GBIF Global Biodiversity Information Facility

Foundations for a digitization project
Terminology and standards

Sophie Pamerlon

(based on the presentation done by Sharon Grant for the BID trainings)

Which kind of biodiversity data?

GBIF

Checklists and taxonomical resources

Checklist: a simple list of taxa present in a given area

Taxon list of reference: a valid list of species/taxa present in a given area, with their hierarchy and synonyms

Red list (at national, regional or local level): a list of taxa present in a given area with their vulnerability status

GBIF

GBIF Template for Taxon data

	A	B	C	D	E	F	G
1	taxonID	parentNameUsageID	parentNameUsage	acceptedNameUsageID	acceptedNameUsage	scientificName	nameAccordingToID
2	73			73	Equisetopsida C. Aghard	Equisetopsida C. Aghard	http://dx.doi.org/10.1111/j.1095-8339.2009.010
3	26	73	Equisetopsida C. Aghard	26	Equisetidae Warming	Equisetidae Warming	http://dx.doi.org/10.1111/j.1095-8339.2009.010
4	25	26	Equisetidae Warming	25	Equisetales de Candolle ex Berchtold & J. Presl	Equisetales de Candolle ex Berchtold & J. Presl	http://www.jstor.org/stable/25065646
5	128	25	Equisetales de Candolle ex Berchtold & J. Presl	128	Equisetaceae Michaux ex de Candolle	Equisetaceae Michaux ex de Candolle	http://www.jstor.org/stable/25065646
6	1142	128	Equisetaceae Michaux ex de Candolle	1142	Equisetum Linnaeus	Equisetum Linnaeus	http://www.efloras.org/volume_page.aspx?volur
7	2004	1142	Equisetum Linnaeus	2004	Equisetum subg. Equisetum	Equisetum subg. Equisetum	http://www.efloras.org/volume_page.aspx?volur
8	5467	2004	Equisetum subg. Equisetum	5467	Equisetum fluviatile Linnaeus	Equisetum fluviatile Linnaeus	http://www.efloras.org/volume_page.aspx?volur
9	5466	2004	Equisetum subg. Equisetum	5466	Equisetum arvense Linnaeus	Equisetum arvense Linnaeus	http://www.efloras.org/volume_page.aspx?volur
10	5472	2004	Equisetum subg. Equisetum	5472	Equisetum pratense Ehrhart	Equisetum pratense Ehrhart	http://www.efloras.org/volume_page.aspx?volur
11	5471	2004	Equisetum subg. Equisetum	5471	Equisetum palustre Linnaeus	Equisetum palustre Linnaeus	http://www.efloras.org/volume_page.aspx?volur
12	5474	2004	Equisetum subg. Equisetum	5474	Equisetum sylvaticum Linnaeus	Equisetum sylvaticum Linnaeus	http://www.efloras.org/volume_page.aspx?volur
13	5482	2004	Equisetum subg. Equisetum	5482	Equisetum ×litorale Kühlewein ex Ruprecht	Equisetum ×litorale Kühlewein ex Ruprecht	http://www.efloras.org/volume_page.aspx?volur
14	5476	2004	Equisetum subg. Equisetum	5476	Equisetum telmateia Ehrhart	Equisetum telmateia Ehrhart	http://www.efloras.org/volume_page.aspx?volur
15	15836	5476	Equisetum telmateia Ehrhart	15836	Equisetum telmateia subsp. braunii (J. Milde) H	Equisetum telmateia subsp. braunii (J. Milde) Haul	http://www.efloras.org/volume_page.aspx?volur
16	5481	2004	Equisetum subg. Equisetum	5481	Equisetum ×font-queri Rothmaler	Equisetum ×font-queri Rothmaler	http://www.efloras.org/volume_page.aspx?volur
17	2005	1142	Equisetum Linnaeus	2005	Equisetum subg. Hippochaete (J. Milde) Baker	Equisetum subg. Hippochaete (J. Milde) Baker	http://www.efloras.org/volume_page.aspx?volur
18	5473	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5473	Equisetum scirpoides Michaux	Equisetum scirpoides Michaux	http://www.efloras.org/volume_page.aspx?volur
19	5484	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5484	Equisetum ×nelsonii (A.A. Eaton) J.H. Schaffner	Equisetum ×nelsonii (A.A. Eaton) J.H. Schaffner	http://www.efloras.org/volume_page.aspx?volur
20	5478	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5478	Equisetum variegatum Schleicher ex F. Weber &	Equisetum variegatum Schleicher ex F. Weber & D	http://www.efloras.org/volume_page.aspx?volur
21	5477	5478	Equisetum variegatum Schleicher ex F. Weber & D. Mohr	5477	Equisetum variegatum subsp. alaskanum (A.A.	Equisetum variegatum subsp. alaskanum (A.A. Eat	http://www.efloras.org/volume_page.aspx?volur
22	5479	5478	Equisetum variegatum Schleicher ex F. Weber & D. Mohr	5479	Equisetum variegatum Schleicher ex F. Weber &	Equisetum variegatum Schleicher ex F. Weber & D	http://www.efloras.org/volume_page.aspx?volur
23	5470	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5470	Equisetum laevigatum A. Braun	Equisetum laevigatum A. Braun	http://www.efloras.org/volume_page.aspx?volur
24	5480	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5480	Equisetum ×ferrissii Clute	Equisetum ×ferrissii Clute	http://www.efloras.org/volume_page.aspx?volur
25	5469	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5469	Equisetum hyemale Linnaeus	Equisetum hyemale Linnaeus	http://www.efloras.org/volume_page.aspx?volur
26	5468	5469	Equisetum hyemale Linnaeus	5468	Equisetum hyemale subsp. affine (Engelmann)	Equisetum hyemale subsp. affine (Engelmann) Cal	http://www.efloras.org/volume_page.aspx?volur
27	5483	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5483	Equisetum ×mackayi (Newman) Brichan	Equisetum ×mackayi (Newman) Brichan	http://www.efloras.org/volume_page.aspx?volur

Used for sharing **taxonomic information**: red lists, checklists…

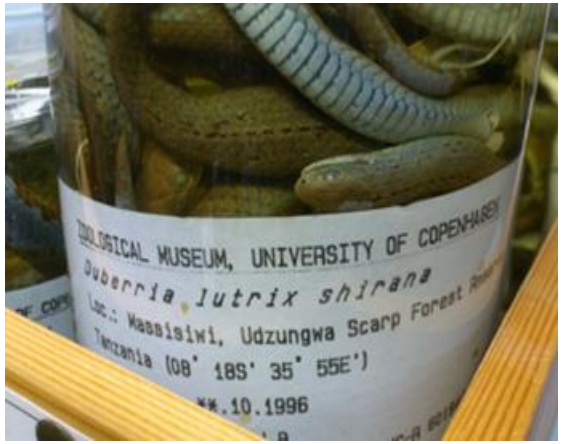
Each line = 1 taxon (not necessarily to the species level); you cannot have the same taxon twice in your list

Fields = all taxonomic levels (from kingdom to subspecies), with authorship, references and additional information (endangerment status, geographic details, etc.)

GBIF

Specimens and materials

Herbarium sheets and vegetal materials (seeds, foliage, branches, bark, dried/preserved fruits…)

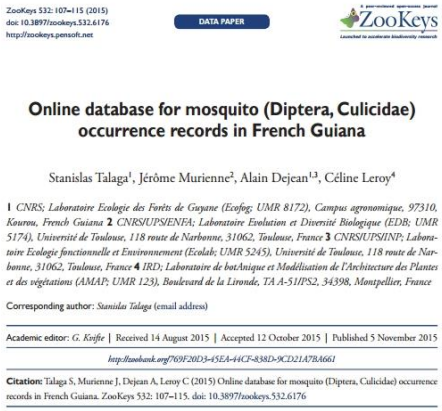


Preserved specimens in formol, alcohol (fishes, herpetology collections…); mounted specimens (birds, mammals, insects)

Fossils and other paleontological materials (amber, teeth, bones…); animal or vegetal samples (DNA, organs, skin, fur, faeces…)

GBIF

Literature documents



ZooKeys 532: 107–115 (2015)
doi: 10.3897/zookeys.532.6176
http://zookeys.pensoft.net

DATA PAPER

ZooKeys

Online database for mosquito (Diptera, Culicidae) occurrence records in French Guiana

Stanislas Talaga[1], Jérôme Murienne[2], Alain Dejean[1,3], Céline Leroy[4]

1 *CNRS; Laboratoire Ecologie des Forêts de Guyane (Ecofog; UMR 8172), Campus agronomique, 97310, Kourou, French Guiana* 2 *CNRS/UPS/ENFA; Laboratoire Evolution et Diversité Biologique (EDB; UMR 5174), Université de Toulouse, 118 route de Narbonne, 31062, Toulouse, France* 3 *CNRS/UPS/INP; Laboratoire Ecologie fonctionnelle et Environnement (Ecolab; UMR 5245), Université de Toulouse, 118 route de Narbonne, 31062, Toulouse, France* 4 *IRD; Laboratoire de botAnique et Modélisation de l'Architecture des Plantes et des végétations (AMAP; UMR 123), Boulevard de la Lironde, TA A-51/PS2, 34398, Montpellier, France*

Corresponding author: *Stanislas Talaga* (email address)

Academic editor: *G. Knife* | Received 14 August 2015 | Accepted 12 October 2015 | Published 5 November 2015

http://zoobank.org/769F20D3-45EA-44CF-838D-9CD21A7BA661

Citation: Talaga S, Murienne J, Dejean A, Leroy C (2015) Online database for mosquito (Diptera, Culicidae) occurrence records in French Guiana. ZooKeys 532: 107–115. doi: 10.3897/zookeys.532.6176

