

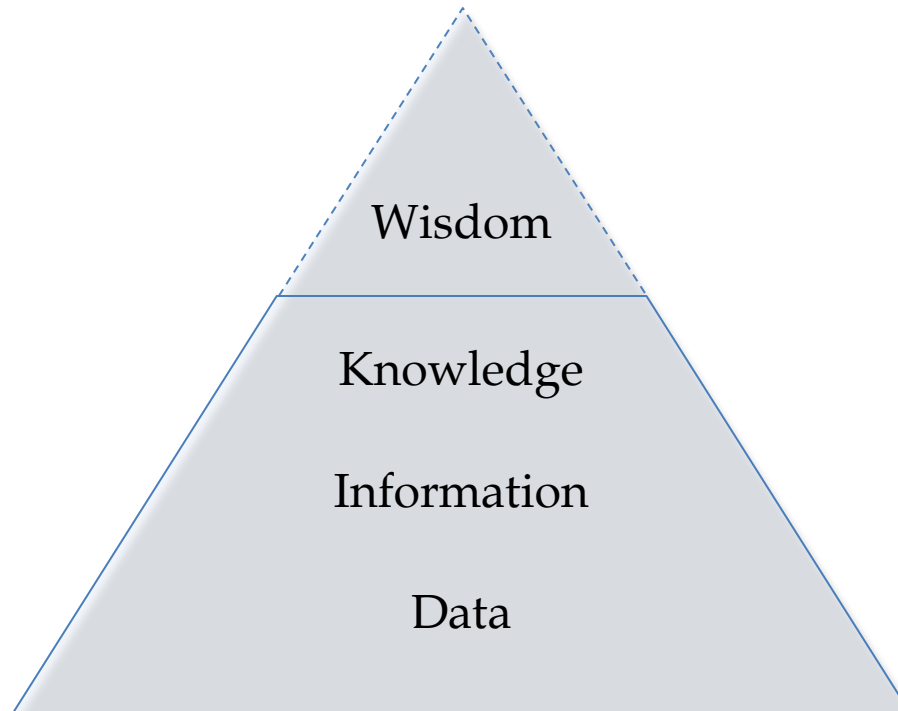
Managing Data, Information and Knowledge on C.AGB

Pierre Martin

Knowledge Management

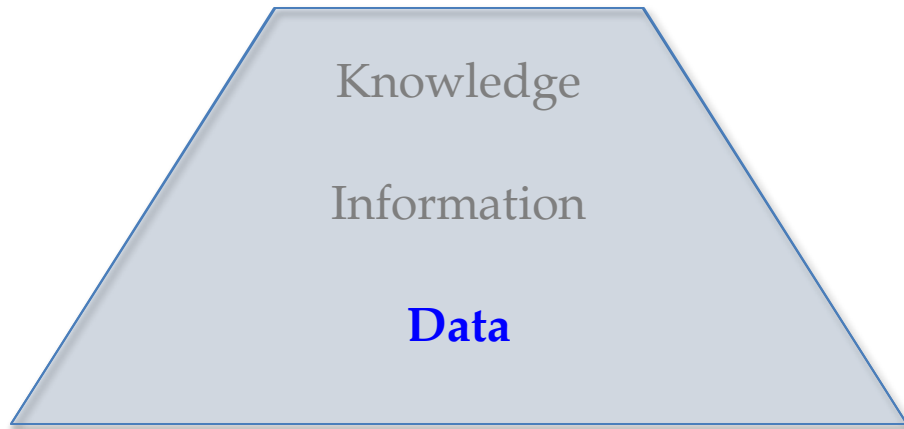
- What is it Discipline at the intersection
 between Management and Computer Science
- Objective Identify, create, represent, and distribute insights and experiences
 in an organization
- How Methodologies and tools enabling a formal representation of
 knowledge
- Application Organization monitoring, semantic web, collaborative learning...

