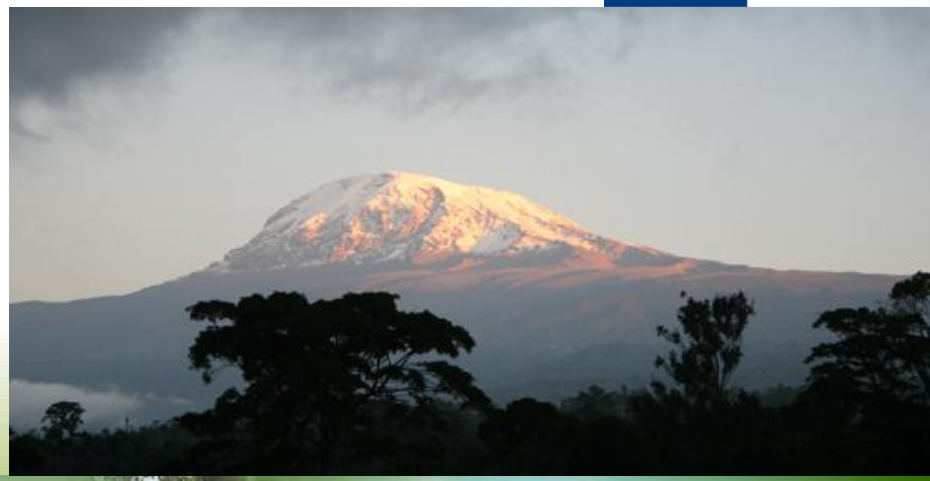
Central database design and data synthesis in the DFG Research Unit FOR 1246 – Kilimanjaro ecosystems under global change (KiLi)

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

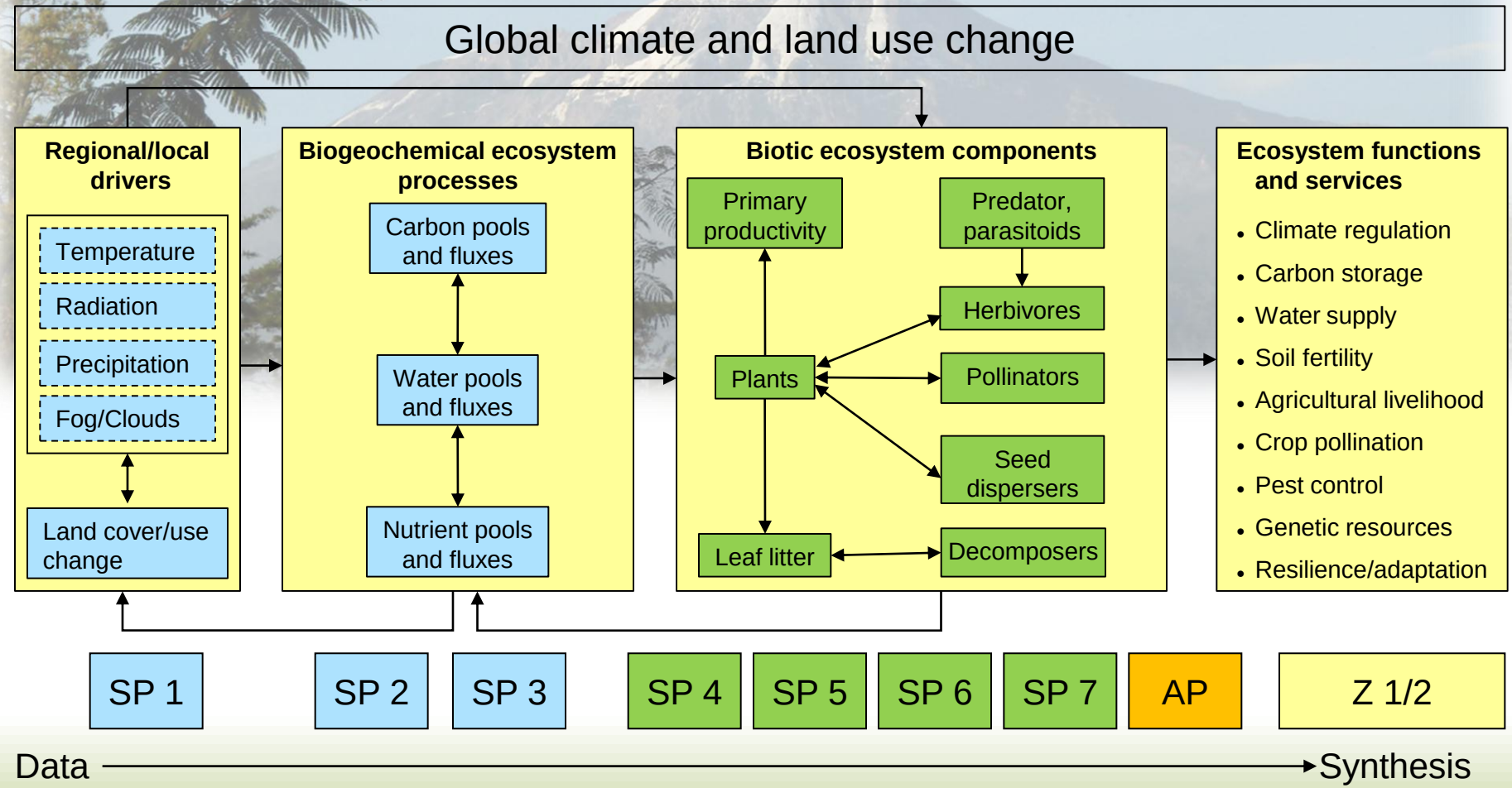
Outline

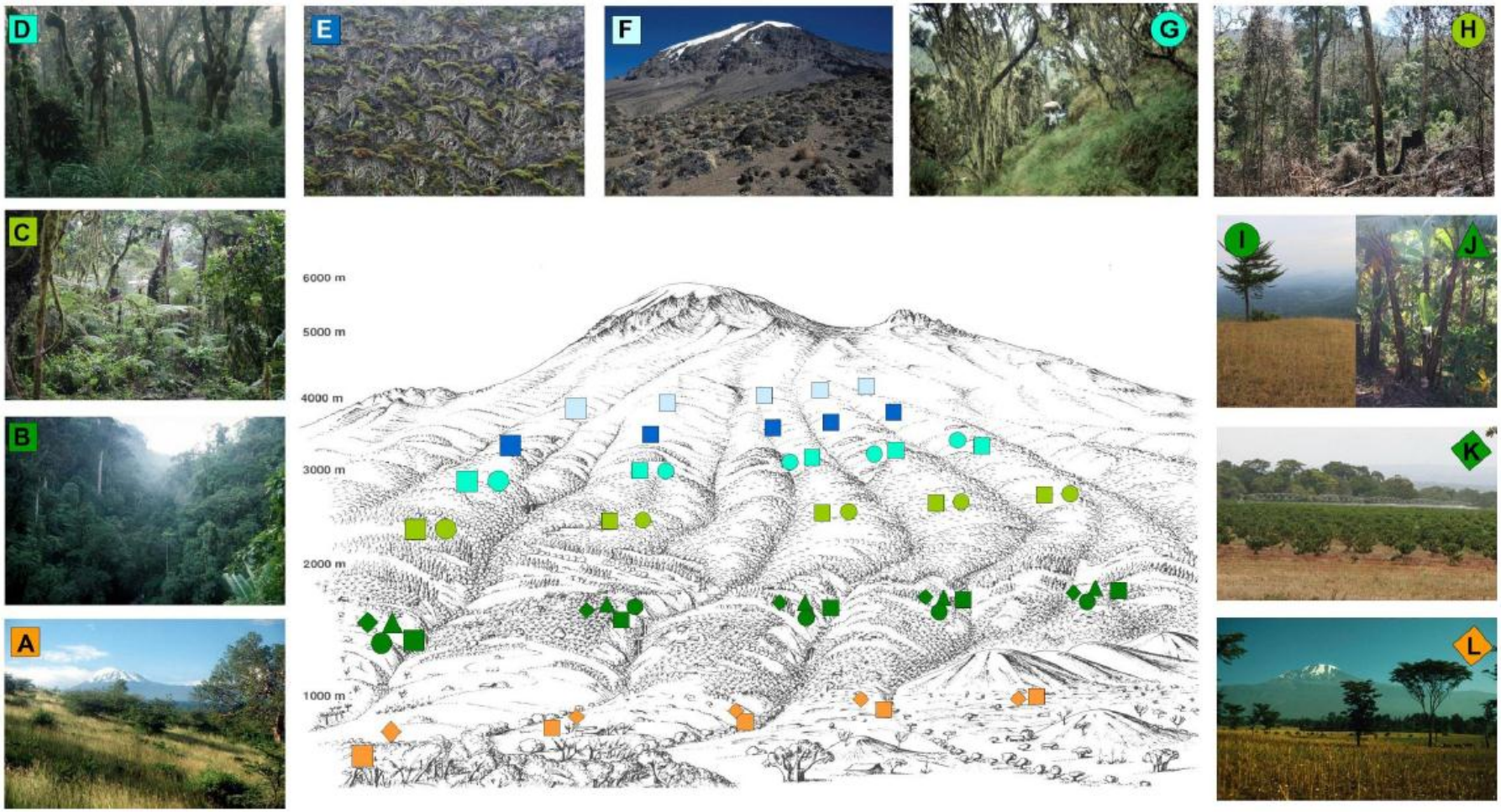
- KiLi project introduction
- KiLi central database design
  - ✓ Hopes fulfilled
  - ✓ Fears conquered
  - ✓ Challenge accepted
  - ✓ Future Outlook

