

The West African Data and Metadata Repository – a long-term data archive for ecological datasets from West Africa

Eva-Maria Gerstner¹, Yvonne Bachmann^{2,3}, Karen Hahn³, Anne Mette Lykke⁴, Marco Schmidt^{1,3,5}

¹ Senckenberg Biodiversity and Climate Research Centre (BiK-F), Data and Modelling Centre, Senckenberganlage 25, 60325 Frankfurt am Main, Germany.

² Darmstadt University of Technology, Center for Teacher Education, Alexanderstr. 6, 64283 Darmstadt, Germany.

³ Goethe University Frankfurt, Institute of Ecology, Evolution and Diversity, Max-von-Laue-Str. 13, 60438 Frankfurt am Main, Germany.

⁴ Department of Bioscience, Aarhus University, Vejlsovej 25, 8600 Silkeborg, Denmark

⁵ Senckenberg Research Institute and Natural History Museum Frankfurt, Botany and molecular evolution, Senckenberganlage 25, 60325 Frankfurt am Main, Germany.

Email: marco.schmidt@senckenberg.de

Summary: Although there is an increasing need for data in ecological studies, many datasets are still lost or not sufficiently visible due to a lack of appropriate data archives. With the West African Data and Metadata Repository, we present a secure long-term archive for a data-poor region allowing detailed documentation by metadata following the EML standard and giving data holders the opportunity to define levels of data access and conditions of use. This article gives an overview of structure, functions and content. The repository is online at the URL <http://westafricandata.senckenberg.de>.

Key words: West Africa, data archiving, ecological data

LE WEST AFRICAN DATA AND METADATA REPOSITORY - UNE ARCHIVE DE DONNÉES À LONG TERME POUR LES ENSEMBLES DE DONNÉES ÉCOLOGIQUES DE L'AFRIQUE DE L'OUEST

Résumé: Bien qu'il existe un besoin croissant de données dans les études écologiques, de nombreux ensembles de données sont encore perdues ou pas suffisamment visibles en raison d'un manque d'archives de données appropriées. Avec le West African Data and Metadata Repository, nous présentons une archive sécurisée à long terme pour une région pauvre en données permettant une documentation détaillée par des métadonnées suivant la norme EML et donnant aux propriétaires de données la possibilité de définir des niveaux de l'accès aux données et les conditions d'utilisation. Cet article donne un aperçu de la structure, des fonctions et du contenu. Le référentiel est en ligne à l'adresse URL <http://westafricandata.senckenberg.de>.

Mots clés: Afrique de l'Ouest, archivage des données, données écologiques

DAS WEST AFRICAN DATA AND METADATA REPOSITORY – EIN LANGZEITDATENARCHIV FÜR ÖKOLOGISCHE DATEN AUS WESTAFRIKA

Zusammenfassung: Obwohl zunehmend Daten für ökologische Studien gebraucht werden, gehen noch immer viele Daten verloren oder sind mangels geeigneter Datenarchive nicht sichtbar genug. Mit dem West African Data and Metadata Repository stellen wir ein Langzeitdatenarchiv für eine datenarme Region vor, das eine detaillierte Dokumentation mit Metadaten nach dem EML-Standard erlaubt und Datenhaltern die Möglichkeit gibt, Datenzugangsebenen und Nutzungsbedingungen zu bestimmen. Dieser Artikel gibt einen Überblick zu Struktur, Funktionen und Inhalt. Das Repositorium ist online unter der URL <http://westafricandata.senckenberg.de>.

Schlagworte: ökologische Daten, Datenarchivierung, Westafrika

1 INTRODUCTION

Plant ecology is continually moving into a direction of 'big data'. This has become possible by the increasing availability of large collaborative datasets such as species occurrence data from the Global Biodiversity Information Facility (GBIF; EDWARDS et al. 2000), nomenclature and synonymy datasets like the African Plant Database (KLOPPER et al. 2006) or plant trait data from TRY (KATTGE et al. 2011). Complementary to these large scale organismic data are continental or global environmental datasets for climate, soil, or satellite data and its derivatives, like vegetation indices or tree cover. Increased data availability, common standards and exchange formats have helped to bridge scientific disciplines and tackle urgent environmental issues from local to global scales more efficiently. Consequently, data-driven approaches become increasingly important in ecological and biodiversity studies (KELLING et al. 2009).