The DIKW pyramid



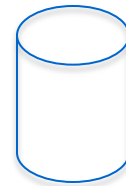
T.S.Eliot, 1934. The Rock (literature Nobel, 1948)

DIKW Description



Symbols or signals

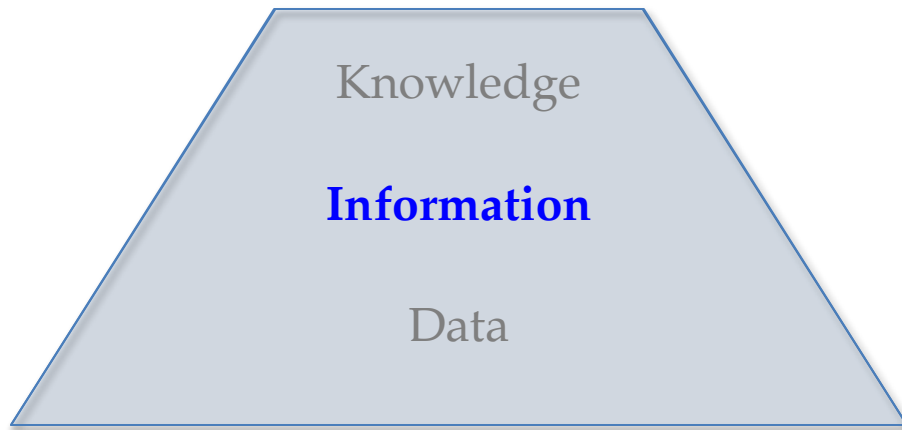
15



Dataset
=> Database

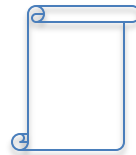
DIKW Description

Data that is given meaning by way of context

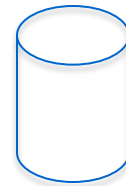


Room temperature

15



Metadata, Taxonomy ...
=> Information System



Dataset
=> Database

Metadata for the Ecological Sciences (1997)

Descriptors	Examples
Class I. Data set descriptors	
A. Data set identity	Title or theme of data set
B. Data set identification code	Database accession numbers or site-specific codes used to uniquely identify data set
C. Data set description	
1. Originator(s)	Names and addresses of principal investigator(s) associated with data set
2. Abstract	Descriptive abstract summarizing research objectives, data contents (including temporal, spatial, and thematic domain), context and potential uses of data set
D. Key words	Location (spatial scale), time period and sampling frequency (temporal scale), theme or contents (thematic scale)
Class II. Research origin descriptors	
A. "Overall" project description	[Note: this section may be essential if data set represents a component of a larger or more comprehensive database; otherwise, relevant items may be incorporated into II.B.]
1. Identity	Project title or theme
2. Originator(s)	Name(s) and address(es) of principal investigator(s) associated with project
3. Period of study	Date commenced, date terminated, or expected duration
4. Objectives	Scope and purpose of research program
5. Abstract	Descriptive abstract summarizing broader scientific scope of "overall" research project
6. Source(s) of funding	Grant and contract numbers, names and addresses of funding sources
B. "Specific subproject" description	
1. Site description	
a. Site type	Descriptive (e.g., short-grass prairie, blackwater stream, etc.)
b. Geography	Location (e.g., latitude/longitude), size
c. Habitat	Detailed characteristics of habitats sampled
d. Geology, landform	Soils, slope/elevation/aspect, terrain/physiography, geology/lithology
e. Watersheds, hydrology	Size, boundaries, receiving streams, etc.
f. Site history	Site management practices, disturbance history, etc.
g. Climate	Descriptive summary of site climatic characteristics
2. Experimental or sampling design	
a. Design characteristics	Description of statistical/sampling design
b. Permanent plots	Dimension, location, general vegetation characteristics (if applicable).
c. Data collection period, frequency, etc.	Information necessary to understand temporal sampling regime
3. Research methods	
a. Field/laboratory	Description or reference to standard field/laboratory methods
b. Instrumentation	Description and model/serial numbers
c. Taxonomy and systematics	References for taxonomic keys, identification and location of voucher specimens, etc.
d. Permit history	References to pertinent scientific and collecting permits
e. Legal/organizational requirements	Relevant laws, decision criteria, compliance standards, etc.
4. Project personnel	Principal and associated investigator(s), technicians, supervisors, students
Class III. Data set status and accessibility	
A. Status	
1. Latest update	Date of last modification of data set
2. Latest archive date	Date of last data set archival
3. Metadata status	Date of last metadata update and current status
4. Data verification	Status of data quality assurance checking
B. Accessibility	
1. Storage location and medium	Pointers to where data reside (including redundant archival sites)
2. Contact person(s)	Name, address, phone, fax, electronic mail
3. Copyright restrictions	Whether copyright restrictions prohibit use of all or portions of the data set
4. Proprietary restrictions	Any other restrictions that may prevent use of all or portions of data set
a. Release date	Date when proprietary restrictions expire
b. Citation	How data may be appropriately cited
c. Disclaimer(s)	Any disclaimers that should be acknowledged by secondary users
5. Costs	Costs associated with acquiring data (may vary by size of data request, desired medium, etc.)
Class IV. Data structural descriptors	
A. Data set file	
1. Identity	Unique file names or codes
2. Size	Number of records, record length, total number of bytes, etc.
3. Format and storage mode	File type (e.g., ASCII, binary, etc.), compression schemes employed (if any), etc.

