.org>

efloras <http://www.efloras.org>

Encyclopedia of Life <http://www.eol.org>

Species level information



Plinian Core

- www.pliniancore.org
- www.gbif.es/pliniancore

Accessibility of biodiversity information

The challenges

- To cross the digital divide

From paper to digital form

Moving our science into e-taxonomy

(Maybe from digital to paper too!)

- To integrate resources

within and between areas (names-specimens-species-literature)

- To cross the science - society divide

make science knowledge accessible to society
and used to make sound political decisions

Intellectual Property Rights (IPR)

GBIF IPR principles:

- GBIF is an open-access facility, but providers may want to block access to certain sensitive data (e.g., location of last populations of an endangered species)
- To the greatest extent possible, GBIF-affiliated data should be, and remain in, the public domain
- IPR ownership remains with the data providers
- GBIF will not assert any IPR in the data of affiliated data providers
- GBIF will seek to ensure that data sources are acknowledged by subsequent users

<http://www.gbif.org/DataProviders/Agreements/GBIFdataIPRprinciples.html>

Intellectual Property Rights

- Some work done:

- REMIB (CONABIO)

- http://www.conabio.gob.mx/remib/cgi-bin/clave_remib.cgi?lengua=es-MX

- GBIF

- Data use agreement

- <http://www.gbif.org/DataProviders/Agreements/DUA>

- Data sharing agreement

- <http://www.gbif.org/DataProviders/Agreements/DSA>

- GBIF expert meeting on IPR (2004) & GBIF's Pro-Bono legal Advisory Group (Proleg) 2007

The image contains two overlapping screenshots of GBIF web pages. The top screenshot shows a page titled 'ACCESO A LA INFORMACIÓN' with a navigation breadcrumb 'Conabio » Remib » Acceso a la información'. It features a globe icon and text stating 'UNA RED CON 6,475,880 REGISTROS'. Below this, it lists 'LINEAMIENTOS Y DECLARACIONES' and begins to outline consultation rules: 'Las consultas se harán bajo los siguientes lineamientos: a. Consultas: el usuario que consulte información de la RED...'. The bottom screenshot shows a page titled 'Data Use Agreement' with a 'Background' section. It discusses the goals of making biodiversity data openly available through the GBIF network, mentioning the Memorandum of Understanding on GBIF (MoU) and the importance of a framework for data sharing.

Intellectual Property Rights in GBIF

3. **Intellectual Property Rights to Biodiversity Data**

GBIF should encourage the free dissemination of biodiversity data and, in particular:

- (a) should not assert any Intellectual Property Rights in the data in databases that are developed by other organisations and that subsequently become affiliated to GBIF;
- (b) should seek, to the greatest extent possible, to place in the public domain any data commissioned, created or developed by GBIF; and
- (c) should respect conditions set by data providers that affiliate their databases to GBIF.

When establishing affiliations or linkages with other databases, GBIF should seek to ensure that the data so made available will, in effect, be in the public domain, and will not be subject to limitations on its further non-commercial use and dissemination, apart from due attribution.

4. **Attribution**

GBIF should seek to ensure that the source of data is acknowledged and should request that such attribution be maintained in any subsequent use of the data.

5. **Access to Specific Data**

Nothing in this MOU should be read to restrict the rights of owners of databases affiliated with GBIF to block access to any data.



THE PRINCIPLES:

THE PRINCIPLES

- Introduction
- Benefits
- Endorsement
- Principles

Supporters of the Conservation Commons agree to the following principles:

Principle 1 Open Access

The Conservation Commons promotes free and open access to data, information and knowledge for conservation purposes.

Principle 2 Mutual Benefit

The Conservation Commons welcomes and encourages participants to both use these resources and to contribute data, information and knowledge.

Principle 3 Rights and Responsibilities

Contributors to the Conservation Commons have full right to attribution for any uses of their data, information, or knowledge, and the right to ensure that the original integrity of their contribution to the Commons is preserved.

Users of the Conservation Commons are expected to comply, in good faith, with terms of uses specified by contributors and in accordance with these Principles.

Endorse the Principles

<http://www.conservationcommons.org>

top ↖







Intellectual Property Rights

Creative commons: Licensing Options

You're probably familiar with the phrase, "All rights reserved," and the little (c) that goes along with it. Creative Commons wants to help copyright holders send a different message: "Some rights reserved."

For example, if you don't mind people copying and distributing your online image so long as they give you credit, we'll have a license that helps you say so. If you want people to copy your band's MP3 but don't want them to profit off it without your permission, use one of our licenses to express that preference. Our licensing tools will even help you mix and match such preferences from a menu of options:

-  **Attribution.** Permit others to copy, distribute, display, and perform the work and derivative works based upon it only if they give you credit.
-  **Noncommercial.** Permit others to copy, distribute, display, and perform the work and derivative works based upon it only for noncommercial purposes.
-  **No Derivative Works.** Permit others to copy, distribute, display and perform only verbatim copies of the work, not derivative works based upon it.
-  **Copyleft.** Permit others to distribute derivative works only under a license identical to the license that governs your work.