INVENTAIRE ET UTILISATION DURABLE DE LA FAUNE MAMMALIENNE EN MILIEU FORESTIER EQUATORIAL : CAS DU SECTEUR OUEST DE LA RESERVE DE LA BIOSPHERE DU DJA (SUD-CAMEROUN)

THESE PRESENTEE PAR

NGANDJUI Germain

Sous la direction de

Charles-Pierre BLANC, Professeur

BUREAU D'ETUDE/CONSULTANT

Études d'impact environnemental

Published or in press scientific articles

PhD or Master thesis

Reports and other written documents

GBIF

Fieldwork records and notes

Surveys, assessments

Logs, field notes with taxa observed or collected; notes about the protocol used on the field

Citizen science logs

GBIF Template for Occurrence data

	A	B	C	D	E	F	G	H	I	J	K
1	occurrenceID	basisOfRecord	eventDate	endDayOfYear	year	month	day	verbatimEventDate	eventRemarks	scientificName	higherClassifica
2	http://arctos.database.m	PreservedSpecimen	1926-04		1926	4		0/4/1926	day of month unknown	Ambystoma maculatum	Animalia; Chord
3	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Desmognathus fuscus	Animalia; Chord
4	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Gyrinophilus porphyriticus	Animalia; Chord
5	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Eurycea bislineata bislineata	Animalia; Chord
6	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Plethodon cinereus	Animalia; Chord
7	http://arctos.database.m	PreservedSpecimen	1953-09-27	270	1953	9	27	27-sept-53		Rana sylvatica	Animalia; Chord
8	http://arctos.database.m	PreservedSpecimen	1979-06-02/1979-06-07					02/06/1979		Eleutherodactylus eneidae	Animalia; Chord
9	http://arctos.database.m	PreservedSpecimen	1981-06-01	152	1981	6	1	01-juin-81		Masticophis flagellum piceus	Animalia; Chord
10	http://arctos.database.m	PreservedSpecimen	2011-06-23	174	2011	6	23	23-juin-11		Rana (Lithobates) clamitans	Animalia; Chord

Occurrence = simple observation in the field or specimen in a collection

Each line = 1 individual or 1 group of individuals (you can have several occurrences of the same species/taxon in your file)

Fields = What? Where? When? How? By whom was the individual(s) observed and/or collected? (+ additional information: habitat, coordinates, associated species, etc.)

GBIF

GBIF Template for Event data

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	eventID	samplingProtocol	samplingEffort	sampleSizeValue	sampleSizeUnit	eventDate	eventTime	startDayOfYear	eventRemarks	country	countryCode	locality	locationID
2	994-tr009-s00	Pollard walks	Average of 30 Minutes walk along transect	250	square metre	2012-10-11	09:28:02Z/10:16:02Z	284	No occurrences	Israel	IL	Sde boker reches halukim	tr009-s00
3	3502-tr056-s6	Pollard walks	Average of 30 Minutes walk along transect	250	square metre	2015-10-19	12:25:02Z/13:10:02Z	291		Israel	IL	Nahal Kovshim Beer Sheva	tr056-s6
4	3502-tr056-s9	Pollard walks	Average of 30 Minutes walk along transect	250	square metre	2015-10-19	12:25:02Z/13:10:02Z	291		Israel	IL	Nahal Kovshim Beer Sheva	tr056-s9

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	eventID	occurrenceID	basisOfRecord	individualCount	organismQuantity	organismQuantityType	occurrenceStatus	scientificName	kingdom	phylum	class	order	family	infraspecificEpithet	taxonRank	recordedBy
2	1382-tr009-s00	1382-tr009-s00-0	HumanObservation	0	0	individuals	absent	Lepidoptera	Animalia	Arthropoda	Insecta	Lepidoptera			order	Eviatar Fingo
3	3502-tr056-s6	3502-tr056-s6-21114	HumanObservation	3	3	individuals	present	Azanus jesous	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
4	3502-tr056-s6	3502-tr056-s6-21126	HumanObservation	1	1	individuals	present	Melitaea trivia	Animalia	Arthropoda	Insecta	Lepidoptera	Nymphalidae		species	Zvika Avni
5	3502-tr056-s6	3502-tr056-s6-21127	HumanObservation	3	3	individuals	present	Deudorix livia	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
6	3502-tr056-s6	3502-tr056-s6-21129	HumanObservation	1	1	individuals	present	Azanus ubaldus	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
7	3502-tr056-s6	3502-tr056-s6-21132	HumanObservation	1	1	individuals	present	Lycaena thersamon	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
8	3502-tr056-s9	3502-tr056-s9-21116	HumanObservation	1	1	individuals	present	Azanus jesous	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
9	3502-tr056-s9	3502-tr056-s9-21122	HumanObservation	1	1	individuals	present	Tarucus balkanica	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
10	3502-tr056-s9	3502-tr056-s9-21131	HumanObservation	1	1	individuals	present	Azanus ubaldus	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni

Used for sharing **more complex information** about a sampling event on the field: area description, protocols used, occurrences collected or observed, variables recorded…

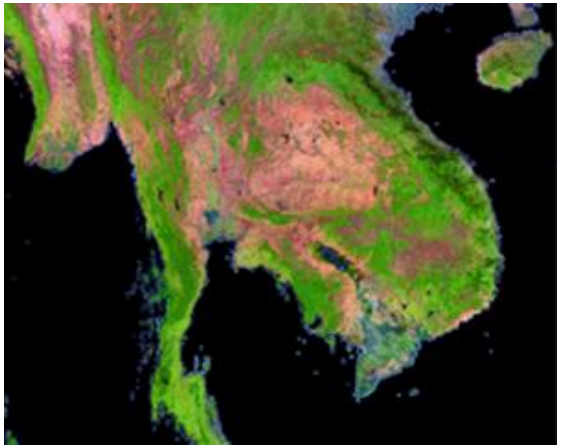
Event data often come **in several sheets**: data about the event itself (transect, trap, quadrat…), data about the occurrences recorded for each event, data about variables, etc.

Each line in the event sheet = 1 event (e.g. a camera trap, a transect, a vegetation plot…)

Each field = description information (size of the plot, protocol, coordinates…)

GBIF

Other origins of data



Remote sensing data: GPS, radar or satellite data; camera traps;

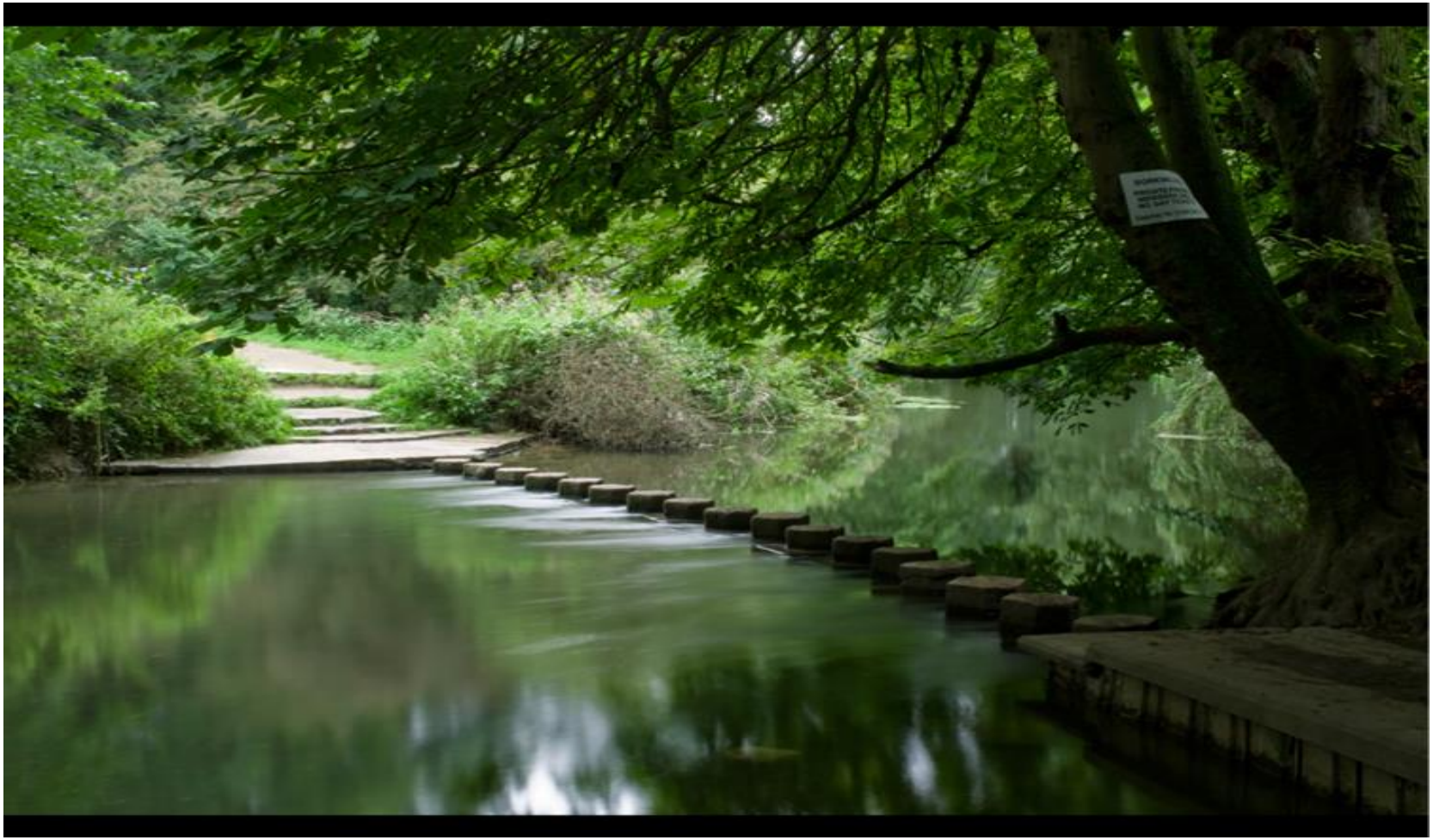
Paper maps or atlas; prints of satellite pictures

Other supports: pictures, audio, video recordings

From data to understanding…

Oceans of data…

…rivers of information…



… streams of knowledge…

…droplets of understanding

Fitness for use - Definition

Data quality is a relative concept that depends on the use of these data.