Descriptors	Examples
4. Header information	Description of any header data or information attached to file [Note: may include elements related to "variable information" (IV.B.); if so, could be linked to appropriate section(s)]
5. Alphanumeric attributes	Mixed, upper, or lower case
6. Special characters/fields	Methods used to denote comments, "flag" modified or questionable data, etc.
7. Authentication procedures	Digital signature, checksum, actual subset(s) of data, and other techniques for assuring accurate transmission of data to secondary users
B. Variable information	
1. Variable identity	Unique variable name or code
2. Variable definition	Precise definition of variables in data set
3. Units of measurement	Units of measurement associated with each variable
4. Data type	
a. Storage type	Integer, floating point, character, string, etc.
b. List and definition of variable codes	Description of any codes associated with variables
c. Range for numeric values	Minimum, maximum
d. Missing value codes	Description of how missing values are represented in data set
e. Precision	Number of significant digits
5. Data format	
a. Fixed, variable length	Start column, end column
b. Columns	
c. Optional number of decimal places	
C. Data anomalies	Description of missing data, anomalous data, calibration errors, etc.
Class V. Supplemental descriptors	
A. Data acquisition	
1. Data forms or acquisition methods	Description or examples of data forms, automated data loggers, digitizing procedures, etc.
2. Location of completed data forms	
3. Data entry verification procedures	Procedures employed to verify that digital data set is error free
B. Quality assurance/quality control procedures	Identification and treatment of outliers, description of quality assessments, calibration of reference standards, equipment performance results, etc.
C. Related materials	References and locations of maps, photographs, videos, GIS data layers, physical specimens, field notebooks, comments, etc.
D. Computer programs and data-processing algorithms	Description or listing of any algorithms used in deriving, processing, or transforming data
E. Archiving	
1. Archival procedures	Description of how data are archived for long-term storage and access
2. Redundant archival sites	Locations and procedures followed
F. Publications and results	Electronic reprints, lists of publications resulting from or related to the study, graphical/statistical data representations, etc.
G. History of data set usage	
1. Data request history	Log of who requested data, for what purpose, and how data set was actually used
2. Data set update history	Description of any updates performed on data set
3. Review history	Last entry, last researcher review, etc.
4. Questions and comments from secondary users	Questionable or unusual data discovered by secondary users, limitations or problems encountered in specific applications of data, unresolved questions or comments

Aim:
towards a common description
of resources within a community

Metadata for the Ecological Sciences (1997)