While the availability of biodiversity collection data, genetic sequences or environmental grid data have considerably increased, other biological data (which tend to be of a more complex and heterogeneous structure) and their digitization efforts and

availability lag behind. In consequence, this also means that datasets from ecological case studies not fitting easily into the schemes of larger databases are usually stored on scattered desktop computers or storage media instead of being properly archived. Consequently, they are more prone to be lost in the long-term and also less visible to the scientific community. This is even more pronounced in world regions that lack the infrastructure of the industrialized nations.

To improve this situation for ecological datasets from West Africa, we implemented the West African Data and Metadata Repository at the Senckenberg Biodiversity and Climate Research Center as a data warehouse for ecological datasets from West Africa. Long-term support of the database is ensured through integration in the wider structure of the BiK-F Data and Metadata Repository maintained at Senckenberg. This new repository will complement our set of biological databases for West Africa (West African Vegetation: SCHMIDT et al. 2012; African Plants – a photo guide: DRESSLER et al. 2014) and give researchers the opportunity of secure long-term archiving and increased visibility of their datasets, while keeping control of data access levels according to their needs.

2 TECHNICAL STRUCTURE

The West African Data and Metadata Repository (<http://westafricandata.senckenberg.de>, see Fig. 1) is based on Metacat (BERKLEY et al. 2001), a data warehouse for ecological and environmental data. Metacat is available under a free public license and allows easy entry and retrieval of data and metadata.

Technically, Metacat is a Java servlet application that runs in conjunction with a relational database such as in our case PostgreSQL and a Web server. Metadata are stored in EML (Ecological Metadata Language, FEGRAUS et al. 2005), an XML schema, which can be used in a modular and extensible manner to document ecological data. Especially data in tabular formats can be described in detail including row and column information. This way, Metacat guarantees a standardized comprehensive documentation of the data.

Metadata entry and editing is possible via a configurable web form or by using Morpho, a Java-based EML editor, that has explicitly been designed for this purpose (LEINFELDER et al. 2011). In addition to the metadata, the data tables themselves as well as further data entities can be optionally included (see Fig.2). This way, data users are able to fully understand the data sets. In addition, semantic web approaches in the ecological sciences are also supported.

Metacat also allows users to define different data access levels for each individual data set, including public (free access to all other users), private (no access to other users) and access restricted to selected registered users of the repository.

By linking Metacat repositories, a parallel search of data in different databases can be facilitated. Metacat is being used by large international data repositories such as the ones from the Knowledge Network for Biocomplexity (KNB), the National Center for Ecological Analysis and Synthesis (NCEAS), the Long-Term Ecological Research Network

(LTER) or the Data Observation Network for Earth (DataONE).

3 FEATURES

Our local Metacat installation has a specifically adapted interface for the West African Data and Metadata Repository. It consists of a semantic and geographic search tool, a data entry and editing form, and a web map service for visualization. Furthermore, via personal request to the data manager, each inserted dataset can be supplied with a digital object identifier (DOI), as we have registered our Metacat repository with DataCite.

A simple search is available by title, abstract, keywords, persons or by pre-defined frequently used terms of botany, geography and others (see Fig.2). Seven mandatory and seven additional optional EML categories with further sub-elements are supported by the data entry and editing form (Fig.3). The integrated web map service (WMS) is based on GeoServer and OpenLayers and allows a visualization of the center points or bounding boxes of data sets. A geographic search can be performed by clicking on a point location or by drawing a bounding box into the map (Fig.4).