<http://www.creativecommons.org/>

Data quality

- There is not such a thing as good data and bad data
- Instead: the question is if the data is fit for use
- In other words, the quality of the data depends on the use you are going to put it for

Data quality

“fit for use”

Data is of high quality if it is fit for its intended uses (by customers) in operations, decision making and planning (after Juran).

Free of defects:

- accessible
- accurate
- timely
- complete
- consistent with other sources

***Data
that's fit
for use***

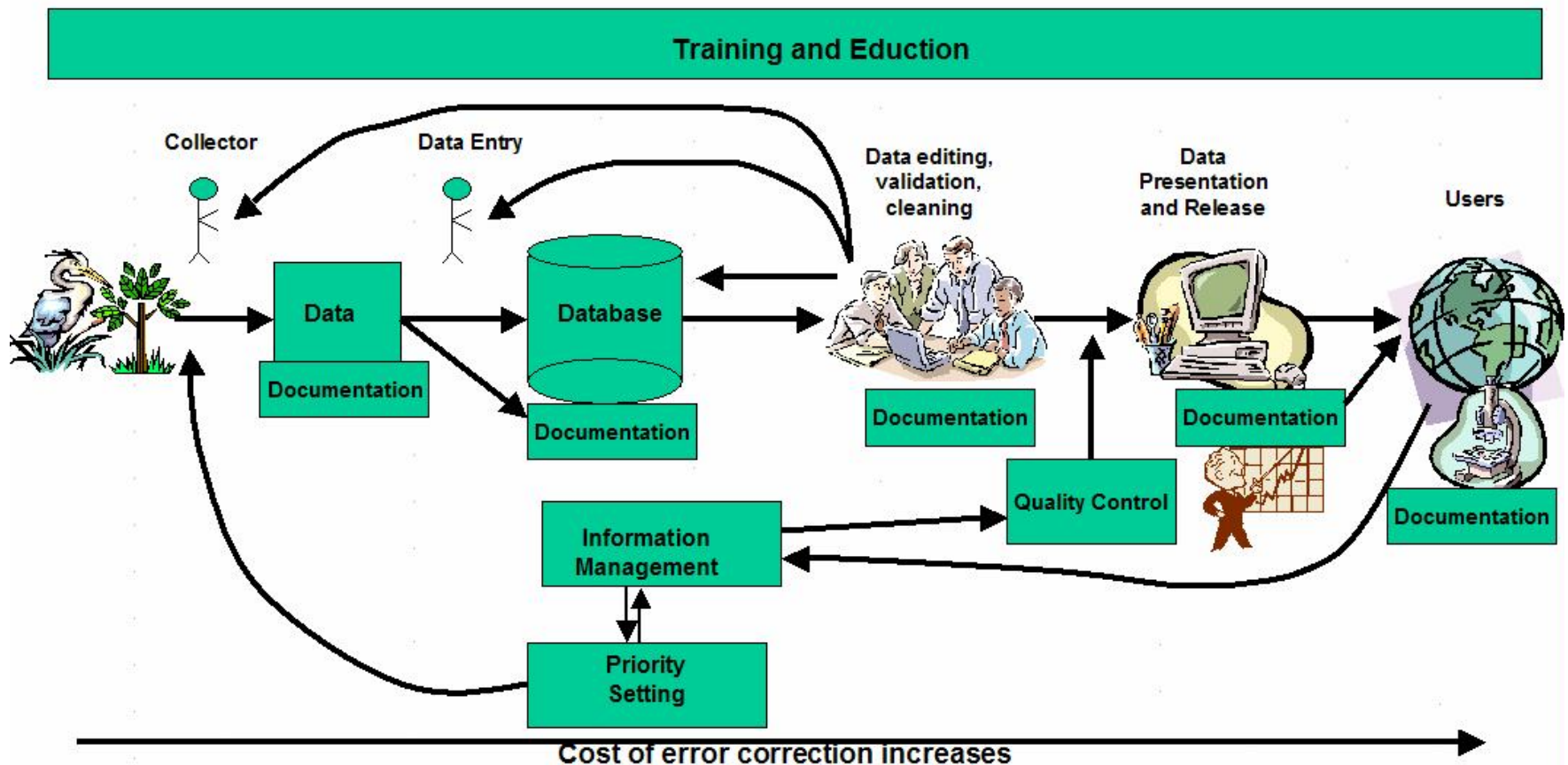
Possesses desired features:

- relevant
- comprehensive
- proper level of detail
- easy to read
- easy to interpret

From Thomas C. Redman, Ph.D., *Data Quality: The Field Guide*, Woburn, MA, Butterworth-Heinemann, 2001, p. 74.