Descriptors	Examples
Class I. Data set descriptors	
A. Data set identity	Title or theme of data set
B. Data set identification code	Database accession numbers or site-specific codes used to uniquely identify data set
C. Data set description	Information associated with data set contents (including temporal, spatial, and thematic domain), context and potential uses of data set
1. Originator(s)	Name(s) and address(es) of principal investigator(s) associated with project
2. Abstract	Summary of data set contents
D. Key words	Location (spatial scale), time period and sampling frequency (temporal scale), theme or contents (thematic scale)
Class II. Research origin descriptors	
A. "Overall" project description	[Note: this section may be essential if data set represents a component of a larger or more comprehensive database; otherwise, relevant items may be incorporated into II.B.]
1. Identity	Project title or theme
2. Originator(s)	Name(s) and address(es) of principal investigator(s) associated with project
3. Period of study	Date commenced, date terminated, or expected duration
4. Objectives	Goals or purposes of project
5. Abstract	Summary of "overall" research sources
6. Source(s) of funding	Identifying sources
B. "Specific subproject" description	
1. Site description	(e.g., etc.)
a. Site type	
b. Geography	
c. Habitat	
d. Geology, landform	Soils, slope/elevation/aspect, terrain/physiography, geology/lithology
e. Watersheds, hydrology	Size, boundaries, receiving streams, etc.
f. Site history	Site management practices, disturbance history, etc.
g. Climate	Descriptive summary of site climatic characteristics
2. Experimental or sampling design	Description of statistical/sampling design
a. Design characteristics	Dimension, location, general vegetation characteristics (if applicable).
b. Permanent plots	Information necessary to understand temporal sampling regime
c. Data collection period, frequency, etc.	
3. Research methods	Description or reference to standard field/laboratory methods
a. Field/laboratory	Description and model/serial numbers
b. Instrumentation	References for taxonomic keys, identification and location of voucher specimens, etc.
c. Taxonomy and systematics	References to pertinent scientific and collecting permits
d. Permit history	Relevant laws, decision criteria, compliance standards, etc.
e. Legal/organizational requirements	Principal and associated investigator(s), technicians, supervisors, students
4. Project personnel	
Class III. Data set status and accessibility	
A. Status	
1. Latest update	Date of last modification of data set
2. Latest archive date	Date of last data set archival
3. Metadata status	
4. Data verification	
B. Accessibility	
1. Storage location	(e.g., archival sites)
2. Contact person	
3. Copyright notice	Full or portions of the data set
4. Proprietary restrictions	Full or portions of data set
a. Release date	
b. Citation	How data may be appropriately cited
c. Disclaimer(s)	Any disclaimers that should be acknowledged by secondary users
5. Costs	Costs associated with acquiring data (may vary by size of data request, desired medium, etc.)
Class IV. Data structural descriptors	
A. Data set file	
1. Identity	Unique file names or codes
2. Size	Number of records, record length, total number of bytes, etc.
3. Format and storage mode	File type (e.g., ASCII, binary, etc.), compression schemes employed (if any), etc.

Data set descriptors

Research origin descriptors

Data set states and accessibility

Descriptors	Examples
4. Header information	Description of any header data or information attached to file [Note: may include elements related to "variable information" (IV.B.); if so, could be linked to appropriate section(s)]
5. Alphanumeric	
6. Special characters	
7. Authentication	
B. Variable information	
1. Variable identifier	
2. Variable definition	
3. Units of measurement	
4. Data type	
a. Storage type	Integer, floating point, character, string, etc.
b. List and definition of variable codes	Description of any codes associated with variables
c. Range for numeric values	Minimum, maximum
d. Missing value codes	Description of how missing values are represented in data set
e. Precision	Number of significant digits
5. Data format	
a. Fixed, variable length	
b. Columns	Start column, end column
c. Optional number of decimal places	
C. Data anomalies	Description of missing data, anomalous data, calibration errors, etc.
Class V. Supplemental descriptors	
A. Data acquisition	
1. Data forms or acquisition methods	Description or examples of data forms, automated data loggers, digitizing procedures, etc.
2. Location of completed data forms	
3. Data entry verification procedures	Procedures employed to verify that digital data set is error free
B. Quality assurance procedures	Procedures used for quality assessments, performance results, etc.
C. Related materials	Other data files, maps, GIS data layers, etc.
D. Computer program algorithms	Procedures used for deriving, processing, or analyzing data
E. Archiving procedures	Procedures used for storage and access
1. Archival procedures	
2. Redundant archival sites	Locations and procedures involved
F. Publications and results	Electronic reprints, lists of publications resulting from or related to the study, graphical/statistical data representations, etc.
G. History of data set usage	
1. Data request history	Log of who requested data, for what purpose, and how data set was actually used
2. Data set update history	Description of any updates performed on data set
3. Review history	Last entry, last researcher review, etc.
4. Questions and comments from secondary users	Questionable or unusual data discovered by secondary users, limitations or problems encountered in specific applications of data, unresolved questions or comments

Data structural descriptors

Supplementary descriptors

Aim: towards a common description of resources within a community

Essential Biodiversity Variables (2012)

GEO BON (Group on Earth Observation - Biodiversity Observation Network, www.earthobservations.org)

=> measurements required for study, reporting, and management of biodiversity change.