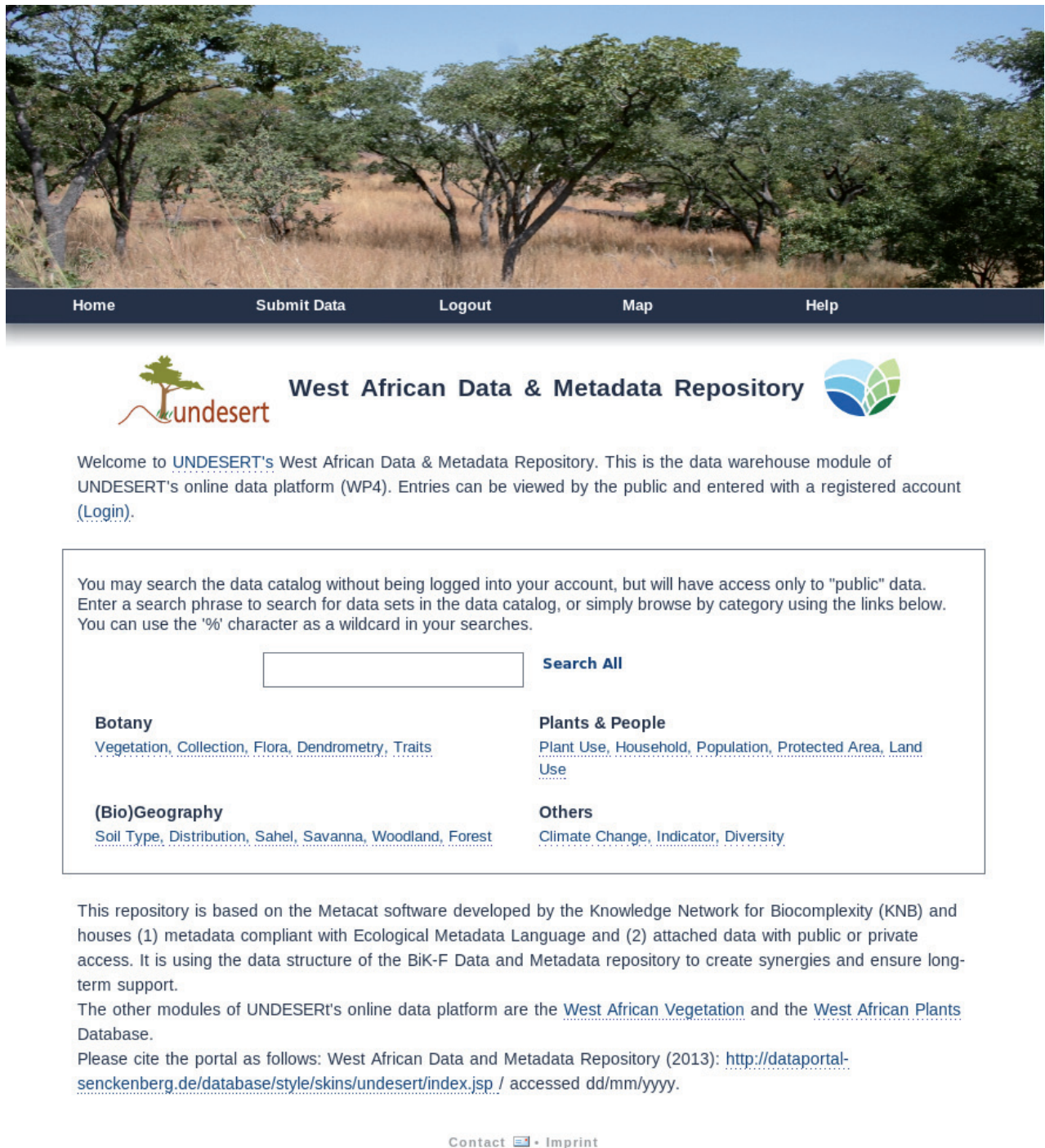
4 CONTENTS

Presently, the repository contains 50 ecological datasets from West Africa. These datasets are thematically diverse, dealing with land degradation, soil parameters, land use and land cover, environmental raster datasets, non-timber forest products, ethnobotanical interview data, species checklists and introduced species (Fig. 5).



Each single dataset is accompanied with metadata. This includes a citation comprising authors, title, document ID and a URL to the complete set of metadata. General information includes an abstract and keywords, upload date and eventually URLs to the data itself. Further sections of the metadata set cover involved parties (dataset owners, curators, contact persons, etc.), geographic coverage, temporal coverage, methods and information on data usage rights and the type of medium the data is distributed with. Datasets described with Morpho include even more detailed information on the data itself.

ACKNOWLEDGEMENTS

We acknowledge funding by the UNDESERT project "Understanding and combating desertification to mitigate its impact on ecosystem services" (EU FP7: 243906), which is funded by the European Commission, Directorate General for Research and Innovation, Environment Programme. Additionally, we acknowledge funding by the Biodiversity and Climate Research Center (BiK-F), part of the LOEWE program of the state of Hesse.



Home Submit Data Logout Map Help

 **West African Data & Metadata Repository** 

Welcome to [UNDESERT's](#) West African Data & Metadata Repository. This is the data warehouse module of UNDESERT's online data platform (WP4). Entries can be viewed by the public and entered with a registered account ([Login](#)).