"The general intent of describing the quality of a particular dataset or record is to describe the fitness of that dataset or record for a particular use that one may have in mind for the data."

Chrisman, 1991

Fitness for Use in the Real World

"How well does a thing do what it's supposed to and what is that anyway?"

Fitness for Use in the Real World

"How well does a thing do what it's supposed to and what is that anyway?"

Fitness for Use in the Real World

"How well does a thing do what it's supposed to and what is that anyway?"

Fitness for Use - Data

Do you understand your data and can you explain its purpose to someone else?

1. accessibility,
2. accuracy,
3. timeliness,
4. completeness / comprehensiveness,
5. consistency,
6. relevancy,
7. well documented [outside of your head],
8. easy to read and easy to interpret

Reference: Chapman, A. D. 2005. Principles of Data Quality, version 1.0. Report for GBIF, Copenhagen. ISBN 87-92020-03-8.

Data Processing and Quality

Each institution should have:

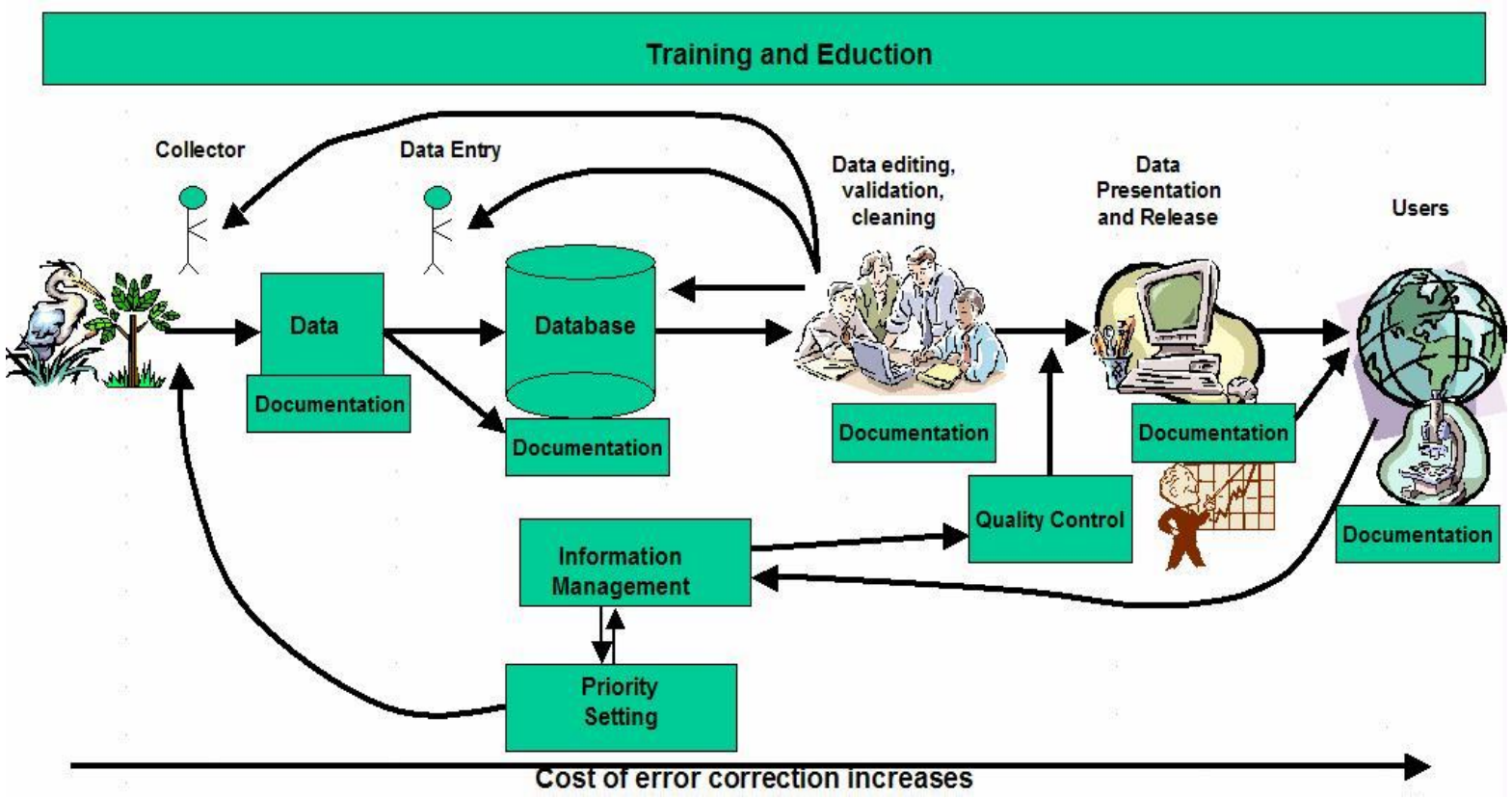
1.A **vision** targeted on data quality

- **Don’t « reinvent the wheel » and use standards**
- **Seek efficiency (in collecting data and quality checks) and avoid duplicating efforts**
- **Promote sharing (data, informations, tools, standards…)**
- **Think at a large scale**
- **Cater to users and their needs**
- Invest in **documentation** and **metadata**

1.A **policy** implementing this vision

1.An **implementation strategy** for this policy (precise goals at short, mean and long term)

Data Processing and Quality



Quality loss happens at every step.
The responsibility in terms of data quality has to be assigned at the earlier possible step of the process.

Sharing responsibilities

Collector

Labels and logs are as **correct, complete and readable** as possible

Collection methods are vastly documented

Remarks are clear and non-ambiguous

Curator

Retranscription quality in the database

Regular validation **tests**.

Data regularly saved and **archived**

Keep precedent **versions**

Ensure **respect** of private life, intellectual rights, local traditions and sensibilities ...

Provide quality **documentation** (including known issues about the data)

Take **feedback** into account

Responsibility for maintenance but also moral responsibility to improve data quality (if possible) for future uses and users.

User

Inform data curators about **mistakes** and omissions in data and **documentation.**

Provide **feedback** to define future priorities

When using data, determine whether **data are adequate for intended use** and not use them if this is not the case.

Data quality and Data capture

Metadata :

- Description of the whole dataset (title, summary, contacts, licence, logo…)

Taxonomic information :

- Scientific names, vernacular names, reference lists used

Spatial information :

- Coordinates, locality, altitude, depth...

Collection (event) information :

- Collector's name, habitat, date...

Descriptive information

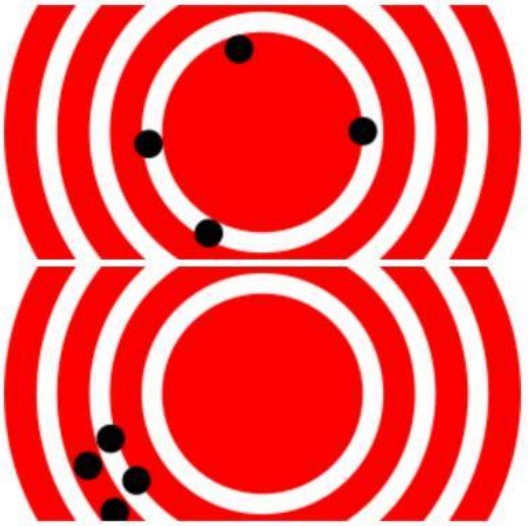
- Age, sex, behaviour, quantity of individuals observed/collected...

Measures of Quality

"All data include error – there is no escaping it! It is knowing what the error is that is important, and knowing if the error is within acceptable limits for the purpose to which the data are to be put. (Chapman 2005)"

- **Correctness (Accuracy)**
  - How close is the recorded value to the actual value?

- **Consistency (Precision)**
  - How often do you get it right?



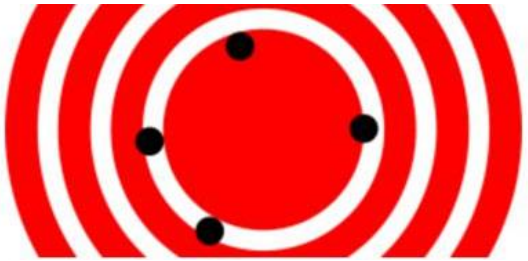
Reference: http://www.businessdictionary.com/definition/data-cleaning.html

Correctness - examples

"All data include error – there is no escaping it! It is knowing what the error is that is important, and knowing if the error is within acceptable limits for the purpose to which the data are to be put. (Chapman 2005)"

Correctness (Accuracy)

How close is the recorded value to the actual value?