EXAMPLES OF CANDIDATE ESSENTIAL BIODIVERSITY VARIABLES					
EBV class	EBV examples	Measurement and scalability	Temporal sensitivity	Feasibility	Relevance for CBD targets and indicators (1,9)
Genetic composition	Allelic diversity	Genotypes of selected species (e.g., endangered, domesticated) at representative locations.	Generation time	Data available for many species and for several locations, but little global systematic sampling.	Targets: 12, 13. Indicators: Trends in genetic diversity of selected species and of domesticated animals and cultivated plants; RLI.
Species populations	Abundances and distributions	Counts or presence surveys for groups of species easy to monitor or important for ES, over an extensive network of sites, complemented with incidental data.	1 to >10 years	Standardized counts under way for some taxa but geographically restricted. Presence data collected for more taxa. Ongoing data integration efforts (Global Biodiversity Information Facility, Map of Life).	Targets: 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15. Indicators: LPI; WBI; RLI; population and extinction risk trends of target species, forest specialists in forests under restoration, and species that provide ES; trends in invasive alien species; trends in climatic impacts on populations.
Species traits	Phenology	Timing of leaf coloration by RS, with in situ validation.	1 year	Several ongoing initiatives (Phenological Eyes Network, PhenoCam, etc.)	Targets: 10, 15. Indicators: Trends in extent and rate of shifts of boundaries of vulnerable ecosystems.
Community composition	Taxonomic diversity	Consistent multitaxa surveys and metagenomics at select locations.	5 to >10 years	Ongoing at intensive monitoring sites (opportunities for expansion). Metagenomics and hyperspectral RS emerging.	Targets: 8, 10, 14. Indicators: Trends in condition and vulnerability of ecosystems; trends in climatic impacts on community composition.
Ecosystem structure	Habitat structure	RS of cover (or biomass) by height (or depth) globally or regionally.	1 to 5 years	Global terrestrial maps available with RS (e.g., Light Detection and Ranging). Marine and freshwater habitats mapped by combining RS and in situ data.	Targets: 5, 11, 14, 15. Indicators: Extent of forest and forest types; mangrove extent; seagrass extent; extent of habitats that provide carbon storage.
Ecosystem function	Nutrient retention	Nutrient output/input ratios measured at select locations. Combine with RS to model regionally.	1 year	Intensive monitoring sites exist for N saturation in acid-deposition areas and P retention in affected rivers.	Targets: 5, 8, 14. Indicators: Trends in delivery of multiple ES; trends in condition and vulnerability of ecosystems.

Essential Biodiversity Variables (2012)



- EBVs facilitate the harmonization of existing monitoring schemes and guide the implementation of new monitoring schemes, especially in gap areas where information on biodiversity change is still very sparse.
- Examples of essential variables are the allelic diversity of selected wild and domestic species, the population abundances for groups of species representative of some taxa (e.g. birds), the three-dimensional structure of habitats, and the nutrient retention rate in sensitive ecosystems
- EBVs are crucial for robust estimation of the indicators to assess progress towards the 2020 targets of the Convention on Biological Diversity. They can also provide the foundation for developing scenarios of the future of biodiversity under different policy and management options.

EBVs are sensitive to change over time

EBVs are focused on 'state' variables

EBVs are defined at a level of specificity intermediate between that of low-level (primary) observations and high-level indicators of biodiversity change

...

Taxonomy, Thesaurus, ...

Species
2000



Catalogue of Life: 18th March 2015

indexing the world's known species



English French Spanish Chinese Russian Portuguese Dutch German Polish Lithuanian Thai Vietnamese