You may search the data catalog without being logged into your account, but will have access only to "public" data. Enter a search phrase to search for data sets in the data catalog, or simply browse by category using the links below. You can use the '%' character as a wildcard in your searches.

Search All

Botany
[Vegetation](#), [Collection](#), [Flora](#), [Dendrometry](#), [Traits](#)

(Bio)Geography
[Soil Type](#), [Distribution](#), [Sahel](#), [Savanna](#), [Woodland](#), [Forest](#)

Plants & People
[Plant Use](#), [Household](#), [Population](#), [Protected Area](#), [Land Use](#)

Others
[Climate Change](#), [Indicator](#), [Diversity](#)

This repository is based on the Metacat software developed by the Knowledge Network for Biocomplexity (KNB) and houses (1) metadata compliant with Ecological Metadata Language and (2) attached data with public or private access. It is using the data structure of the BiK-F Data and Metadata repository to create synergies and ensure long-term support.

The other modules of UNDESERT's online data platform are the [West African Vegetation](#) and the [West African Plants Database](#).

Please cite the portal as follows: West African Data and Metadata Repository (2013): <http://dataportal-senckenberg.de/database/style/skins/undesert/index.jsp> / accessed dd/mm/yyyy.


Contact  • Imprint

Fig. 1: Start page of the West African data and Metadata Repository with text-based search functions. The repository can be accessed online via the URL <http://westafricandata.senckenberg.de>. / Page de démarrage du West African data and Metadata Repository avec des fonctions de recherche textuelle. L'archive peut être consulté en ligne via l'URL <http://westafricandata.senckenberg.de>.