Outline

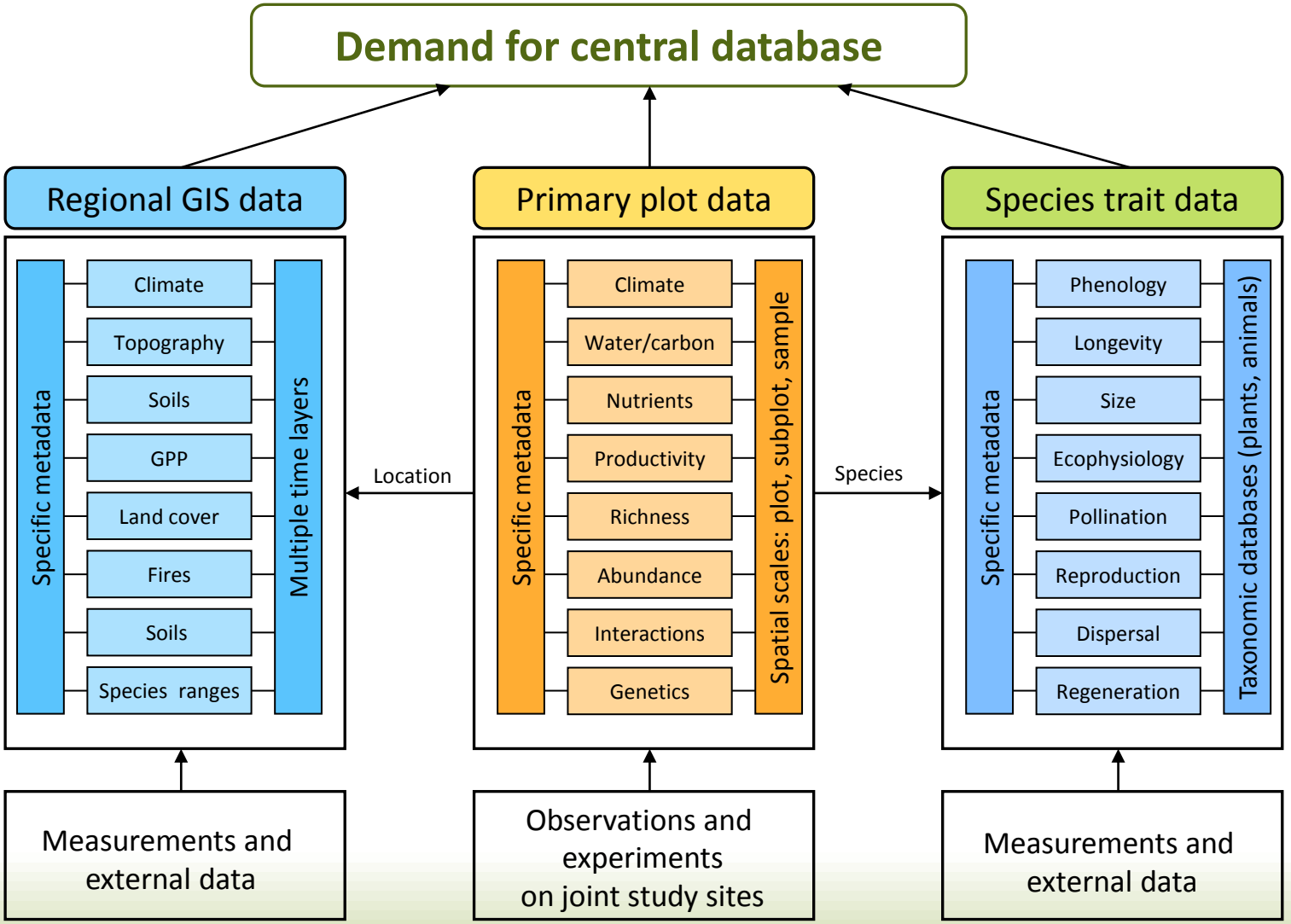
- KiLi project introduction
- KiLi central database design
  - ✓ Hopes fulfilled
  - ✓ Fears conquered
  - ✓ Challenge accepted
  - ✓ Future Outlook

KiLi – Kilimanjaro ecosystems under global change





60 study sites, 12 land cover types, 3700 m altitude, 54 km longitude


Demand for central database
Regional GIS data
Primary plot data
Species trait data
Specific metadata
Climate
Topography
Soils
GPP
Land cover
Fires
Soils
Species ranges
Multiple time layers
Location
Specific metadata
Climate
Water/carbon
Nutrients
Productivity
Richness
Abundance
Interactions
Genetics
Spatial scales: plot, subplot, sample
Species
Specific metadata
Phenology
Longevity
Size
Ecophysiology
Pollination
Reproduction
Dispersal
Regeneration
Taxonomic databases (plants, animals)
Measurements and external data
Observations and experiments on joint study sites
Measurements and external data

Outline

- KiLi project introduction
- KiLi central database design
  - ✓ Hopes fulfilled
  - ✓ Fears conquered
  - ✓ Challenges accepted
  - ✓ Future Outlook