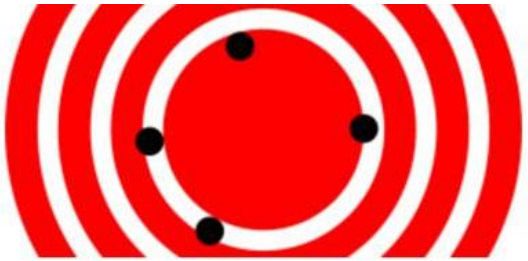
Reference: http://www.businessdictionary.com/definition/data-cleaning.html

Correctness - example 1

"All data include error – there is no escaping it! It is knowing what the error is that is important, and knowing if the error is within acceptable limits for the purpose to which the data are to be put. (Chapman 2005)"

Correctness (Accuracy)

How close is the recorded value to the actual value?



A dataset contains fossil specimens from the Triassic period.

The recorded taxa for a specimen Is *Thismia.*

Is Thismia a fossil bird?

Reference: http://www.businessdictionary.com/definition/data-cleaning.html

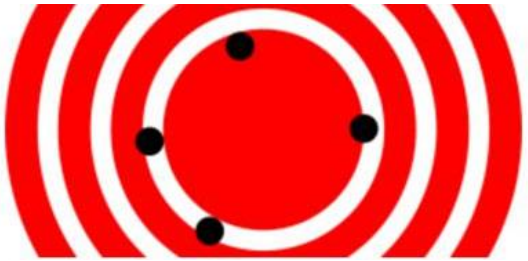
GBIF

Correctness - example 1

"All data include error – there is no escaping it! It is knowing what the error is that is important, and knowing if the error is within acceptable limits for the purpose to which the data are to be put. (Chapman 2005)"

Correctness (Accuracy)

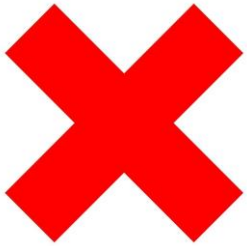
How close is the recorded value to the actual value?



A dataset contains fossil specimens from the Triassic period.

The recorded taxa for a specimen Is *Thismia.*

Is Thismia a fossil bird?



Reference: http://www.businessdictionary.com/definition/data-cleaning.html

Consistency - example

"All data include error – there is no escaping it! It is knowing what the error is that is important, and knowing if the error is within acceptable limits for the purpose to which the data are to be put. (Chapman 2005)"

Consistency (Precision)

How often do you get it right?

A botanical dataset has specimens collected by:

Full Name = Joseph Dalton Hooker

Full Name = Hooker, J.

Full Name = W. J. Hooker

Full Name = Hook.f.

Full Name = Hook.

How many unique collectors are there?

Reference: http://www.businessdictionary.com/definition/data-cleaning.html

Consistency - example

"All data include error – there is no escaping it! It is knowing what the error is that is important, and knowing if the error is within acceptable limits for the purpose to which the data are to be put. (Chapman 2005)"

Consistency (Precision)

How often do you get it right?

A botanical dataset has specimens collected by:

Full Name = Joseph Dalton Hooker

Full Name = Hooker, J.

Full Name = W. J. Hooker

Full Name = Hook.f.

Full Name = Hook.

How many unique collectors are there?

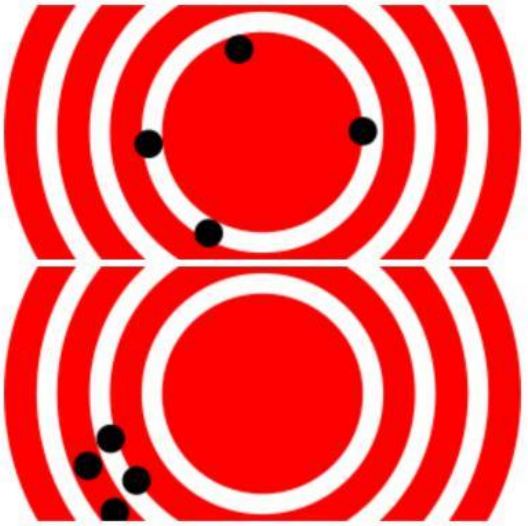
Reference: http://www.businessdictionary.com/definition/data-cleaning.html

Data Cleaning

Data cleaning is the process of correcting or removing dirty data caused by contradictions, disparities, keying mistakes, missing bits, etc. It also includes validation of the changes made, and may require normalization.

- **Correctness (Accuracy)**
  - How close is the recorded value to the actual value?

- **Consistency (Precision)**
  - How often do you get it right?



Reference: http://www.businessdictionary.com/definition/data-cleaning.html

What is a Standard?

"An agreed way of doing something."

An agreed way of doing something, to provide clarity and help communication.

- Norm
- Convention
- Specification
- Requirement
- Restriction
- Rule

Everyday Standards

"The main purpose for standards is to create a framework to ease sharing. They should provide clarity and help communication."

Some examples of standards that you use often:

- Units of Measurement (Metric, Imperial)
- Numeral Systems (Hindu-Arabic; Roman Numerals)
- Alphabets
- Languages
- Emojis
- Postal Addressing
- Morse Code

Reference(s): https://www.bsigroup.com/en-US/Standards/Information-about-standards/What-is-a-standard/

Natural History Standards

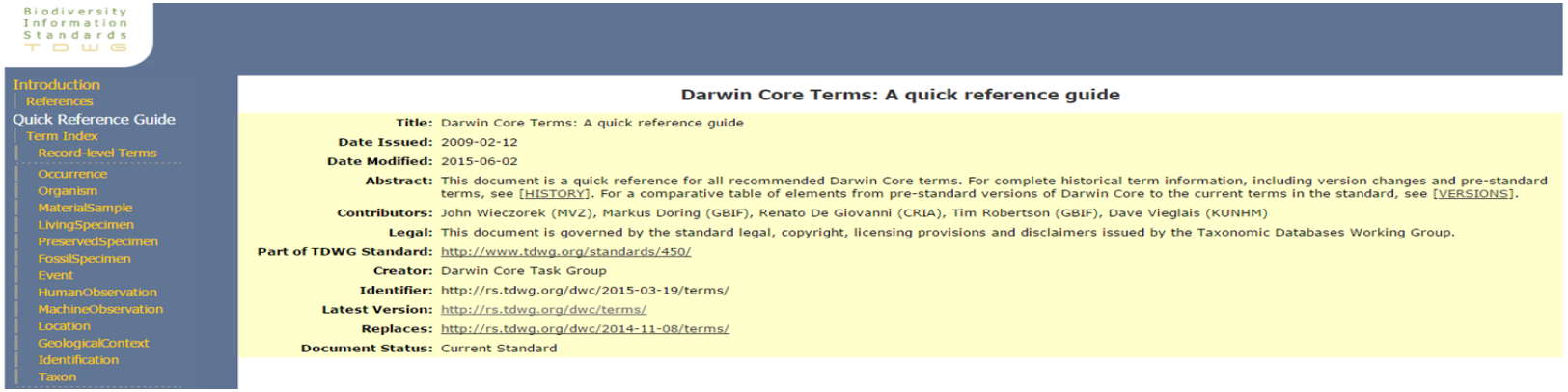
"Data standards are the rules by which data are described and recorded. In order to share, exchange, and understand data, we must standardise the format as well as the meaning." (USGS)

Some standards which already exist:

- Ecological Metadata Language Standard (EML),
- Audubon Media Description (aka Audubon Core),
- Global Genome Biodiversity Network(GGBN),
- Ocean Data Standards and Best Practices Project (ODSBP),
- **Darwin Core**

What is Darwin Core?

"List of fields and their definitions, as they relate to biodiversity data."

