Central database design and data synthesis in the DFG Research Unit FOR 1246 – Kilimanjaro ecosystems under global change (KiLi)

Jie Zhang, Katrin Böhning-Gaese, Markus Fischer, Andreas Hemp, Thomas Nauss, Marcell K. Peters, Ingolf Steffan-Dewenter

GFÖ 2012 – Data publishing, linking, using ecological data online
Outline

- KiLi project introduction
- KiLi central database design
  - Hopes fulfilled
  - Fears conquered
  - Challenge accepted
  - Future Outlook
Outline

- KiLi project introduction

- KiLi central database design
  - Hopes fulfilled
  - Fears conquered
  - Challenge accepted
  - Future Outlook
KiLi – Kilimanjaro ecosystems under global change

Global climate and land use change

Regional/local drivers
- Temperature
- Radiation
- Precipitation
- Fog/Clouds

Biogeochemical ecosystem processes
- Carbon pools and fluxes
- Water pools and fluxes
- Nutrient pools and fluxes

Biotic ecosystem components
- Primary productivity
- Plants
- Leaf litter
- Decomposers
- Predator, parasitoids
- Herbivores
- Pollinators
- Seed dispersers

Ecosystem functions and services
- Climate regulation
- Carbon storage
- Water supply
- Soil fertility
- Agricultural livelihood
- Crop pollination
- Pest control
- Genetic resources
- Resilience/adaptation

Data
- SP 1
- SP 2
- SP 3
- SP 4
- SP 5
- SP 6
- SP 7
- AP

Synthesis
60 study sites, 12 land cover types, 3700 m altitude, 54 km longitude
Demand for central database

Regional GIS data
- Climate
- Topography
- Soils
- GPP
- Land cover
- Fires
- Soils
- Species ranges

Primary plot data
- Climate
- Water/carbon
- Nutrients
- Productivity
- Richness
- Abundance
- Interactions
- Genetics

Spatial scales: plot, subplot, sample

Species trait data
- Phenology
- Longevity
- Size
- Ecophysiology
- Pollination
- Reproduction
- Dispersal
- Regeneration

Taxonomic databases (plants, animals)

Measurements and external data

Multiple time layers

Observations and experiments on joint study sites

Location

Species

Specific metadata

Measurements and external data
Outline

- KiLi project introduction

- KiLi central database design
  - Hopes fulfilled
  - Fears conquered
  - Challenges accepted
  - Future Outlook
### Hopes

- Data management online
- Data sharing policy
- Long term storage
- Data synthesis
- Data linking
- Project activities
- Communications

### Fears

- Time effort
- Data control
- Versioning
- Flexibility
- User friendly
- Acknowledgement

### Challenges

- Data linking within different platforms
- Permanent data storage
Outline

- KiLi project introduction
- KiLi central database design
  - Hopes fulfilled
  - Fears conquered
  - Challenge accepted
  - Future Outlook
<table>
<thead>
<tr>
<th><strong>Hopes</strong></th>
<th><strong>Fears</strong></th>
<th><strong>Challenges</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Data management online</td>
<td>I can edit my data any time</td>
<td>I can edit my data any time</td>
</tr>
<tr>
<td>Data sharing policy</td>
<td>I can delete my data any time</td>
<td>I can delete my data any time</td>
</tr>
<tr>
<td>Long term storage</td>
<td>history of data use</td>
<td>history of data use</td>
</tr>
<tr>
<td>Data synthesis</td>
<td>contact if someone wants to use my data</td>
<td>contact if someone wants to use my data</td>
</tr>
<tr>
<td>Data linking</td>
<td>guaranteed long-term-care of database</td>
<td>guaranteed long-term-care of database</td>
</tr>
<tr>
<td>Project activities</td>
<td>guaranteed data quality</td>
<td>guaranteed data quality</td>
</tr>
<tr>
<td>Communications</td>
<td>user friendly data upload</td>
<td>user friendly data upload</td>
</tr>
<tr>
<td></td>
<td>clear guidelines for reuse of data</td>
<td>clear guidelines for reuse of data</td>
</tr>
</tbody>
</table>

(Enke, 2011)
KiLi central database construction

- Rapidly reduced time effort
- Improve platform compatibility and data linking
- Knowledge sharing and exchange

“If I have been able to see further, it was only because I stood on the shoulders of giants.” Isaac Newton
KiLi central database architecture

KiLi Info-System

DB2 Database

Website

Research Data

Application Data

Project data Documents

Data sharing Upload/download

Station booking

Online GIS

User Management

Communication

Website

IO

ClimateData

XmlData

XmlDataAcces

Fieldbook

Provider

Forum

GFÖ 2012
Data publishing, linking, using ecological data online
KiLi central database architecture

Website

KiLi Info-System

DB2 Database

Project data Documents
Data sharing Upload/download
Station booking
Online GIS
User Management
Communication

Website IO ClimateData XmlData XmlDataAcces Fieldbook Provider Forum

Research Data
Application Data
KiLi central database architecture
KiLi central database architecture

KiLi Info-System

DB2 Database

Website

Research Data

Application Data

Project data Documents
Data sharing Upload/download
Station booking
Online GIS
User Management
Communication