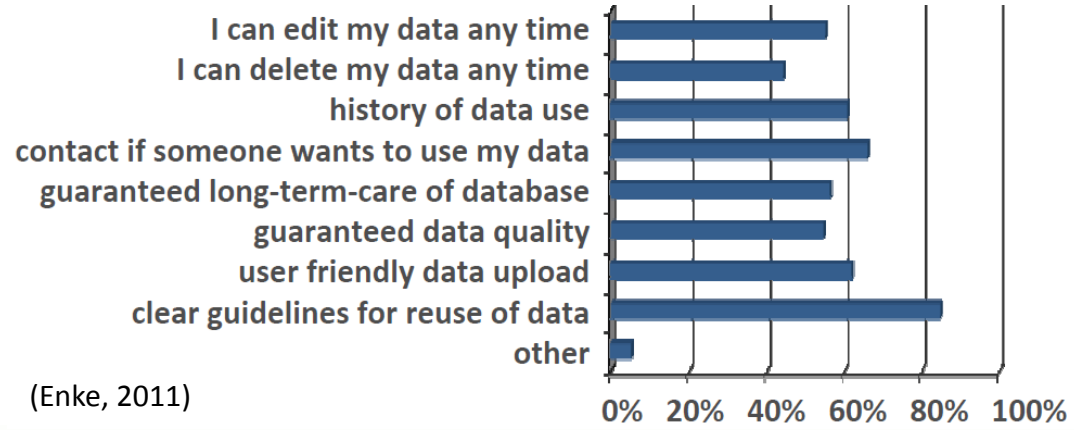
Hopes	Fears	Challenges
		
Data management online	Time effort	Data linking within different platforms
Data sharing policy	Data control	
Long term storage	Versioning	
Data synthesis	Flexibility	
Data linking	User friendly	Permanent data storage
Project activities		
Communications	Acknowledgement	

Outline

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

Hopes	Fears	Challenges

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



(Enke, 2011)

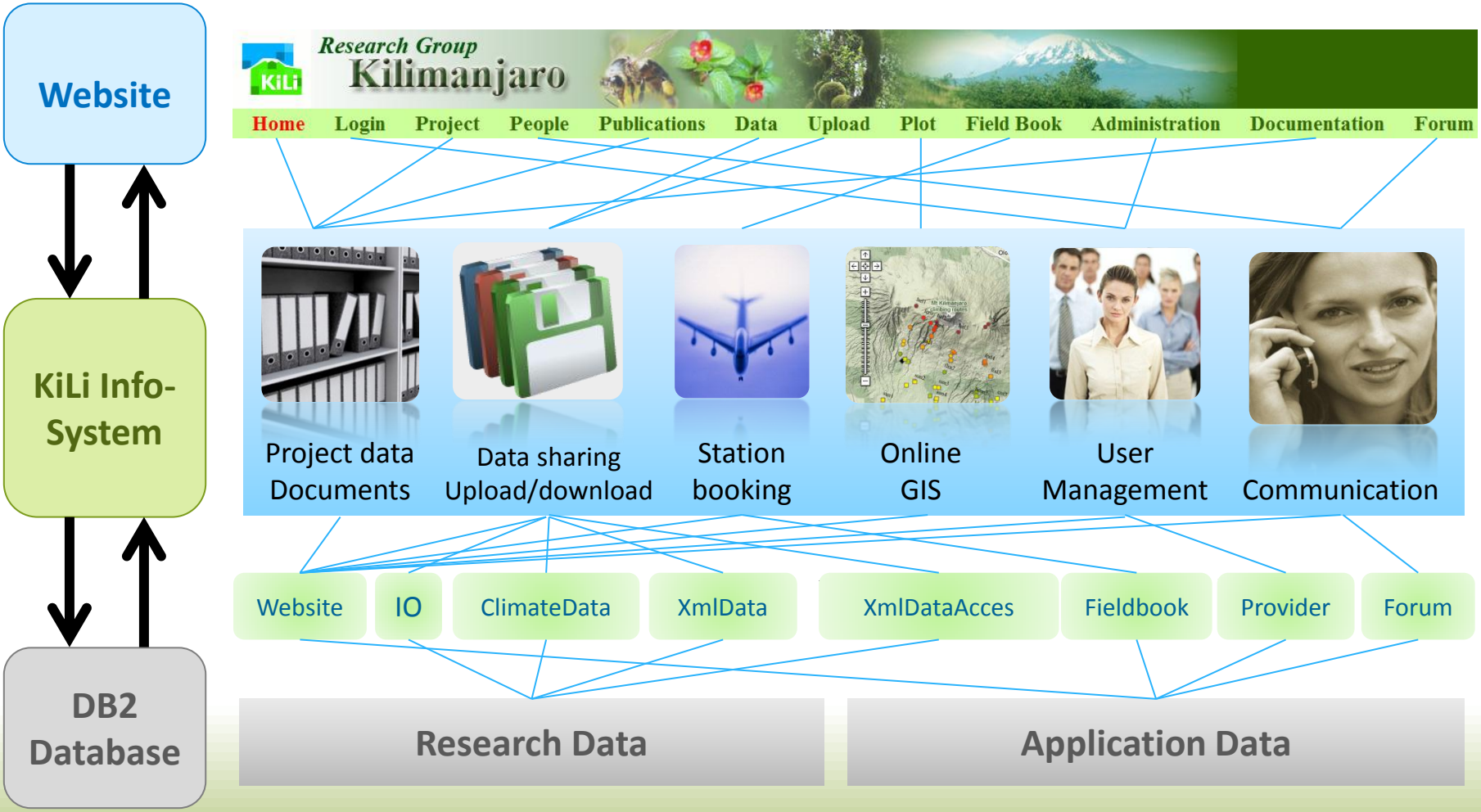
KiLi central database construction

"If I have been able to see further, it was only because I stood on the shoulders of giants." Isaac Newton