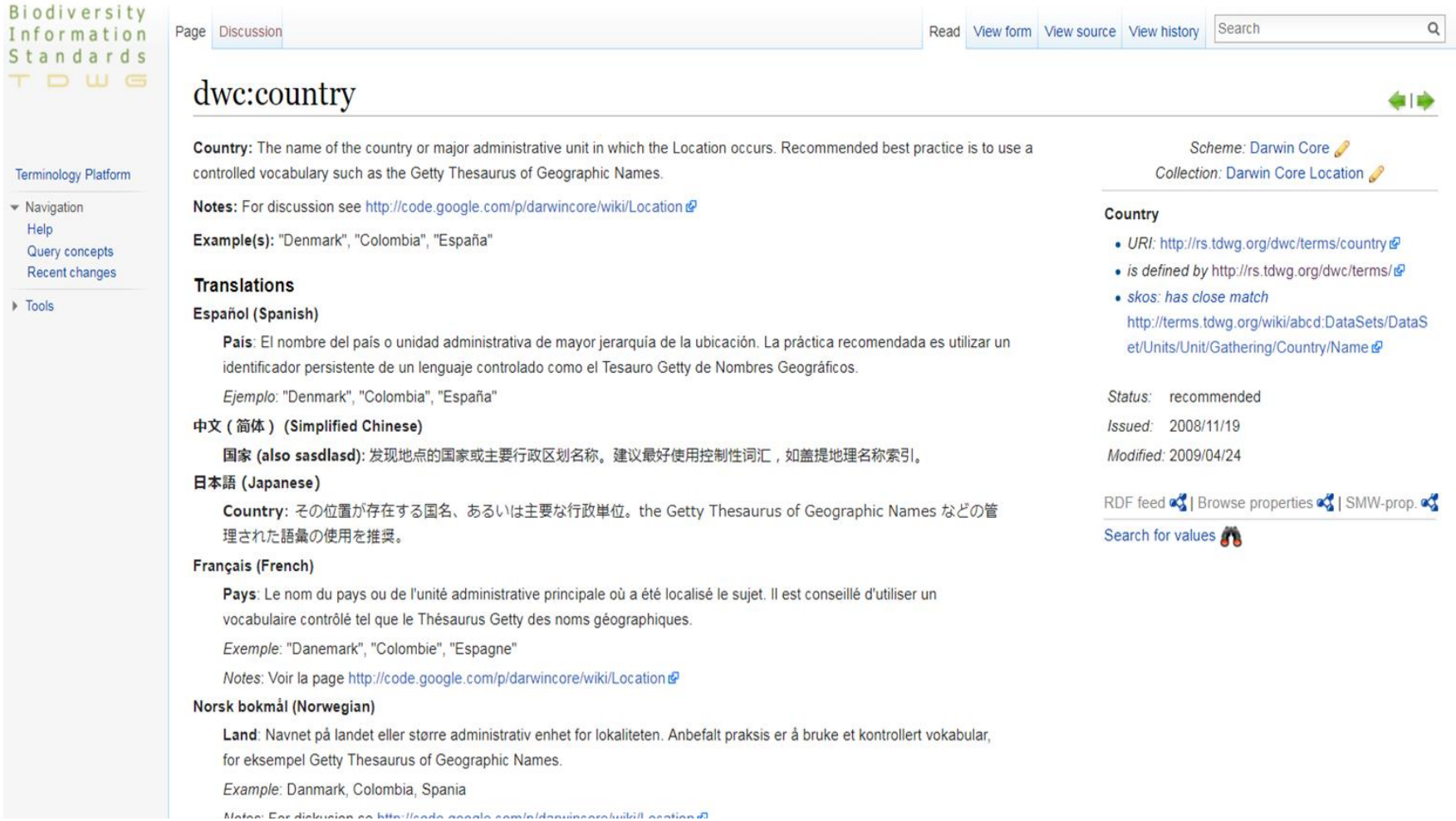


Term Name: country	
Identifier:	http://rs.tdwg.org/dwc/terms/country
Class:	http://purl.org/dc/terms/Location
Definition:	The name of the country or major administrative unit in which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names.
Comment:	Examples: "Denmark", "Colombia", "España". For discussion see http://terms.tdwg.org/wiki/dwc:country
Details:	country

Reference(s): Governance, http://www.tdwg.org; Standard, http://rs.tdwg.org/dwc.

GBIF

What is Darwin Core?



Biodiversity Information Standards TDWG

Terminology Platform

Navigation
Help
Query concepts
Recent changes

Tools

Page | Discussion | Read | View form | View source | View history | Search

dwc:country

Country: The name of the country or major administrative unit in which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names.

Notes: For discussion see http://code.google.com/p/darwincore/wiki/Location

Example(s): "Denmark", "Colombia", "España"

Translations

Español (Spanish)

País: El nombre del país o unidad administrativa de mayor jerarquía de la ubicación. La práctica recomendada es utilizar un identificador persistente de un lenguaje controlado como el Tesauro Getty de Nombres Geográficos.

Ejemplo: "Denmark", "Colombia", "España"

中文（简体）(Simplified Chinese)

国家 (also sasdlasd): 发现地点的国家或主要行政区划名称。建议最好使用控制性词汇，如盖提地理名称索引。

日本語 (Japanese)

Country: その位置が存在する国名、あるいは主要な行政単位。the Getty Thesaurus of Geographic Names などの管理された語彙の使用を推奨。

Français (French)

Pays: Le nom du pays ou de l'unité administrative principale où a été localisé le sujet. Il est conseillé d'utiliser un vocabulaire contrôlé tel que le Thésaurus Getty des noms géographiques.

Exemple: "Danemark", "Colombie", "Espagne"

Notes: Voir la page http://code.google.com/p/darwincore/wiki/Location

Norsk bokmål (Norwegian)

Land: Navnet på landet eller større administrativ enhet for lokaliteten. Anbefalt praksis er å bruke et kontrollert vokabular, for eksempel Getty Thesaurus of Geographic Names.

Example: Danmark, Colombia, Spania

Notes: For diskusjon se http://code.google.com/p/darwincore/wiki/Location

Scheme: Darwin Core
Collection: Darwin Core Location

Country

- *URI:* http://rs.tdwg.org/dwc/terms/country
- *is defined by* http://rs.tdwg.org/dwc/terms/
- *skos: has close match* http://terms.tdwg.org/wiki/abcd:DataSets/DataSet/Units/Unit/Gathering/Country/Name

Status: recommended
Issued: 2008/11/19
Modified: 2009/04/24

RDF feed | Browse properties | SMW-prop.
Search for values

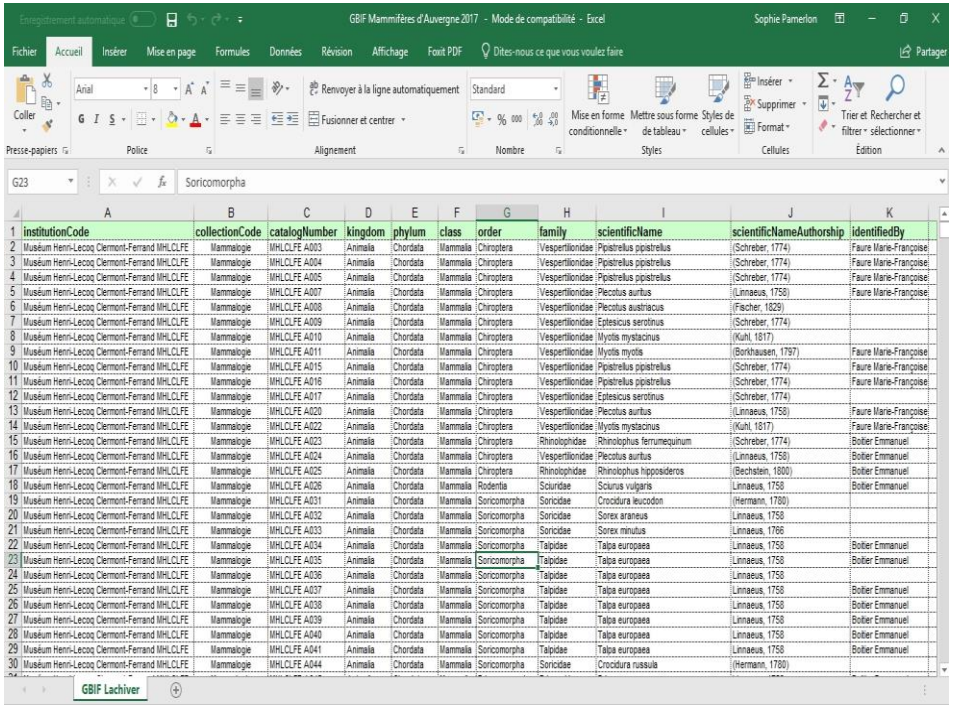
Reference(s): https://terms.tdwg.org/wiki/dwc:country

GBIF

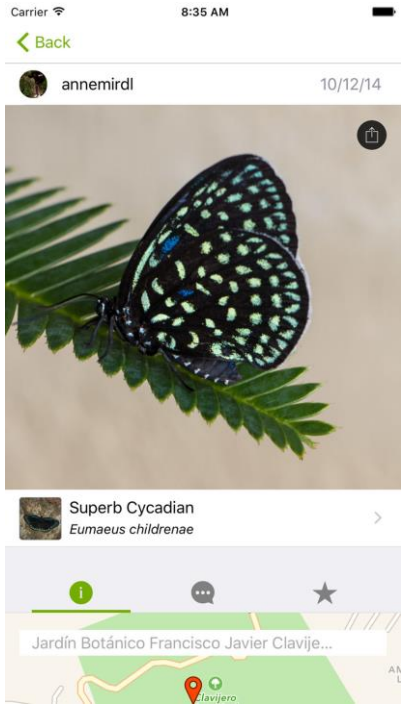
Which software to use?

Softwares for data capture and data management

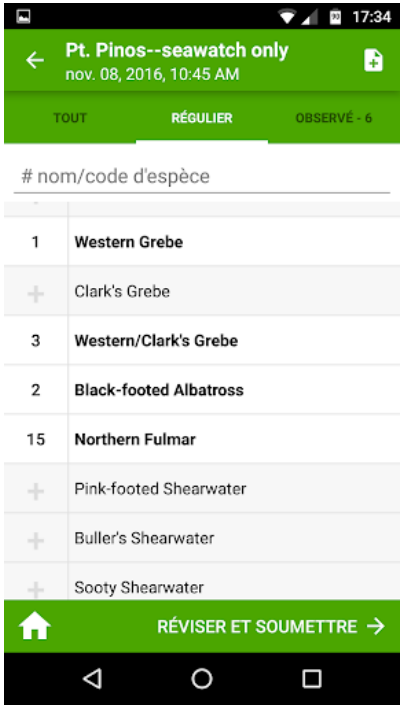
- No need to go “big”: you can capture information from collection specimens or live observations in a simple spreadheet such as Excel, OpenOffice, Google Sheets or through an app (iNaturalist, Biodiversity Data Capture, eBird for ornithology, Memento…)