Website
IO
ClimateData
XmlData
XmlDataAcces
Fieldbook
Provider
Forum

Research Group Kilimanjaro

Data publishing, linking, using ecological data online

GFÖ 2012
KiLi central database architecture

Regional GIS data → Location → Primary plot data

Primary plot data → Species → Species trait data

Standardized Metadata (EML based)
KiLi central database architecture

Regional GIS data \(\xrightarrow{\text{Location}}\) Primary plot data \(\xrightarrow{\text{Species}}\) Species trait data

Standardized Metadata (EML based)

GFÖ 2012
Data publishing, linking, using ecological data online
KiLi central database architecture

Regional GIS data → Location → Primary plot data → Species → Species trait data

Standardized Metadata (EML based)

XSLT

GFÖ 2012
Data publishing, linking, using ecological data online
KiLi central database architecture

Regional GIS data -> Location -> Primary plot data -> Species -> Species trait data

Standardized Metadata (EML based)

GFÖ 2012
Data publishing, linking, using ecological data online
KiLi central database architecture

Regional GIS data \[\rightarrow\] Location \[\rightarrow\] Primary plot data \[\rightarrow\] Species \[\rightarrow\] Species trait data

Standardized Metadata (EML based)

Recycled

XSLT

GFÖ 2012
Data publishing, linking, using ecological data online
Remote sensing images, vector geospatial data

Sensor data, experiment data

Processed data

Primary plot data

Species trait data

Regional GIS data

Location

Species

Standardized Metadata (EML based)

Unstructured data

Tabular table

XML table

Geoserver

serialize
deserialize

XML Shredding

GFÖ 2012
Data publishing, linking, using ecological data online
Outline

- KiLi project introduction
- KiLi central database design
  - Hopes fulfilled
  - Fears conquered
  - Challenge accepted
  - Future Outlook
<table>
<thead>
<tr>
<th>Hopes</th>
<th>Fears</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time effort</td>
<td>Hopes</td>
<td>Fears</td>
</tr>
<tr>
<td>Data control</td>
<td>versioning</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>Versioning</td>
<td>Flexibility</td>
<td></td>
</tr>
<tr>
<td>Flexibility</td>
<td>User friendly</td>
<td></td>
</tr>
<tr>
<td>User friendly</td>
<td>Acknowledgement</td>
<td></td>
</tr>
</tbody>
</table>

(Enke, 2011)
<table>
<thead>
<tr>
<th>Fears</th>
<th>KiLi solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time effort</td>
<td>Metadata recycle; Data structure defined in metadata</td>
</tr>
<tr>
<td>Data control</td>
<td>Author – Subproject – Request for access</td>
</tr>
<tr>
<td>Versioning</td>
<td>Online updating; Quality flag (raw, processed, validated); version number: 1(dataset).2(observation).3(metadata)</td>
</tr>
<tr>
<td>Flexibility</td>
<td>XML flexible data structure; Various formats upload/download; Support of unstructured data</td>
</tr>
<tr>
<td>User friendly</td>
<td>Keyword search; Dataset browse based on Ontology</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>Co-authorship, acknowledgements</td>
</tr>
</tbody>
</table>
Outline

- KiLi project introduction
- KiLi central database design
  - Hopes fulfilled
  - Fears conquered
  - Challenge accepted
  - Future Outlook
<table>
<thead>
<tr>
<th>Hopes</th>
<th>Fears</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data management online</td>
<td>Time effort</td>
<td>Data linking within different platforms</td>
</tr>
<tr>
<td>Data sharing policy</td>
<td>Data control</td>
<td>Permanent data storage</td>
</tr>
<tr>
<td>Long term storage</td>
<td>Versioning</td>
<td></td>
</tr>
<tr>
<td>Data synthesis</td>
<td>Flexibility</td>
<td></td>
</tr>
<tr>
<td>Data linking</td>
<td>User friendly</td>
<td></td>
</tr>
<tr>
<td>Project activities</td>
<td>Acknowledgement</td>
<td></td>
</tr>
<tr>
<td>Communications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hopes</td>
<td>Fears</td>
<td>Challenges</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>Data linking within different platforms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Metadata standard</td>
<td>✓ EML</td>
<td>✓ Cooperation between projects</td>
</tr>
<tr>
<td>• Semantic search</td>
<td>✓ Ontology</td>
<td></td>
</tr>
<tr>
<td>• Online access</td>
<td>✓ Link of web interface</td>
<td>✓ Central homepage for data integration</td>
</tr>
</tbody>
</table>

© 2012
Data publishing, linking, using ecological data online
Data linking within different platforms

- Transformation of data structure  ✓ XML table

Observations

Metadata

Dataset

Hopes

Fears

Challenges
Store data permanently

- Journals (e.g. Ecological Archives) -> permanent, limited
- Project archives -> long term, not permanent
- Combine of existing data centers (universities, libraries, institutes)
Outline

- KiLi project introduction

- KiLi central database design
  - Hopes fulfilled
  - Fears conquered
  - Challenge accepted
  - Future Outlook
KiLi Central Database

Regional GIS data
- Measurements and external data

Primary plot data
- Observations and experiments on joint study sites

Species trait data
- Measurements and external data

Plot ID

Species ID

Online Geo-processing

Semantic Search Ontology

Online Data Analysis Visualization

Data Synthesis

Scientific Workflow
Thank you for your attention!

www.kilimanjaro.biozentrum.uni-wuerzburg.de