- Browse
- Search
- Info

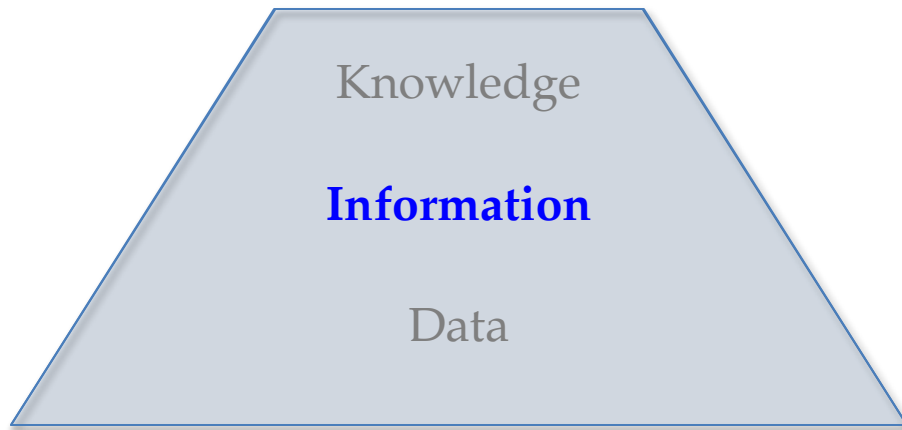
Browse taxonomic tree

Show statistics Show providers

- Animalia
- Archaea
- Bacteria
- Chromista
- Fungi
- Plantae
 - Phylum Bryophyta
 - Phylum Tracheophyta
 - Class Cycadopsida
 - Class Equisetopsida
 - Class Ginkgoopsida
 - Order Ginkgoales
 - Family Ginkgoaceae
 - Genus *Ginkgo*
 - Ginkgo biloba*
 - Class Gnetopsida
 - Class Liliopsida
 - Class Lycopodiopsida
 - Class Magnoliopsida
 - Class Marattiopsida
 - Class Pinopsida
 - Class Polypodiopsida
 - Class Psilotopsida
 - Protozoa

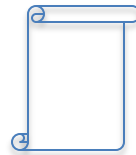
DIKW Description

Data that is given meaning by way of context

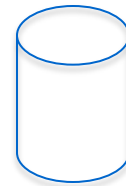


Room temperature

15



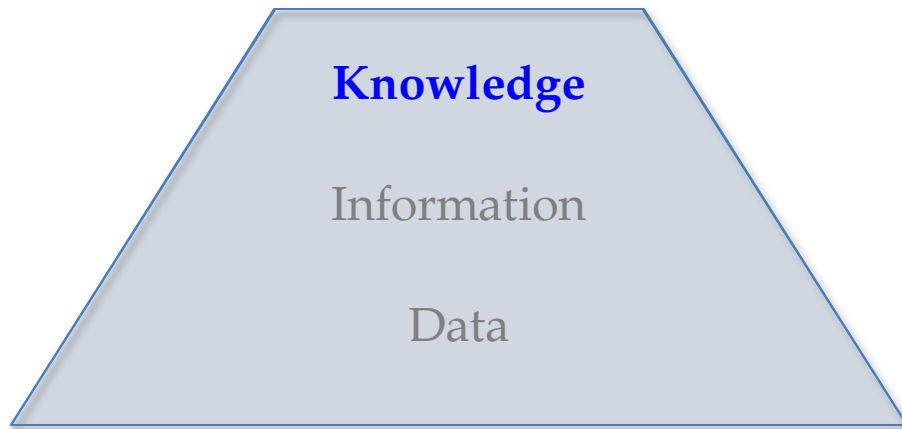
Metadata, Taxonomy ...
=> Information System