Excel spreadsheet for biodiversity data



INaturalist app



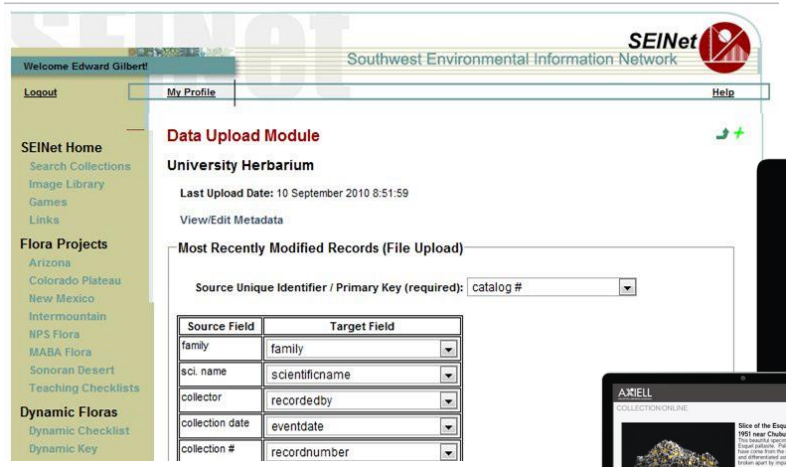
eBird app

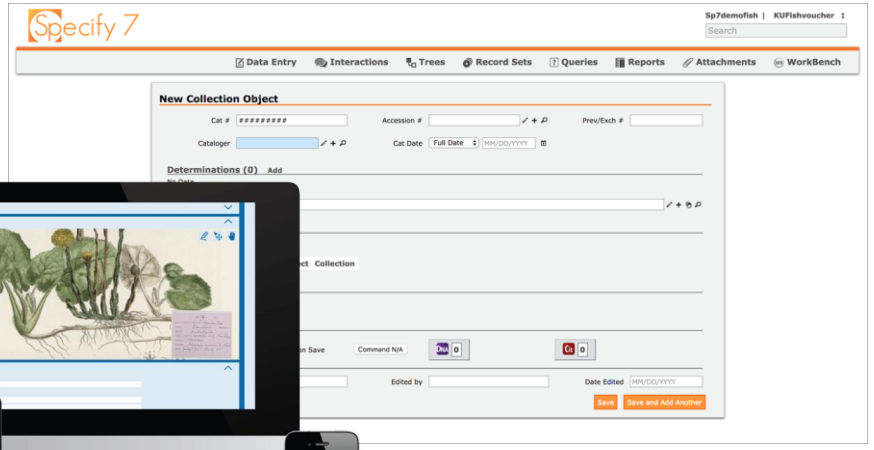
GBIF

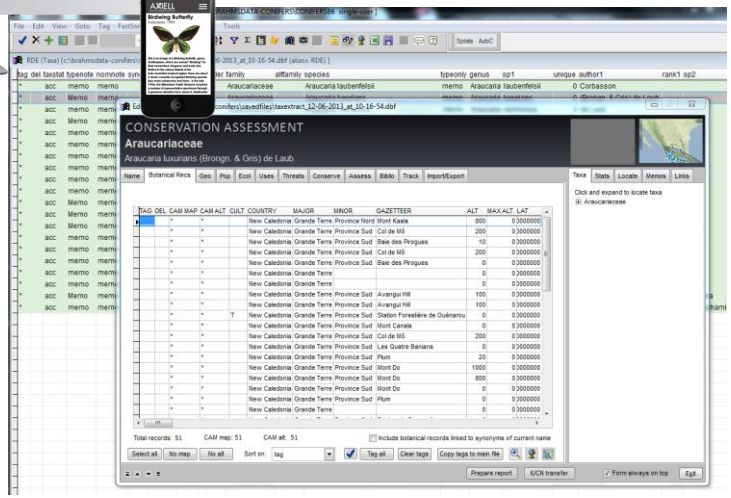
Which software to use?

Softwares for data capture and data management

- Depending on your needs, a more detailed data capture/management software might be the right solution for your data:





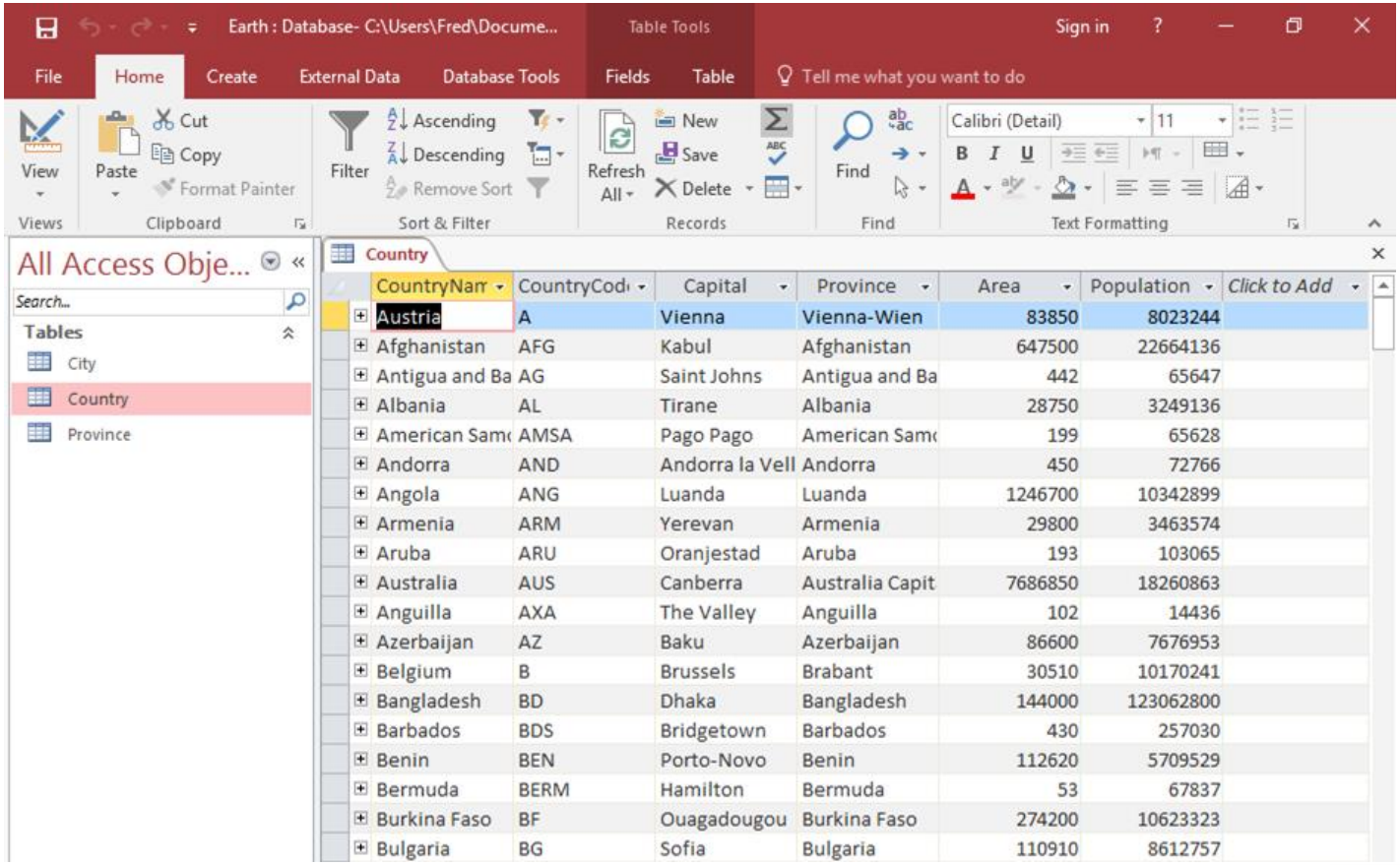


From left to right:
Symbiota, Specify
EMu
Elysia, Brahms

GBIF

Softwares for data capture and data management

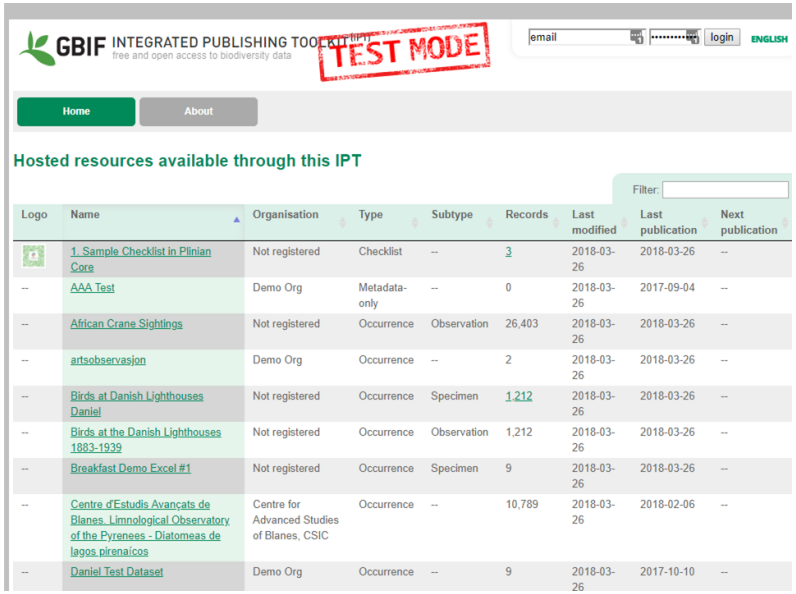
- Relational databases such as Access are useful for storing and managing data