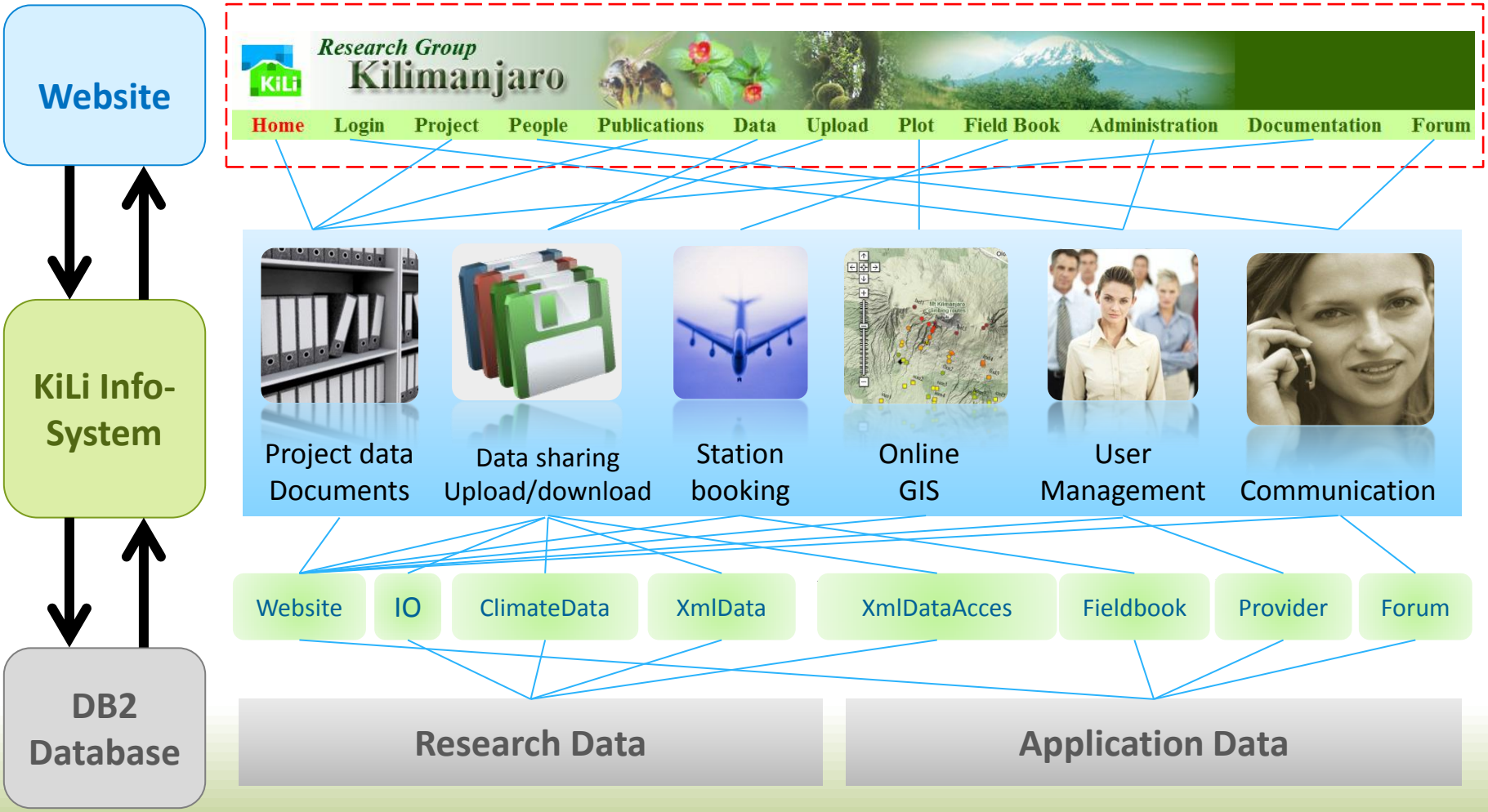
adapt → modify → Simplify → develop

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

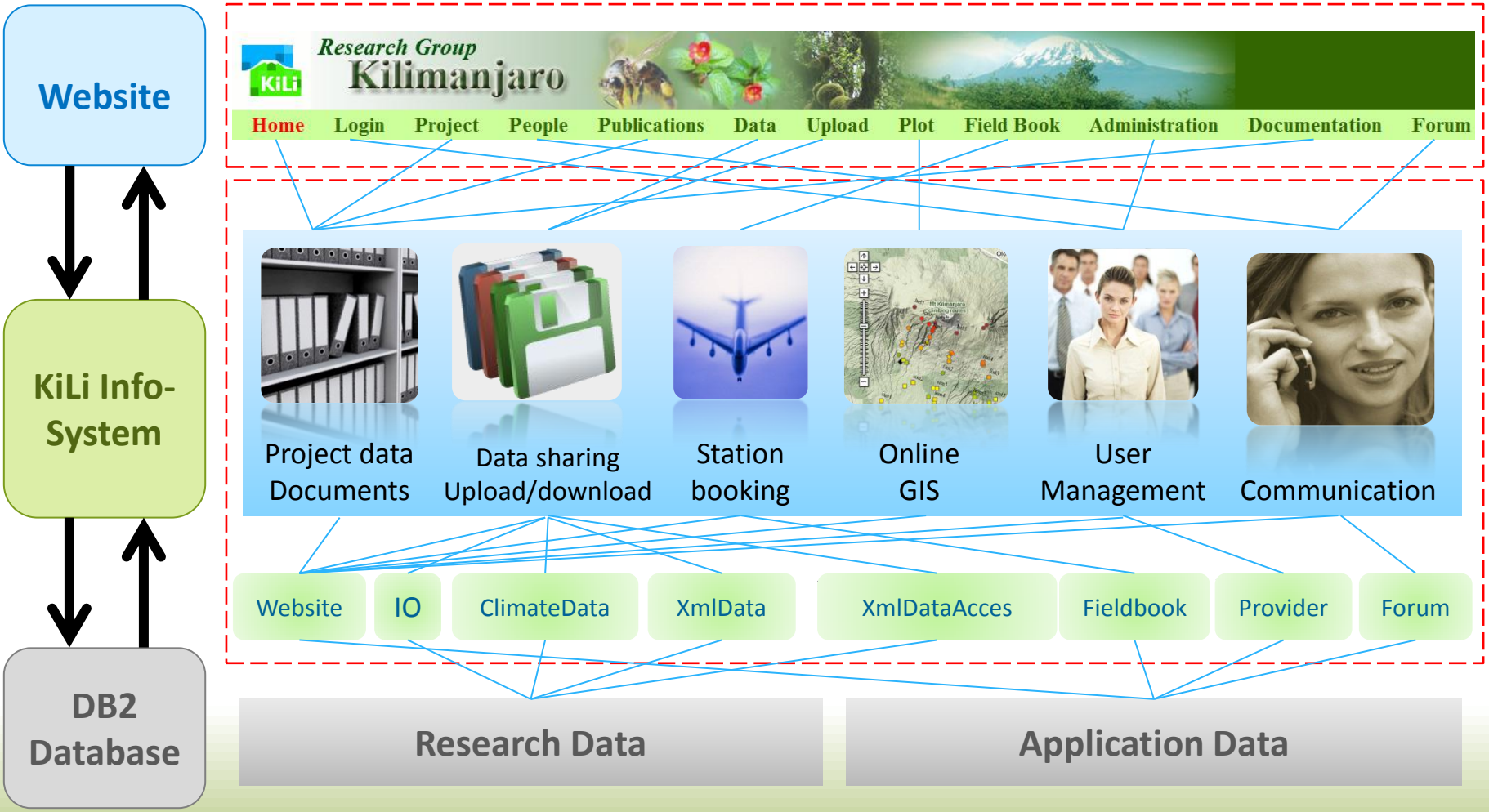
KiLi central database architecture



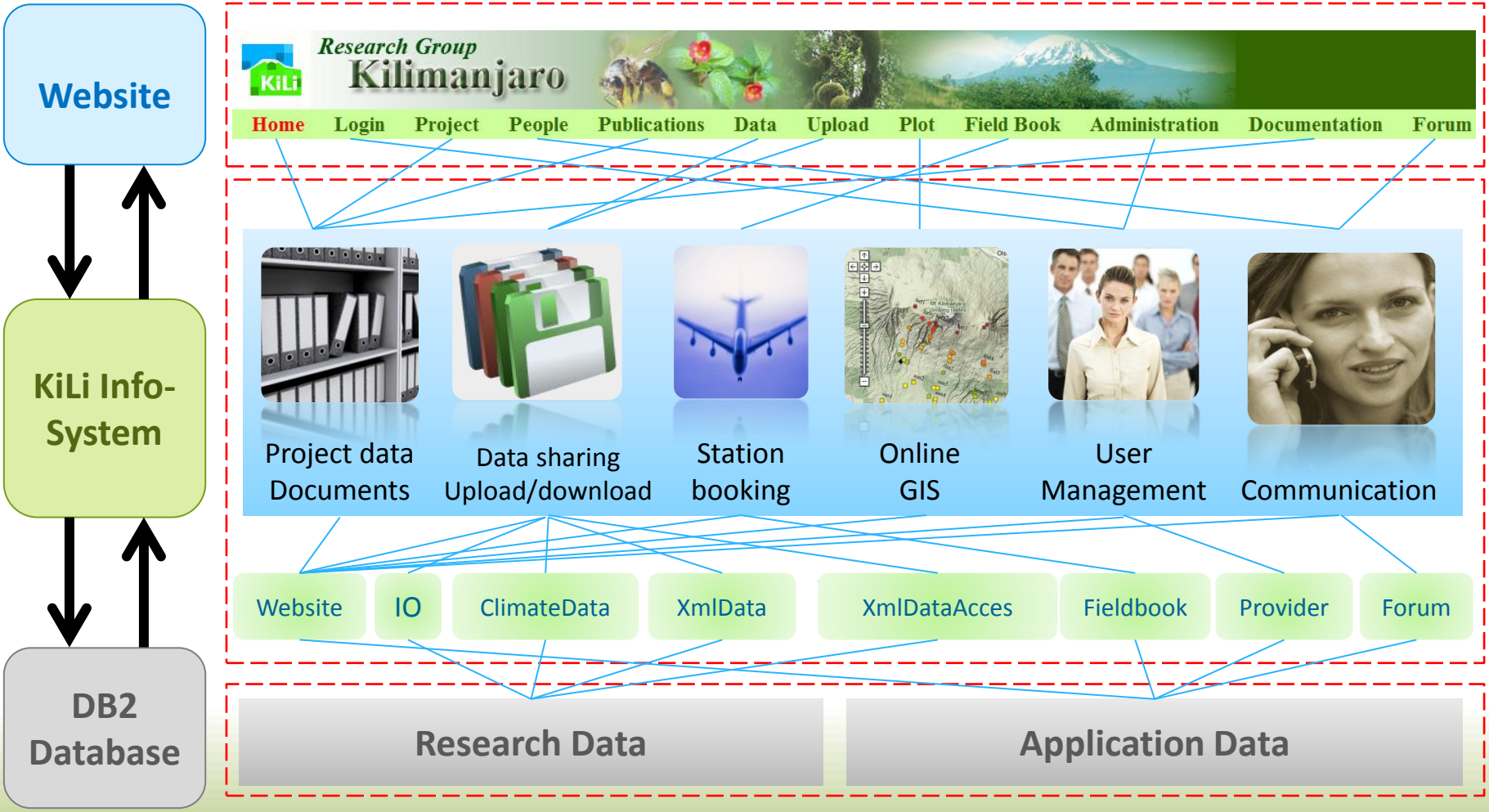
KiLi central database architecture



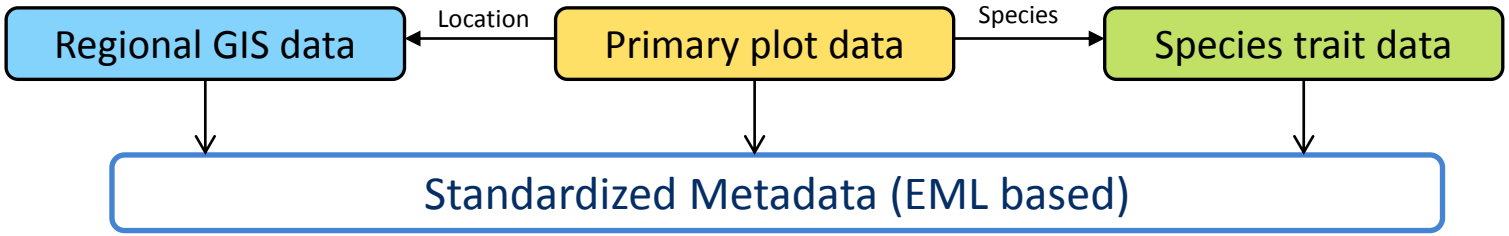
KiLi central database architecture



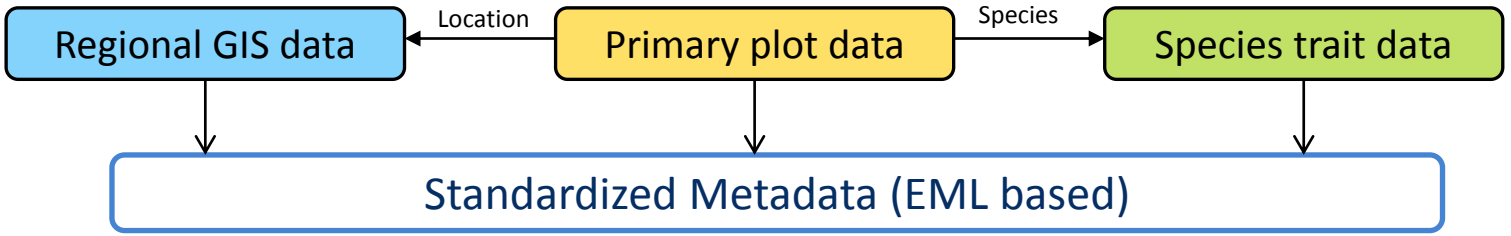
KiLi central database architecture

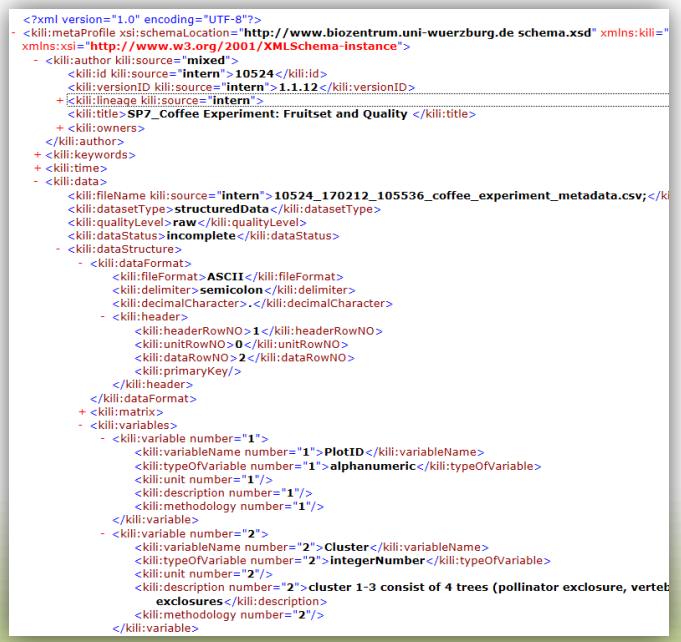


KiLi central database architecture

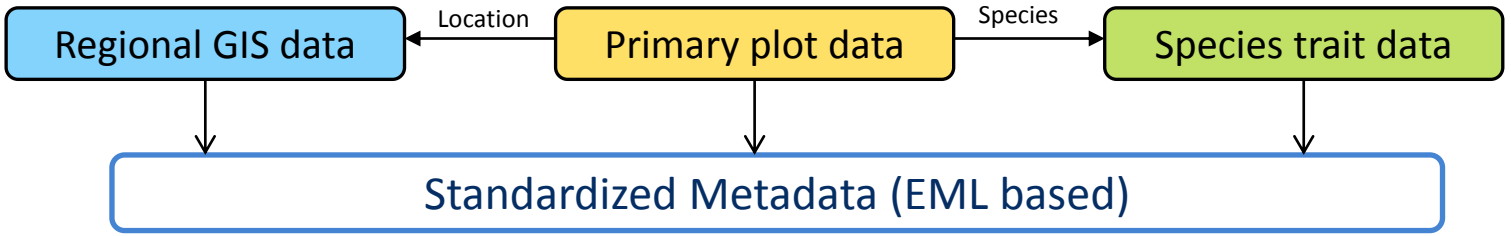


KiLi central database architecture

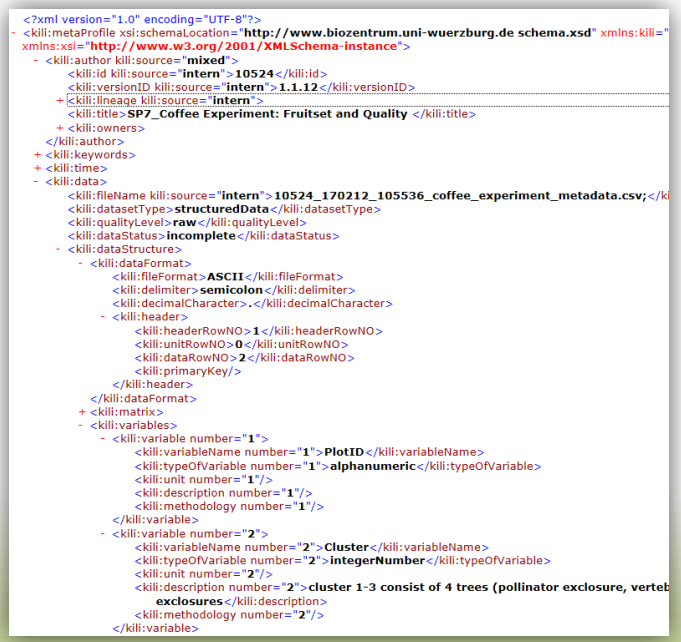




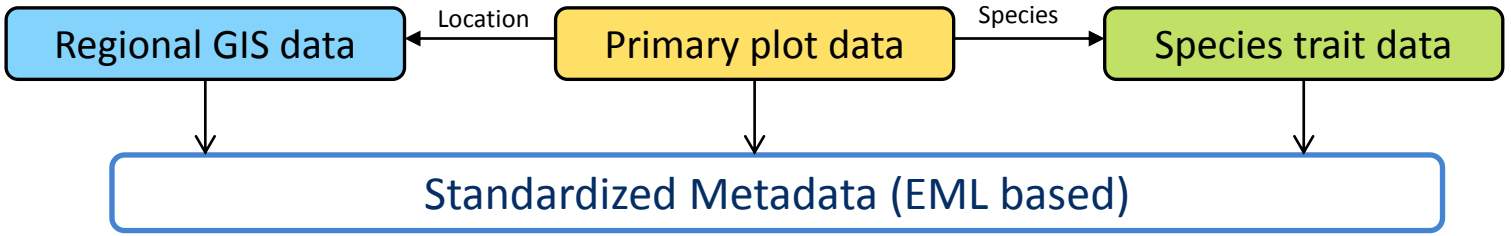
KiLi central database architecture