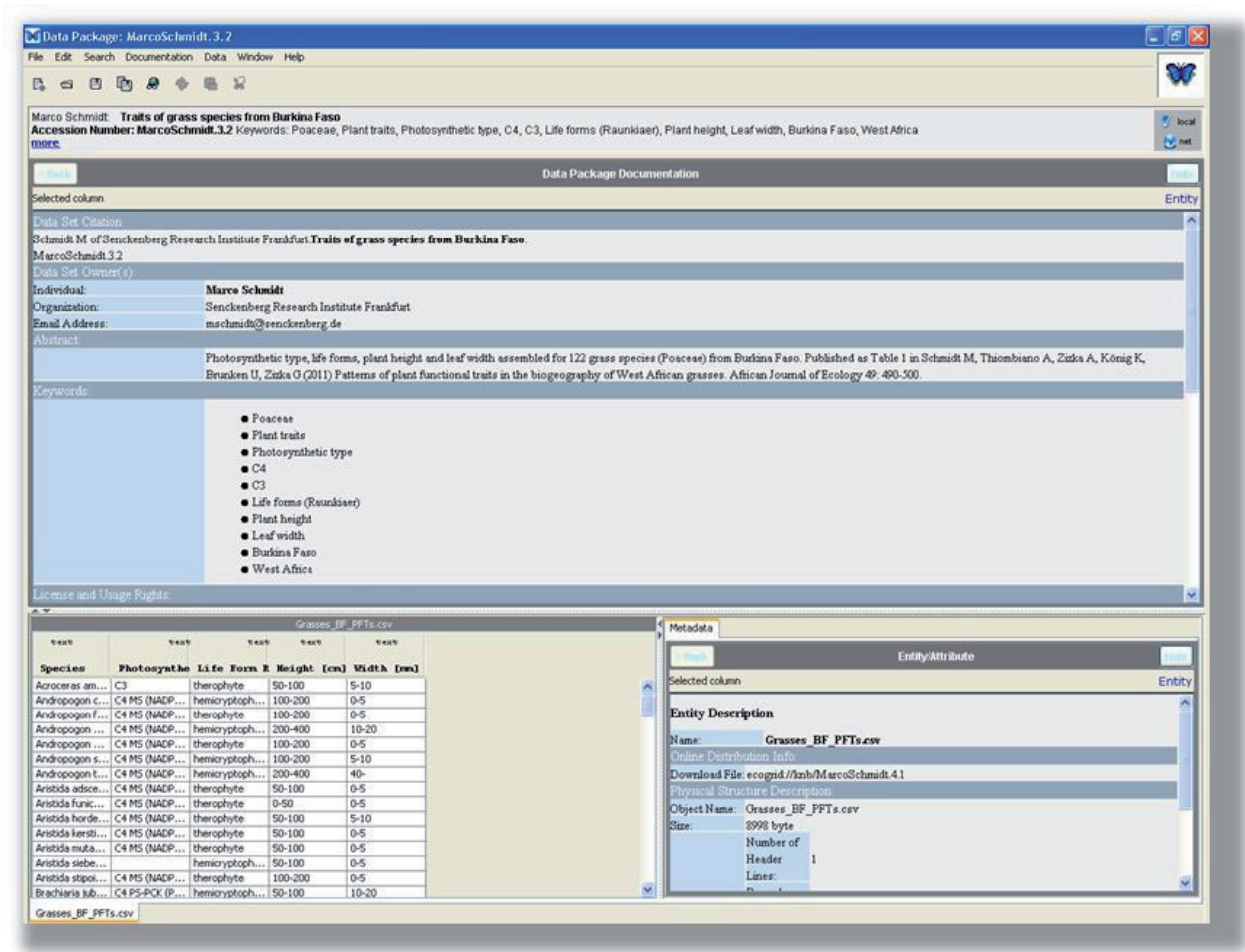



Fig. 2: Using Morpho to describe a tabular dataset in detail. / Utilisation de Morpho pour décrire un ensemble de données sous forme de tableau en détail.

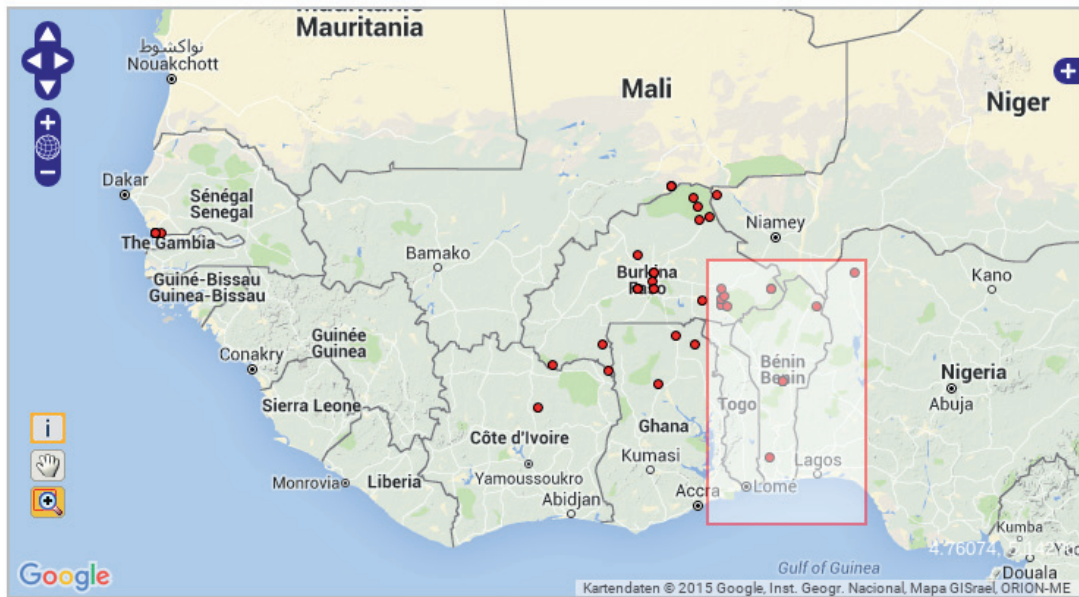
Use this form to edit the data set description that you submitted earlier to the repository.
 The ID of this data set is: undesert.88.3

*Denotes a required field.