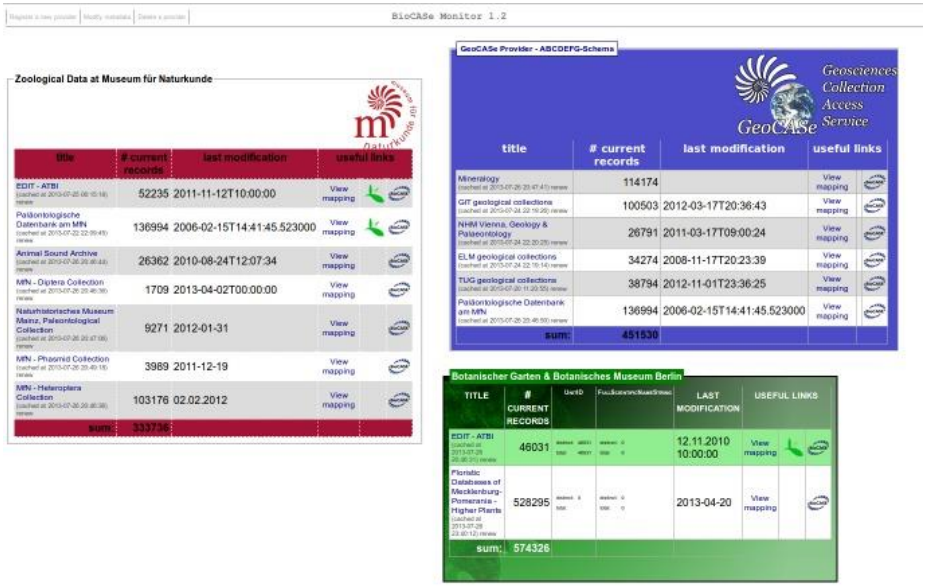
GBIF

Softwares for data capture and data management

- Going further: anticipate data publishing! (More about this in the days to come…) Make sure that your software or tool allows you to export data as csv or txt files. Then you will be able to share them with the international scientific and deciders community through tools such as the GBIF IPT or a BioCASe installation:



IPT (GBIF)



BioCASe

GBIF

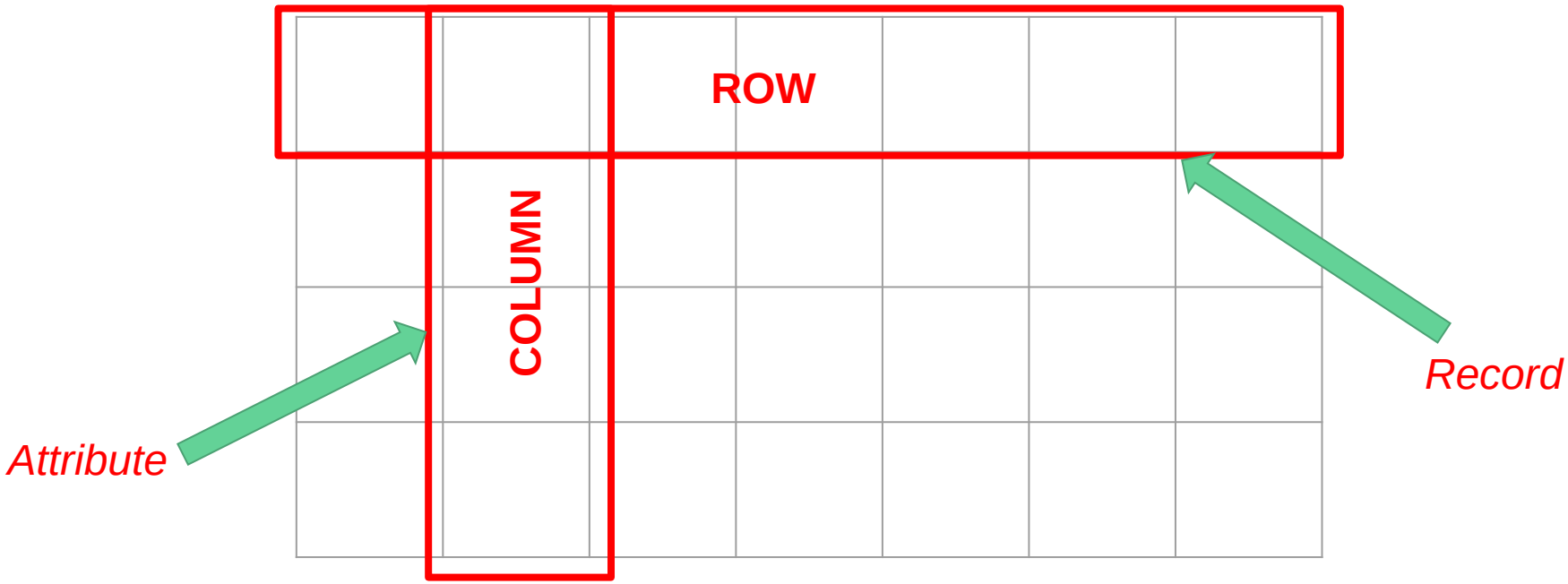
Which software to use?

Softwares for data capture and data management

- The choice is yours!
- Each institution should discuss data workflow, goals, standards, licences, softwares, in order to choose the tools best suited for their needs
- Each software has its pros and cons (price, language, functionalities, documentation, new versions…), but whole user communities are available to help you decide which one is the best for you and your institution
- Feel free to exchange feedback and advice between colleagues!

Data Structures - Tables

"What happens when you add structure to a grid?"



Integrity and Security

"The difference between a spreadsheet and a database table."

Row

Attributes of a record **ALWAYS** stay together.

Column

Any attribute has the **SAME** field/data type for every record.

Table

All data in a table refers to a **SINGLE** concept.

Row

For example a single specimen.

Column

For example the collector's name.

Table

For example everything collected on an expedition

Metadata

”Metadata must be rich enough to allow data (re)use by a third party without them having to refer to the data source.”

Metadata = « Data about the data »

- **Describe content, accessibility, completeness**…
- About the **dataset**
- **Error documentation**
- Documentation of **validation process**, data cleaning and data correcting

Metadata

Experience has shown that treating data as a long-term asset and managing it within a coordinated framework produces considerable savings and ongoing value. (NLWRA 2003).

- title;
- narrative;
- source;
- data lineage;
- accuracy;
- logical consistency;
- date and life expectancy;
- field definitions;
- collection methodology;
- completeness;
- conditions of use and use constraints;
- custodianship;
- contact information

Reference: Chapman, A. D. 2005. Principles of Data Quality, version 1.0. Report for GBIF, Copenhagen. ISBN 87-92020-03-8.

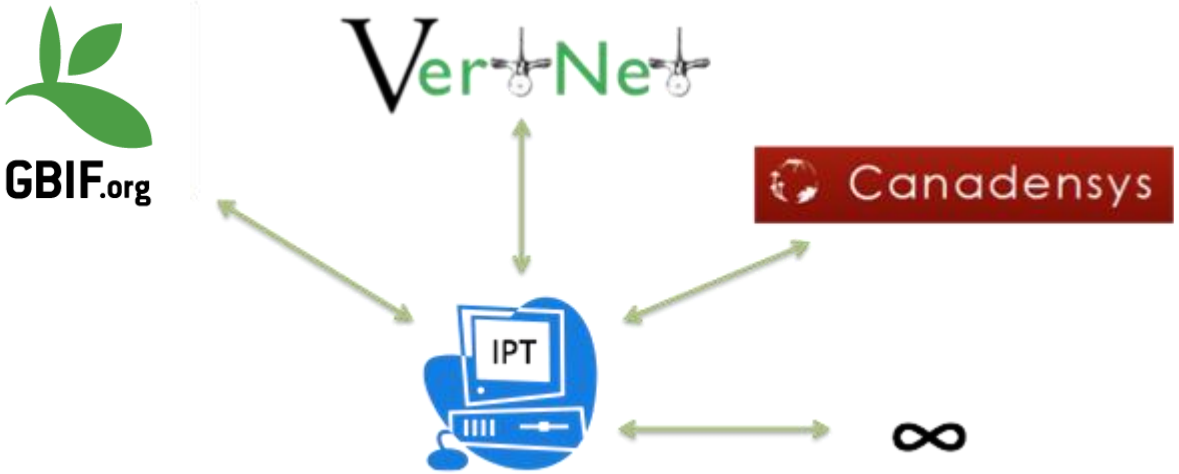
GBIF

Publishing

"Nothing stinks like a pile of unpublished writing" Sylvia Plath

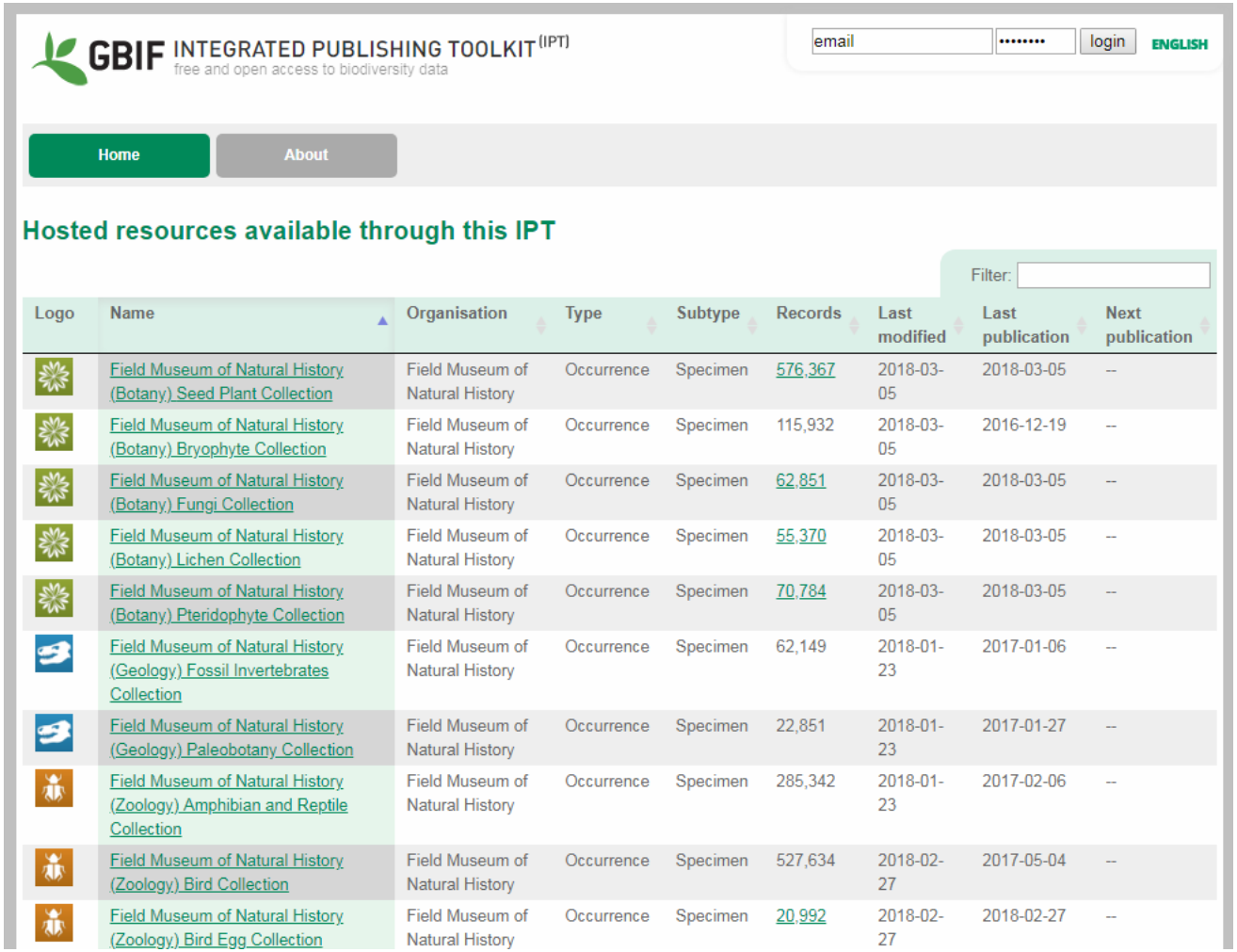
What is Data Publishing?