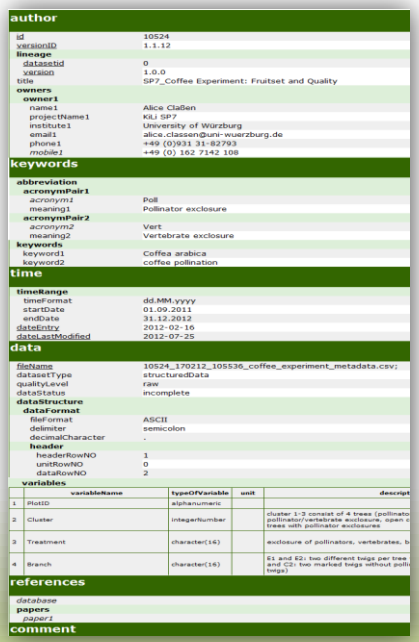


XSLT

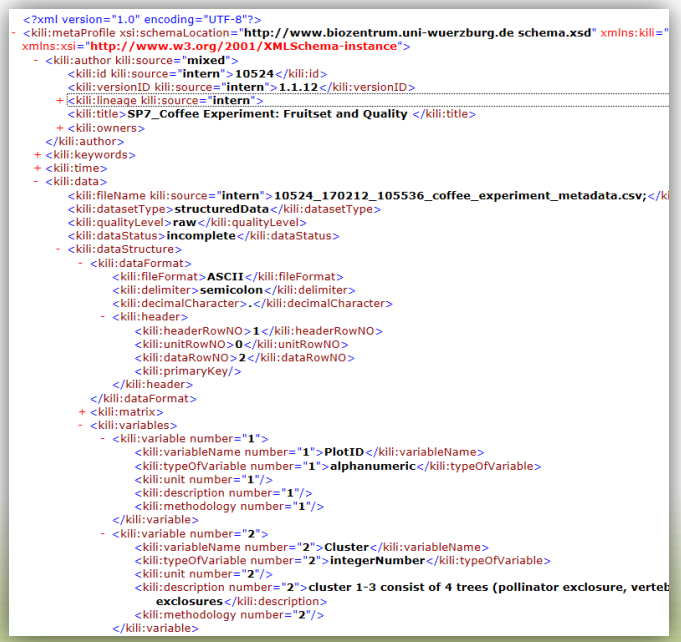


KiLi central database architecture

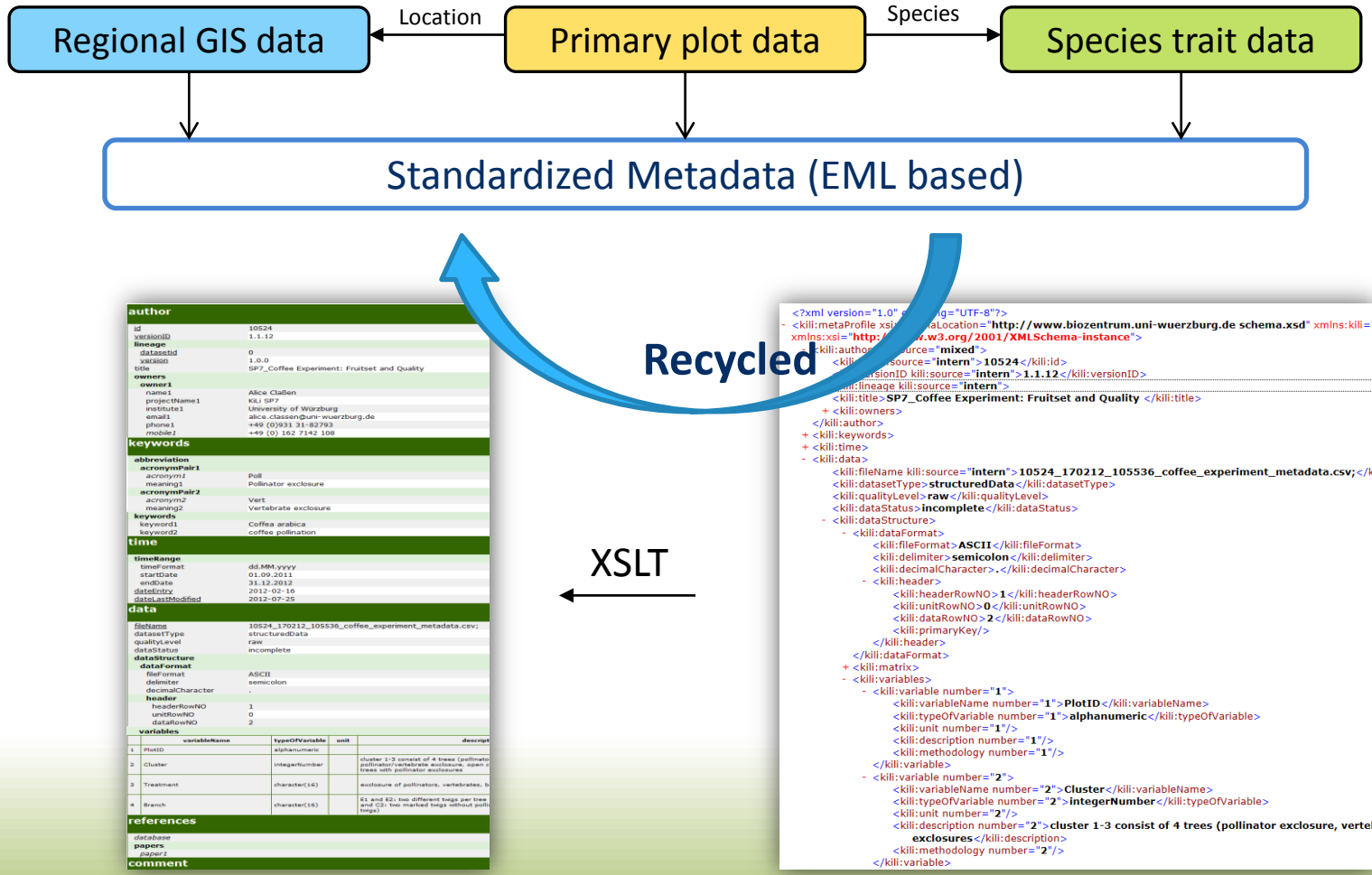


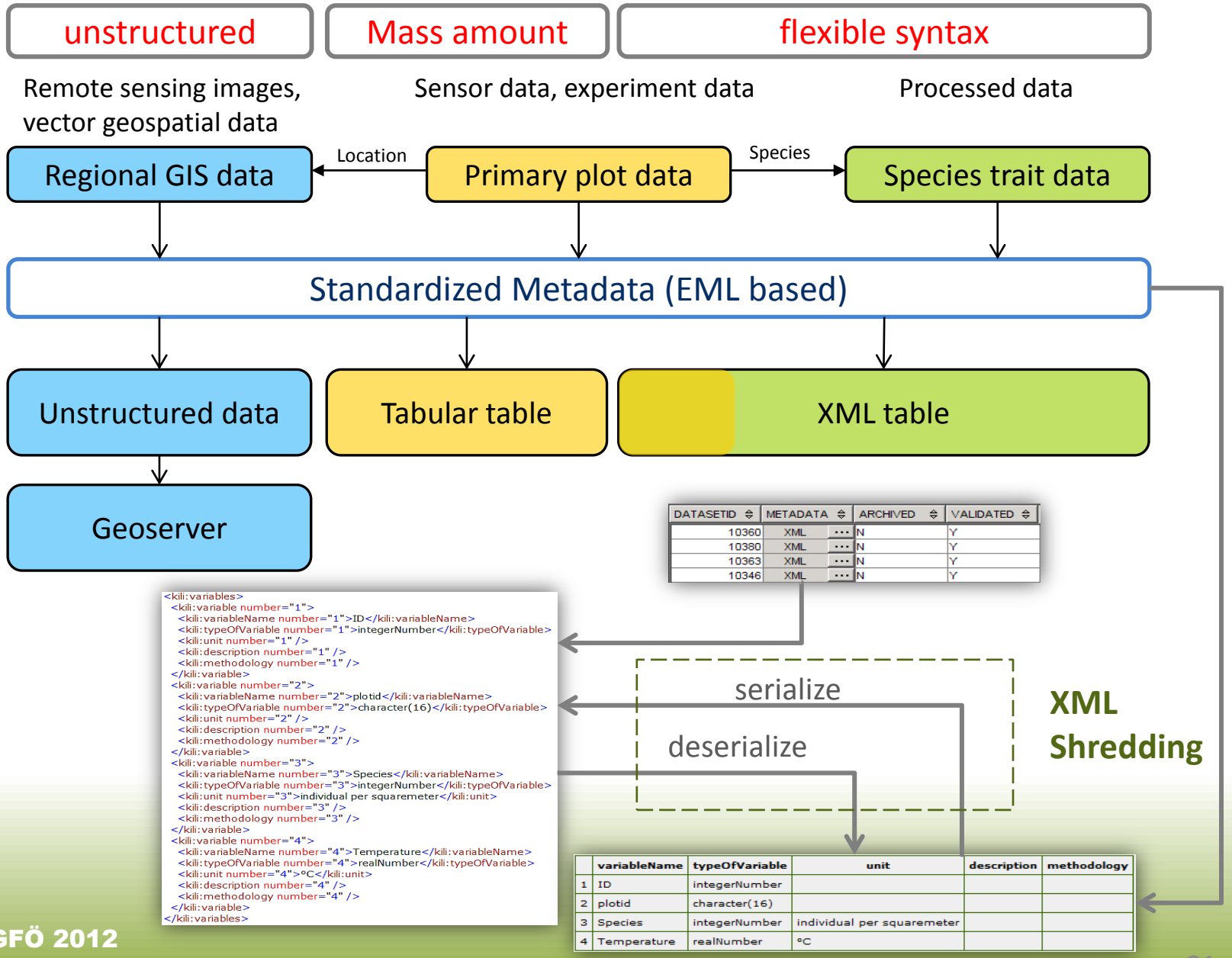


XSLT



KiLi central database architecture



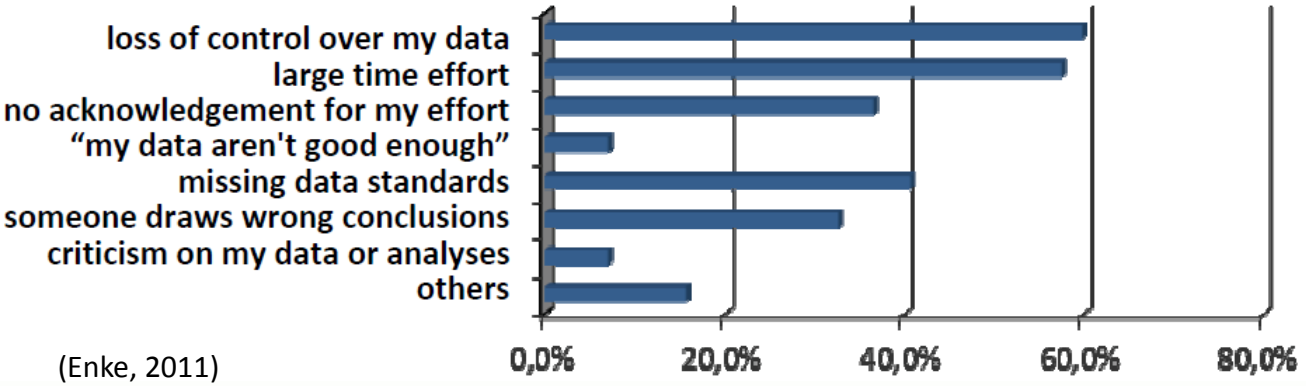

unstructured
Mass amount
flexible syntax
Remote sensing images, vector geospatial data
Sensor data, experiment data
Processed data
Regional GIS data
Location
Primary plot data
Species
Species trait data
Standardized Metadata (EML based)
Unstructured data
Tabular table
XML table
Geoserver
DATASETID METADATA ARCHIVED VALIDATED
10360 XML N Y
10380 XML N Y
10363 XML N Y
10346 XML N Y
<kili:variables>
<kili:variable number="1">