NAME OF SUBMITTER <small>(What's this?)</small>	Show																									
BASIC INFORMATION <small>(What's this?)</small>	Hide																									
<p>*Data Set Title <input type="text" value="Introduced and invasive species of Burkina Faso"/></p> <p>(The title should include taxa, time period and location.)</p>																										
PRINCIPAL DATA SET OWNER <small>(What's this?)</small>	Show																									
ASSOCIATED PARTIES <small>(What's this?)</small>	Show																									
RESPONSIBLE PERSON IN THE LONGRUN (CONTACT) <small>(What's this?)</small>	Show																									
DATA SET ABSTRACT <small>(What's this?)</small>	Hide																									
<p>Data Set Abstract <small>(max. 350 words)</small> <input type="text" value="The list of introduced and invasive species has been extracted from information on distribution and origin of the"/></p> <p>The abstract should include the data type and the purpose of the data collection.</p>																										
KEYWORD INFORMATION <small>(What's this?)</small>	Show																									
TEMPORAL COVERAGE OF DATA <small>(What's this?)</small>	Show																									
TEMPORAL RESOLUTION OF DATA <small>(What's this?)</small>	Show																									
SPATIAL COVERAGE OF DATA <small>(What's this?)</small>	Hide																									
<p>*Geographic Description <input type="text" value="Burkina Faso"/></p> <p>General description of the geographic area in which the data were collected. It can be a simple place name (e.g., North Sea) or a fuller description.</p> <p>COORDINATES</p> <table border="0"> <tr> <td></td> <td>Degrees</td> <td>Minutes</td> <td>Seconds</td> <td></td> </tr> <tr> <td>*Latitude</td> <td><input type="text" value="9"/></td> <td><input type="text" value="0"/></td> <td><input type="text" value="0"/></td> <td><input checked="" type="radio"/> North <input type="radio"/> South</td> </tr> <tr> <td>*Longitude</td> <td><input type="text" value="6"/></td> <td><input type="text" value="0"/></td> <td><input type="text" value="0"/></td> <td><input checked="" type="radio"/> West <input type="radio"/> East</td> </tr> </table> <p>If only this first lat/long pair is entered, this indicates a point location. If both lat/long pairs are entered, then this first pair represents the northwest corner of a bounding box.</p> <table border="0"> <tr> <td>Latitude</td> <td><input type="text" value="15"/></td> <td><input type="text" value="0"/></td> <td><input type="text" value="0"/></td> <td><input checked="" type="radio"/> North <input type="radio"/> South</td> </tr> <tr> <td>Longitude</td> <td><input type="text" value="2"/></td> <td><input type="text" value="0"/></td> <td><input type="text" value="0"/></td> <td><input type="radio"/> West <input checked="" type="radio"/> East</td> </tr> </table> <p>If entered, this lat/long pair represents the southeast corner of a bounding box.</p>			Degrees	Minutes	Seconds		*Latitude	<input type="text" value="9"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input checked="" type="radio"/> North <input type="radio"/> South	*Longitude	<input type="text" value="6"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input checked="" type="radio"/> West <input type="radio"/> East	Latitude	<input type="text" value="15"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input checked="" type="radio"/> North <input type="radio"/> South	Longitude	<input type="text" value="2"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="radio"/> West <input checked="" type="radio"/> East
	Degrees	Minutes	Seconds																							
*Latitude	<input type="text" value="9"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input checked="" type="radio"/> North <input type="radio"/> South																						
*Longitude	<input type="text" value="6"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input checked="" type="radio"/> West <input type="radio"/> East																						
Latitude	<input type="text" value="15"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input checked="" type="radio"/> North <input type="radio"/> South																						
Longitude	<input type="text" value="2"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="radio"/> West <input checked="" type="radio"/> East																						
TAXONOMIC COVERAGE OF DATA <small>(What's this?)</small>	Show																									
DATA COLLECTION METHODS <small>(What's this?)</small>	Show																									
DISTRIBUTION INFORMATION <small>(What's this?)</small>	Show																									
UPLOAD DATA <small>(What's this?)</small>	Show																									
<input type="button" value="Submit Data Set Description"/>																										

Fig. 3: Data entry web form of the West African Data and Metadata Repository. / Forme de web du West African Data and Metadata Repository pour la saisie des données.

Click on the query icon  and draw a bounding box (with left mouse button) to select a region of special interest. You will be redirected to a result page with all available data sets within this region. Click [here](#) for bigger map view.



Click on the data points for feature information

Fig. 4: Map view of the West African Data and Metadata repository showing the locations of datasets in red and allowing direct access to the data descriptions by clicking single points or selecting a rectangular area (as shown in this screenshot around Benin). / Carte interactive du West African Data and Metadata Repository montrant les emplacements des ensembles de données en points rouges et permettant accès direct aux descriptions de données en cliquant sur les points ou en sélectionnant une zone rectangulaire (comme le montre cette capture d'écran).