"Publishing" refers to making biodiversity datasets publicly accessible and discoverable, in a standardized form, via an access point, typically a web address (a URL).



What Does IPT Stand For?

Integrated Publishing Toolkit



GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)
free and open access to biodiversity data

email ••••••• login ENGLISH

Home | About

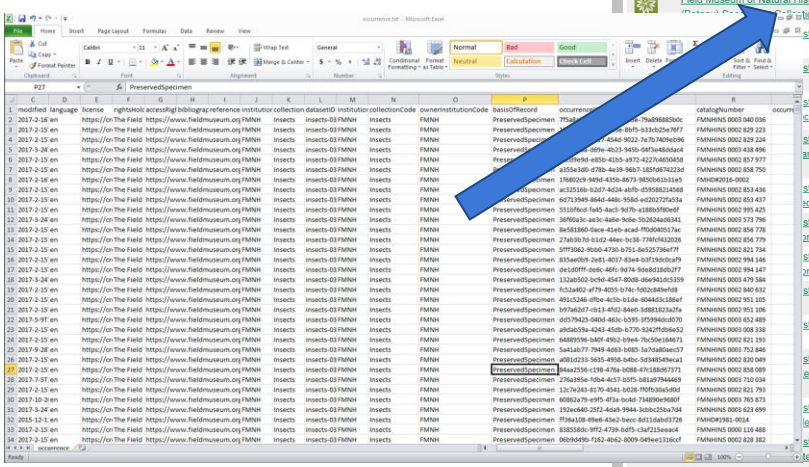
Hosted resources available through this IPT

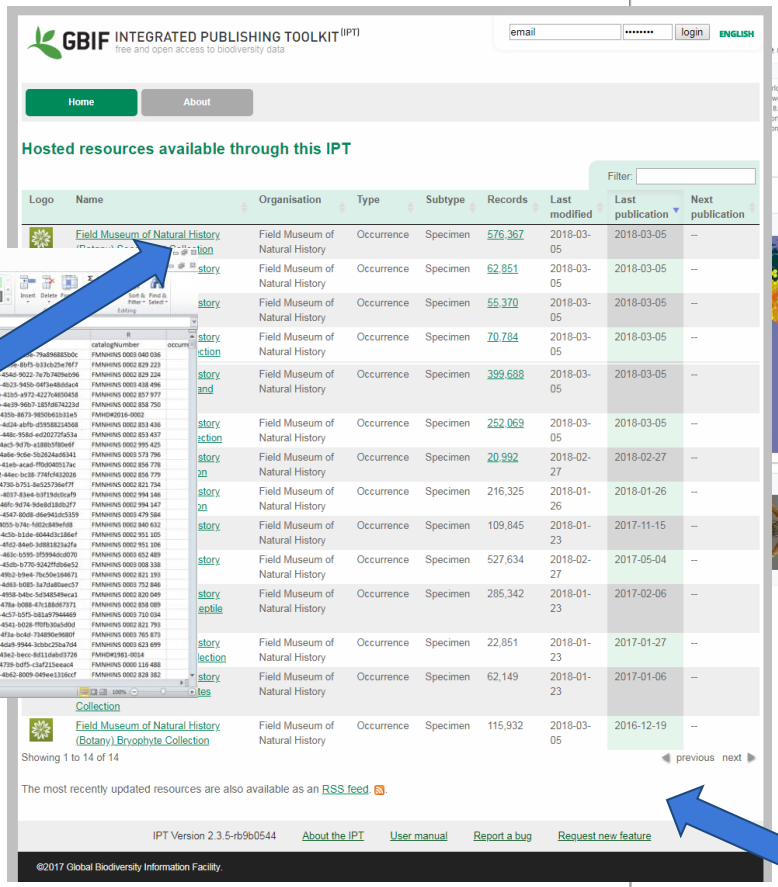
Filter:

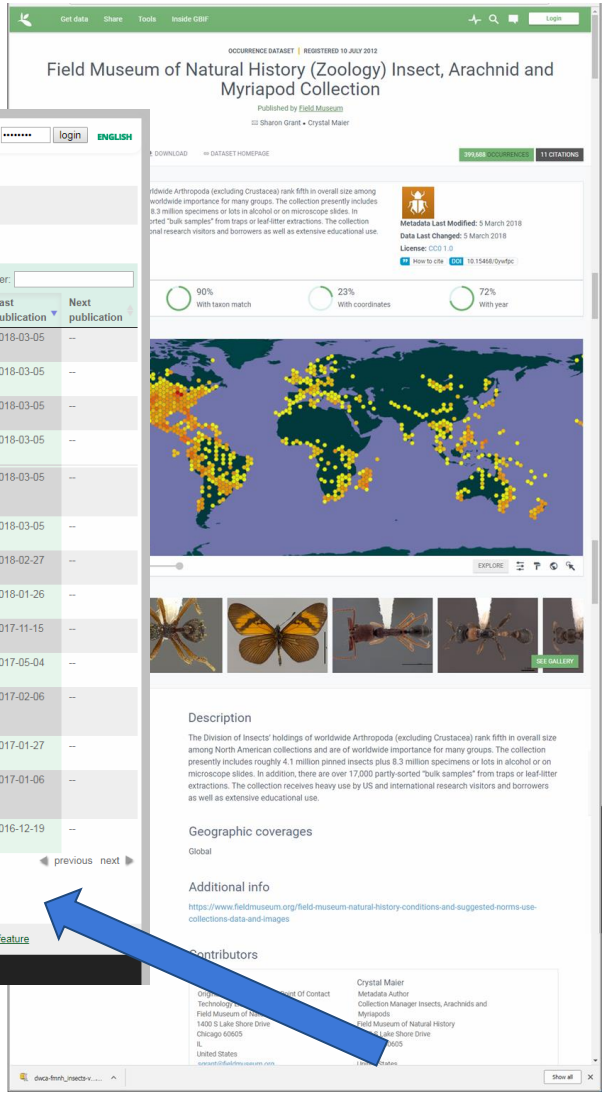
Logo	Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication
	Field Museum of Natural History (Botany) Seed Plant Collection	Field Museum of Natural History	Occurrence	Specimen	576,367	2018-03-05	2018-03-05	--
	Field Museum of Natural History (Botany) Bryophyte Collection	Field Museum of Natural History	Occurrence	Specimen	115,932	2018-03-05	2016-12-19	--
	Field Museum of Natural History (Botany) Fungi Collection	Field Museum of Natural History	Occurrence	Specimen	62,851	2018-03-05	2018-03-05	--
	Field Museum of Natural History (Botany) Lichen Collection	Field Museum of Natural History	Occurrence	Specimen	55,370	2018-03-05	2018-03-05	--
	Field Museum of Natural History (Botany) Pteridophyte Collection	Field Museum of Natural History	Occurrence	Specimen	70,784	2018-03-05	2018-03-05	--
	Field Museum of Natural History (Geology) Fossil Invertebrates Collection	Field Museum of Natural History	Occurrence	Specimen	62,149	2018-01-23	2017-01-06	--
	Field Museum of Natural History (Geology) Paleobotany Collection	Field Museum of Natural History	Occurrence	Specimen	22,851	2018-01-23	2017-01-27	--
	Field Museum of Natural History (Zoology) Amphibian and Reptile Collection	Field Museum of Natural History	Occurrence	Specimen	285,342	2018-01-23	2017-02-06	--
	Field Museum of Natural History (Zoology) Bird Collection	Field Museum of Natural History	Occurrence	Specimen	527,634	2018-02-27	2017-05-04	--
	Field Museum of Natural History (Zoology) Bird Egg Collection	Field Museum of Natural History	Occurrence	Specimen	20,992	2018-02-27	2018-02-27	--

Publishing

Data -> IPT -> GBIF







спасибо!

Questions ?