<kili:variableName number="1">ID</kili:variableName>
<kili:typeOfVariable number="1">integerNumber</kili:typeOfVariable>
<kili:unit number="1" />
<kili:description number="1" />
<kili:methodology number="1" />
</kili:variable>
<kili:variable number="2">
<kili:variableName number="2">plotid</kili:variableName>
<kili:typeOfVariable number="2">character(16)</kili:typeOfVariable>
<kili:unit number="2" />
<kili:description number="2" />
<kili:methodology number="2" />
</kili:variable>
<kili:variable number="3">
<kili:variableName number="3">Species</kili:variableName>
<kili:typeOfVariable number="3">integerNumber</kili:typeOfVariable>
<kili:unit number="3">individual per squaremeter</kili:unit>
<kili:description number="3" />
<kili:methodology number="3" />
</kili:variable>
<kili:variable number="4">
<kili:variableName number="4">Temperature</kili:variableName>
<kili:typeOfVariable number="4">realNumber</kili:typeOfVariable>
<kili:unit number="4">°C</kili:unit>
<kili:description number="4" />
<kili:methodology number="4" />
</kili:variable>
</kili:variables>
serialize
deserialize
XML Shredding
variableName typeOfVariable unit description methodology
1 ID integerNumber
2 plotid character(16)
3 Species integerNumber individual per squaremeter
4 Temperature realNumber °C

Outline

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

Hopes	Fears	Challenges

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



(Enke, 2011)

