Data and knowledge management in BIOTA AFRICA
— Current status and perspectives —

G. Muche[1], K.-H. Lampe[2] & G. Rambold[3]
[1]Biocentre Klein Flottbek and Botanical Garden – University of Hamburg, Germany
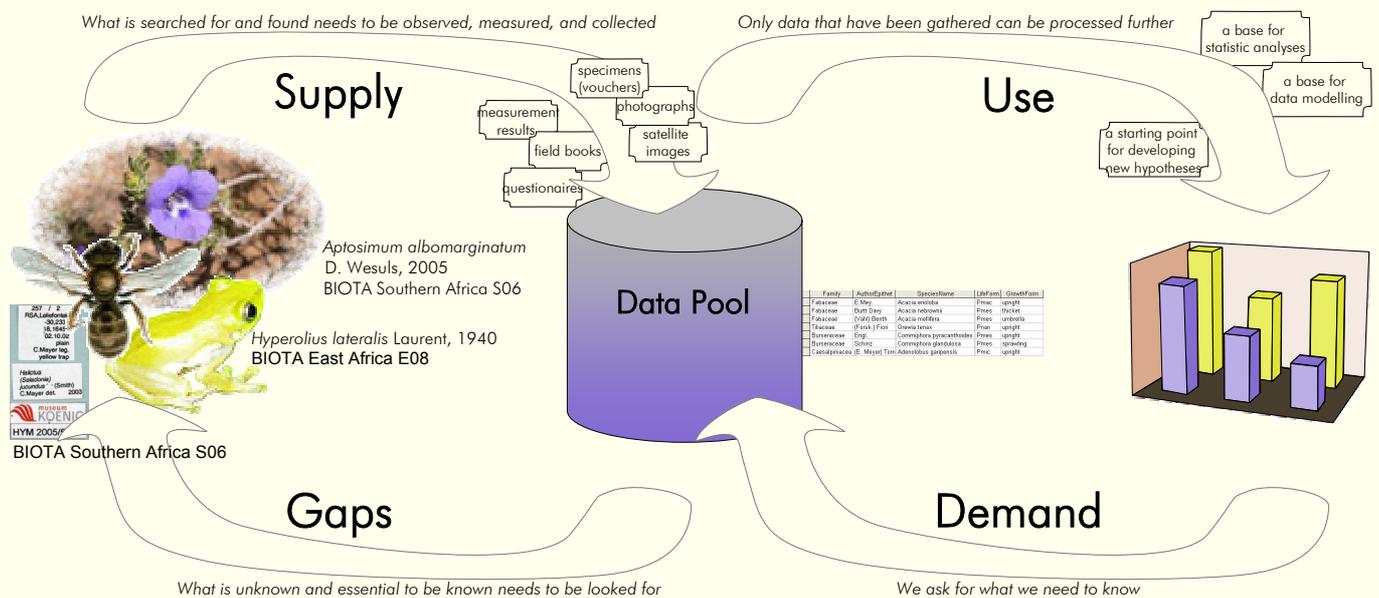[2]ZFMK Zoologisches Forschungsmuseum Alexander Koenig – Bonn, Germany
[3]Dept. of Mycology (LS Plant Systematics) – University of Bayreuth, Germany

BIOTA AFRICA aims to make valid scientific statements about the current status and future development of the biodiversity in hot spot areas of the African continent. For this ambitious objective it is necessary to achieve an absolutely reliable data foundation. Ongoing research activities create a high amount of information, being interesting not only for single working groups or disciplines, but for numerous user communities. Therefore data have to be managed under the aspect of sustainability and to be protected from any chance of loss. Knowledge management supports the process of collecting, refining and supplying of knowledge in BIOTA AFRICA..

Aims of data and knowledge management:
Data have to be usable for a wide circle of scientific users
Data must be a living stock for a long time

What is searched for and found needs to be observed, measured, and collected

Supply

specimens (vouchers), photographs, measurement results, field books, satellite images, questionaires

Aptosimum albomarginatum D. Wesuls, 2005 BIOTA Southern Africa S06

Hyperolius lateralis Laurent, 1940 BIOTA East Africa E08

BIOTA Southern Africa S06

Data Pool

Only data that have been gathered can be processed further

Use

a base for statistic analyses

a base for data modelling

a starting point for developing new hypotheses

Gaps

What is unknown and essential to be known needs to be looked for

Demand

We ask for what we need to know

Existing achievements in pooling BIOTA Africa research data

Since the start of the BIOTA project four years ago, a large amount of digital data have been gathered by the various subprojects. Due to the different kinds and levels of requirements within the various biological disciplines, these are data of relatively heterogeneous structure. Currently the amount of data is rapidly growing in quantity and an increase in quality can be registered. In a first step of data processing and joining, the data files in their different formats are collected and saved in a so-called data pool. For establishing a permanent BIOTA AFRICA data archive, the IT centre at Hamburg University provides sufficient storage capacities. Actually, the considerable part of the existing data pool, maintained by staff of the head office at the Botanical Institute of the University, is hosted there.

BIOTA AFRICA project participants and external co-workers who have signed the Data Sharing Protocol, are provided with original research data and metadata. According to the needs of the authors due to ongoing research activities, public access to the original data via Internet is partly restricted, and will be opened latest after the end of the respective subprojects. A first subset of monitoring data being stored in the BIOTABase database system and ready for open access, has been published recently via GBIF International. In addition, an internal forum and web pages were established for accessing interactive queries for project participant details and publication references.

At ZFMK African zoological voucher specimens from several BIOTA projects are collected and digitized with the BIODAT database. BIODAT data are available online via the BioCASE and GBIF portals. Beside that a field query tool allows easy access to fine-grained information about African specimens housed in ZFMK collections.

Further steps towards a sustainable BIOTA Africa IT-structure

Actually, a web interface for securely accessing a metadatabase providing information about the data pool contents is being implemented in cooperation with an ongoing DFG-funded project Forschergruppe 402: Functionality in a Tropical Mountain Rainforest (http://www.bergregenwald.de).

For the internal and external documentation of the specific terminologies, definitions and methods applied in the various disciplines or subprojects, a glossary of terms is set up as a Wikipedia-like platform in cooperation with the Federal Biological Research Centre for Agriculture and Forestry (BBA). This BIOTA AFRICA Wikipedia will allow registered participants to create and edit web page contents via web browser.

It is considered an essential step in a third project period to emend the structure of the existing data pool with its various types of data for later data transfer into a modularized centralized or decentralized interoperable database system (like the *Diversity Workbench* framework; http://www.diversitycampus.net). It has to be emphasized that all future system needs to follow TDWG standards (http://www.tdwg.org) and principles of the CIDOC CRM (http://cidoc.ics.forth.gr). A framework of database client modules, some being already implemented in subproject BIOTA Southern Africa S04, would enable data authors and users to remotely access, edit and query the contents of the modularized database components.

It has to be considered, that such database pools on the long run are best located at Natural History Collections, being so-called Memory Institutions. In the future, data exchange between the subprojects will be intensified. According to the Data Sharing Protocol, database mirrors will be installed at the African partner organisations.

BIOLOG

Projektträger im DLR

Federal Ministry of Education and Research