Data quality

Implications related to data handling and use:



Data quality

A 58 pp. document that contains all the basic information and the references to more in-depth documentation on the subject for data users and data compilers



Arthur D. Chapman¹

*Although most data gathering disciplines treat error as an embarrassing issue to be expunged, the error inherent in [spatial] data deserves closer attention and public understanding ...because error provides a critical component in judging fitness for use.
(Chrisman 1991).*

http://www.gbif.org/prog/digit/data_quality/DataQuality

Data quality

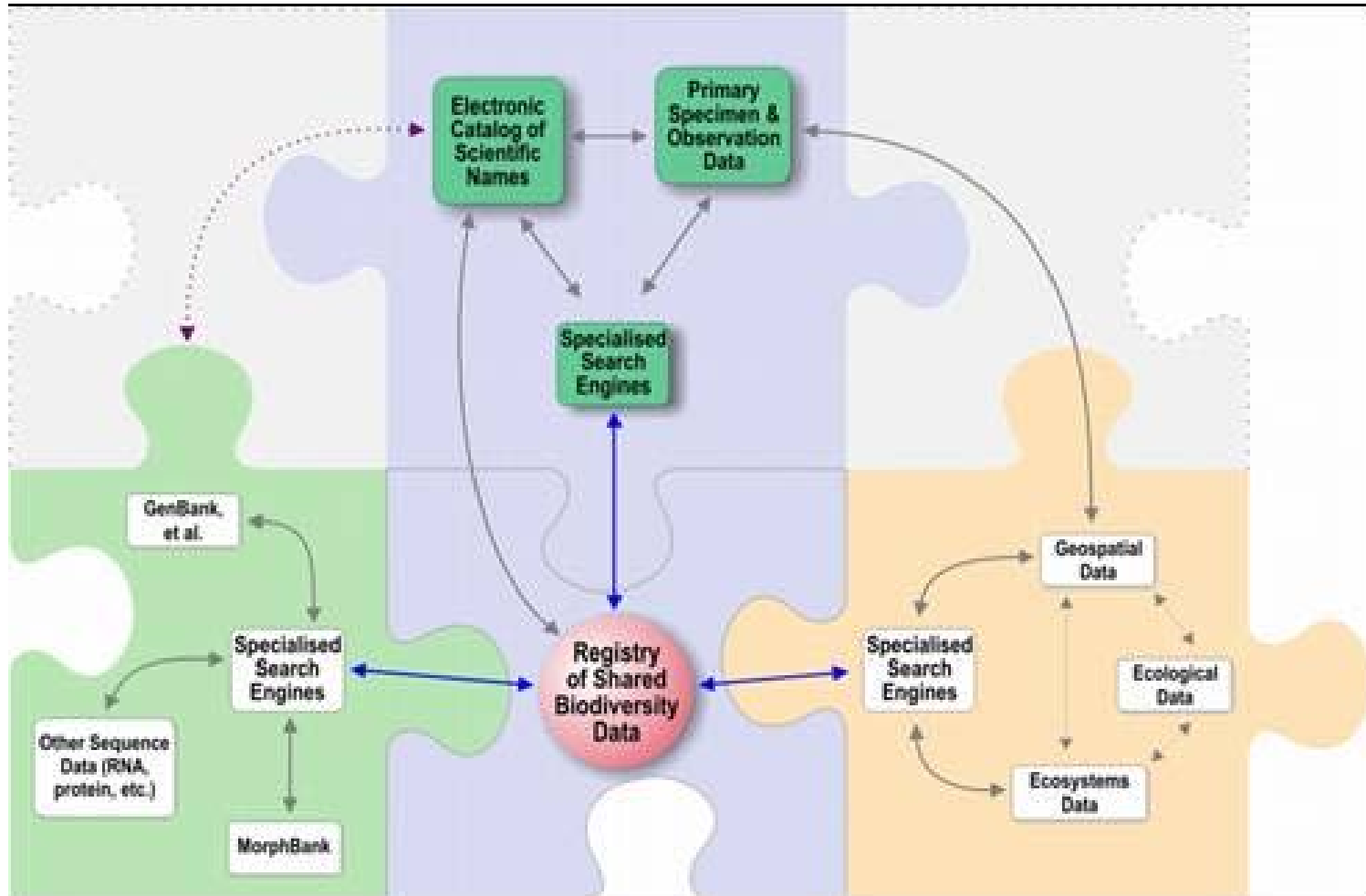
Items for questions or discussion:

- Is any data better than no data?
- Is there a role of the vouchers in biodiversity information?

Environmental data quality and error in data are often neglected issues => data are used uncritically => lead to erroneous results, misleading information => wrong environmental decisions

The nature of biodiversity information

Biodiversity levels, a caveat (2)



The species-level biodiversity information is a central piece, but just a piece of the whole picture

Final consideration

- Any collection (resource) is an important piece to understand biodiversity on Earth and multiple levels
- The way ahead is to provide that "understanding of biodiversity" to the society
- Data providers (collection managers, database scientific administrators, project data managers,...) are in the best situation to make the best use of data

For more information:

Francisco Pando
Node Manager
GBIF Spain
Real Jardín Botánico
Claudio Moyano 1
28014 Madrid, Spain

pando@gbif.es

www.gbif.es

Phone: + 34 91 420 3017

Fax: + 34 91 420 0157