KiLi central database design

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

Outline

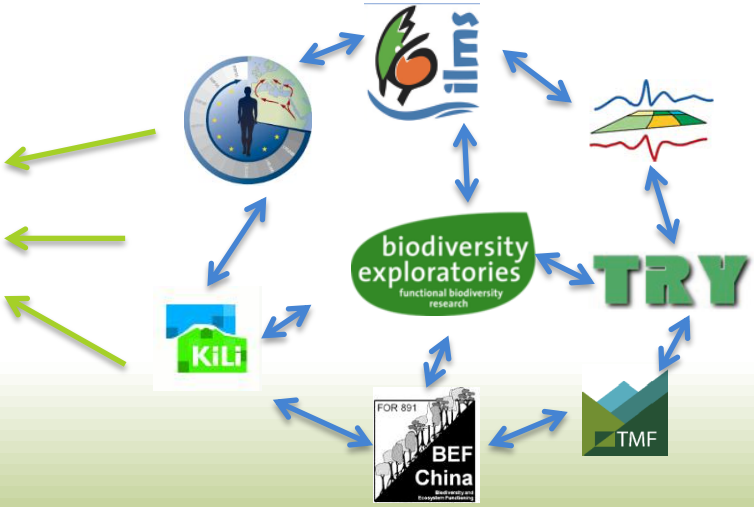
- KiLi project introduction
- KiLi central database design
  - ✓ Hopes fulfilled
  - ✓ Fears conquered
  - ✓ Challenge accepted
  - ✓ Future Outlook

Hopes	Fears	Challenges
		
Data management online	Time effort	Data linking within different platforms
Data sharing policy	Data control	
Long term storage	Versioning	
Data synthesis	Flexibility	
Data linking	User friendly	Permanent data storage
Project activities		
Communications	Acknowledgement	

❖ Data linking within different platforms

- Metadata standard → ✓ EML ✓ Cooperation between projects
- Semantic search → ✓ Ontology
- Online access → ✓ Link of web interface ✓ **Central homepage for data integration**

Central homepage



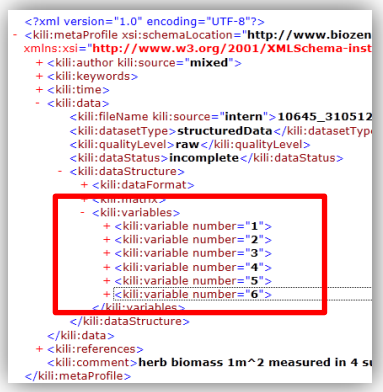
DataONE

LIFEWATCH

❖ Data linking within different platforms

- Transformation of data structure ✓ XML table

Dataset



Metadata

PLOTID	ECOT	SQUA	WEIG	NOTE
MAI5	MAI	4		
MAI5	MAI	3		
MAI4	MAI	1		
FOC5	FOC	4	17,87	18.11.2011 00:00:00
FOC5	FOC	3	19,3	18.11.2011 00:00:00
FOC5	FOC	2	37,85	18.11.2011 00:00:00
FOC5	FOC	1	35,83	18.11.2011 00:00:00
FOC4	FOC	4	12,14	08.11.2011 00:00:00
FOC4	FOC	3	21,23	08.11.2011 00:00:00
FOC4	FOC	2	26,83	08.11.2011 00:00:00

Observations

<obsId>
11392
</obsId>
<ID>
64
</ID>
<PLOTID>
FOC5
</PLOTID>
<ECOT>
FOC
</ECOT>
<SQUA>
4
</SQUA>
<WEIG>
17.87
</WEIG>
<NOTE>
18.11.2011 00:00:00
</NOTE>

OBSID	DATA	DATASETID
11482	XML	10645
11481	XML	10645
11475	XML	10645
11392	XML	10645
11391	XML	10645
11390	XML	10645
11389	XML	10645
11388	XML	10645
11387	XML	10645
11386	XML	10645
11385	XML	10645
11384	XML	10645

One fixed table structure for all data!

❖ Store data permanently

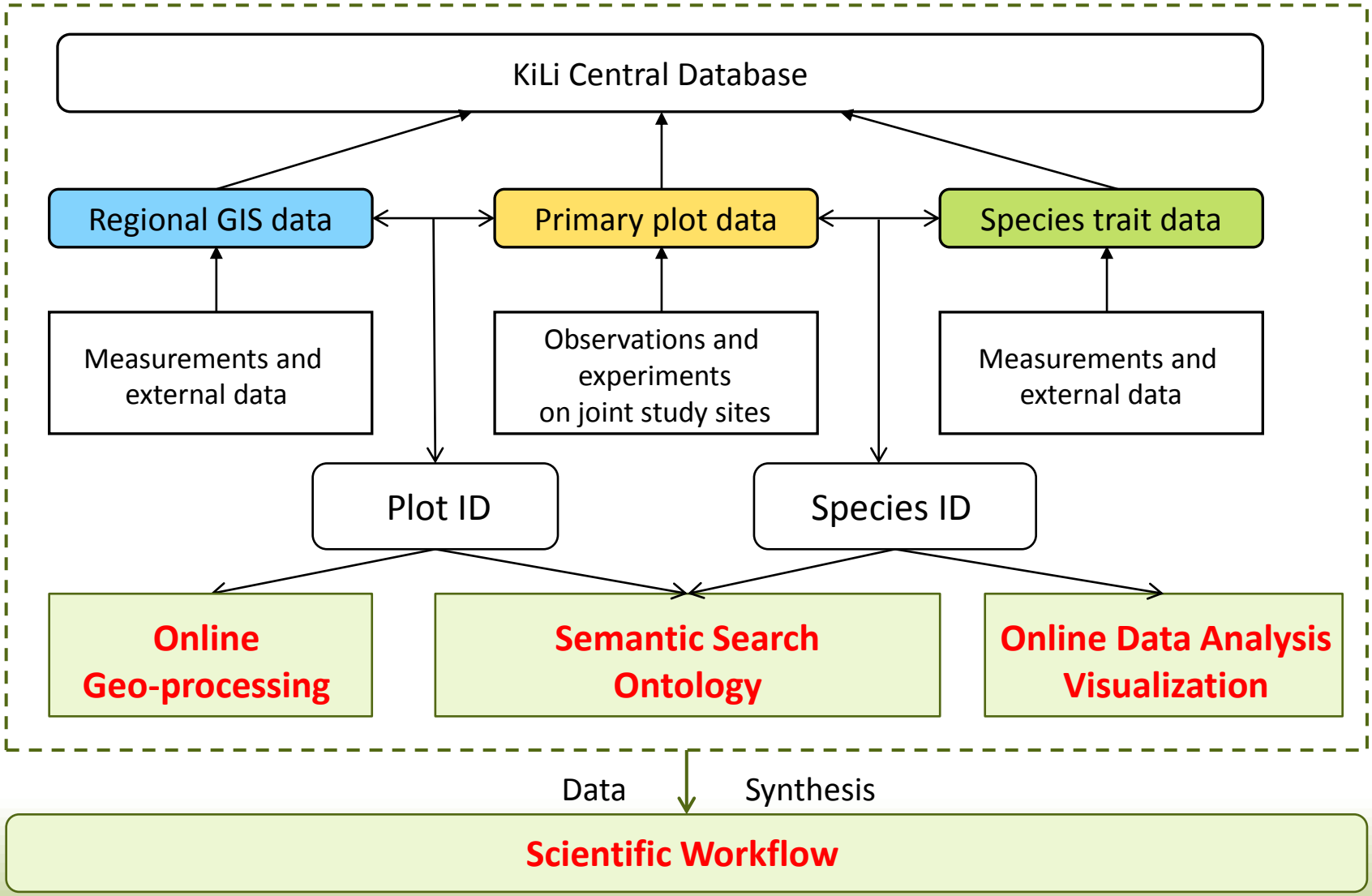
- Journals (e.g. Ecological Archives) -> permanent, limited
- Project archives -> long term, not permanent
- **Combine of existing data centers (universities, libraries, institutes)**

Remote sensing
Climate center

Ecological center

Outline

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


KiLi Central Database
Regional GIS data
Primary plot data
Species trait data
Measurements and external data
Observations and experiments on joint study sites
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