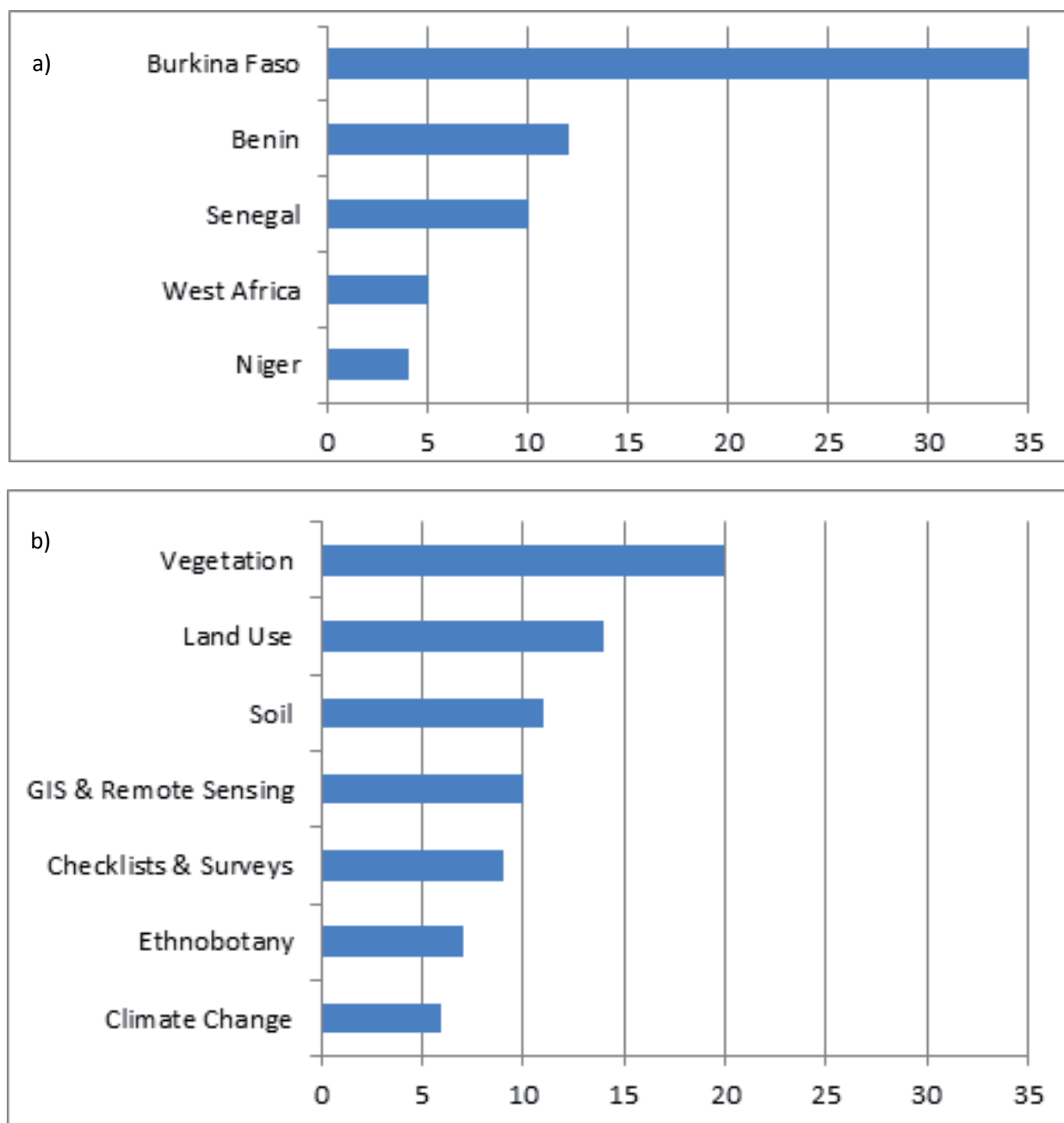


Fig. 5: Number of datasets summarized (a) by individual countries or West Africa as a region, and (b) by thematic fields. As a single dataset may cover several countries and themes, the sum exceeds the total number of datasets. / Nombre d'ensembles de données résumées (a) par des pays individuels ou l'Afrique de l'Ouest en tant que région, et (b) par des champs thématiques. Comme un seul ensemble de données peut couvrir plusieurs pays et thèmes, la somme dépasse le nombre total des ensembles de données.

REFERENCES

- BERKLEY C, JONES M, BOJILOVA J & HIGGINS D (2001): Metacat: A schema-independent XML database system. Proceedings of the 13th Intl. Conference on Scientific and Statistical Database Management: 171p.
- DRESSLER S, SCHMIDT M & ZIZKA G (2014): Introducing African Plants - a Photo Guide - an interactive identification tool for continental Africa. *Taxon* 63: 1159–1161.
- EDWARDS JL, LANE MA & NIELSEN ES (2000): Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. *Sci* 289 (5488): 2312–2314.