Dataset
=> Database

DIKW Description

Appropriation and interpretation of the information by human



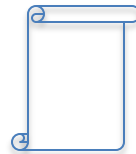
It's cold. To keep warm, up the heat

Room temperature

15



Semantic network ...
=> Knowledge based System

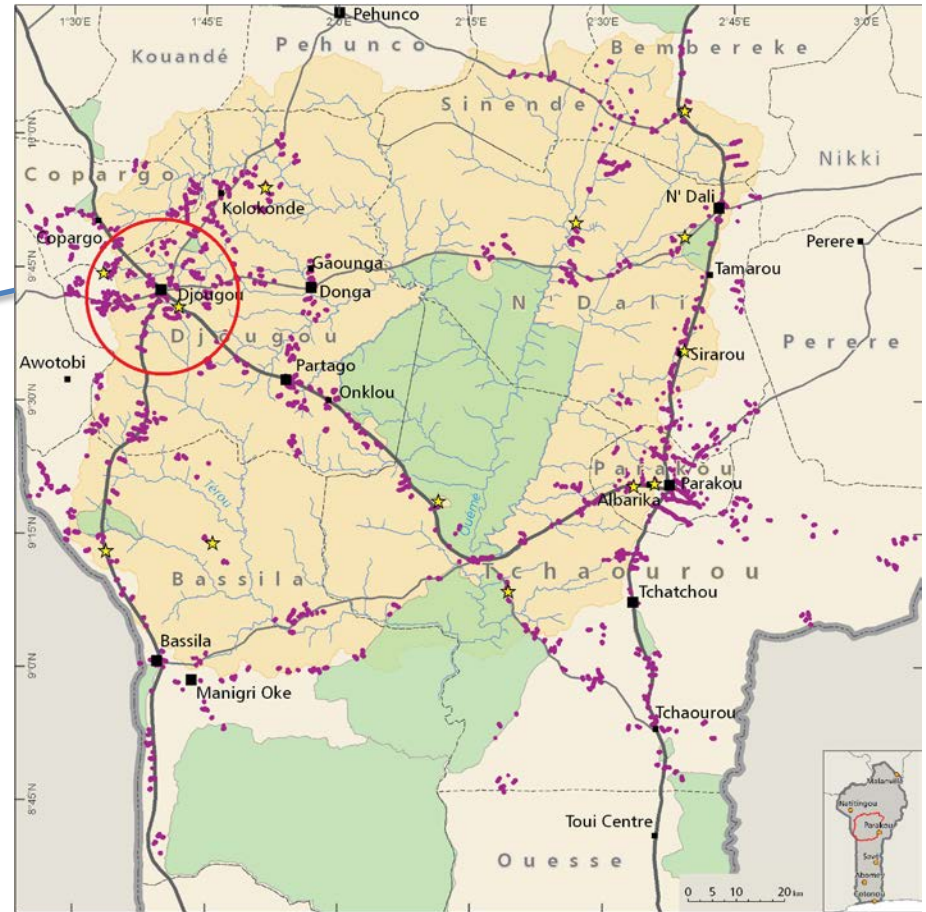
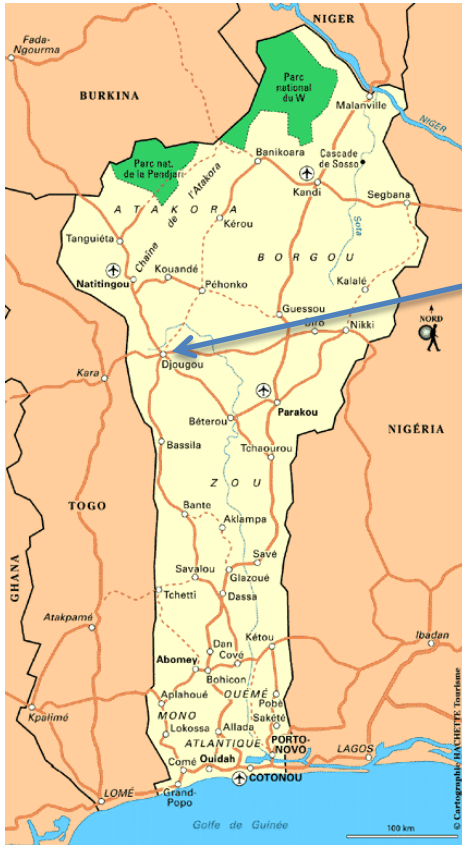


Metadata, Taxonomy ...
=> Information System



Dataset
=> Database

Case study 1 : Cereal stem borers in Pelebina



- | | | | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> ■ Bas-fond ★ Bas-fonds avec des investigations pedologiques — Cours d'eau ■ Forêt classée ■ Bassin versant de l'Ouémé supérieur | <p>Agglomération (habitants)</p> <ul style="list-style-type: none"> ■ 5000 - 10000 ■ > 10000 <p>Unité administrative</p> <ul style="list-style-type: none"> Limite d'Etat Limite de commune | <p>Réseau routier</p> <ul style="list-style-type: none"> Route principale Piste | <p>Sources des données :</p> <ul style="list-style-type: none"> Carte routière du Bénin, IGN 1998 Recensement, INSAE 2003 Enquête sur le terrain (IMPETUS) <p>Traitement et cartographie :</p> <ul style="list-style-type: none"> S. Giertz, G. Steup et H. Voigt |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|



For more informations

Silvie P., Martin P., Gutierrez A., Drieu R., Marnotte P. 2014.
Coupler savoirs académiques et méthodes informatiques pour
mieux orienter les observations de terrain. Proceedings of the X
international conference on Pests in Agriculture, Montpellier, 22-
23/10/2014, 435-444.

Publication in preparation

Case study 2 : NETSEED

Publication in preparation

Conclusion



Dealing with crop agrobiodiversity monitoring implies various objects and relationships to consider

Question is :

which representation and analysis tool
better fit your question/objective ?

But not the opposite ...