FEGRAS E, ANDELMAN S, JONES M & SCHILDHAUER M (2005): Maximizing the value of ecological data with structured metadata: An introduction to Ecological Metadata Language (EML) and principles for metadata creation. *Bull Ecol Soc America* 86: 158–168.

KATTGE J, DÍAZ S, LAVOREL S, PRENTICE IC, LEADLEY P, BÖNISCH G, GARNIER E, WESTOBY M, REICH PB, WRIGHT IJ, CORNELISSEN JHC, VIOLLE C, HARRISON SP, VAN BODEGOM PM, REICHSTEIN M, ENQUIST BJ, SOUDZILOVSKAIA NA, ACKERLY DD, ANAND M, ATKIN O, BAHN M, BAKER TR, BALDOCCHI D, BEKKER R, BLANCO CC, BLONDER B, BOND WJ, BRADSTOCK R, BUNKER DE, CASANOVES F, CAVENDER-BARES J, CHAMBERS JQ, CHAPIN FS, CHAVE J, COOMES D,

- CORNWELL WK, CRAINE JM, DOBRIN BH, DUARTE L, DURKA W, ELSER J, ESSER G, ESTIARTE M, FAGAN WF, FANG J, FERNÁNDEZ-MÉNDEZ F, FIDELIS A, FINEGAN B, FLORES O, FORD H, FRANK D, FRESCHET GT, FYLLAS NM, GALLAGHER RV, GREEN WA, GUTIERREZ AG, HICKLER T, HIGGINS S, HODGSON JG, JALILI A, JANSEN S, JOLY C, KERKHOFF AJ, KIRKUP D, KITAJIMA K, KLEYER M, KLOTZ S, KNOPS JMH, KRAMER K, KÜHN I, KUROKAWA H, LAUGHLIN D, LEE TD, LEISHMAN M, LENS F, LENZ T, LEWIS SL, LLOYD J, LLUSIÀ J, LOUAULT F, MA S, MAHECHA MD, MANNING P, MASSAD T, MEDLYN B, MESSIER J, MOLES AT, MÜLLER SC, NADROWSKI K, NAEEM S, NIINEMETS Ü, NÖLLERT S, NÜSKE A, OGAYA R, OLEKSYN J, ONIPCHENKO VG, ONODA Y, ORDOÑEZ J, OVERBECK G, OZINGA WA, PATIÑO S, PAULA S, PAUSAS JG, PEÑUELAS J, PHILLIPS OL, PILLAR V, POORTER H, POORTER L, POSCHLOD P, PRINZING A, PROULX R, RAMMIG A, REINSCH S, REU B, SACK L, SALGADO-NEGRET B, SARDANS J, SHIODERA S, SHIPLEY B, SIEFERT A, SOSINSKI E, SOUSSANA J-F, SWAINE E, SWENSON N, THOMPSON K, THORNTON P, WALDRAM M, WEIHER E, WHITE M, WHITE S, WRIGHT SJ, YGUEL B, ZAEHLE S, ZANNE AE & WIRTH C (2011): TRY - a global database of plant traits. *Global Change Biol* 17: 2905–2935.
- KELLING S, HOCHACHKA WM, FINK D, RIEDEWALD M, CARUANA R, BALLARD G & HOOKER G (2009): Data-intensive Science: A New Paradigm for Biodiversity Studies. *BioSci* 59: 613–620.
- KLOPPER RR, SMITH GF, GAUTIER L, CHATELAIN C & SPICHIGER R (2006): The African plant checklist and database project. *South African J Bot* 72: 342.
- LEINFELDER B, BOWERS S, O'BRIEN M, JONES MB & SCHILDHAUER M (2011): Using Semantic Metadata for Discovery and Integration of Heterogeneous Ecological Data. *Proceedings of the Environmental Information Management Conference (EIM 2011)* (eds M.B. Jones & C. Gries): 92–97pp.
- SCHMIDT M, JANSSEN T, DRESSLER S, HAHN K, HIEN M, KONATÉ S, LYKKE AM, MAHAMANE A, SAMBOU B, SINSIN B, THIOMBIANO A, WITTIG R & ZIZKA G (2012): The West African Vegetation Database. *Biodiversity Ecol* 4: 105–110.