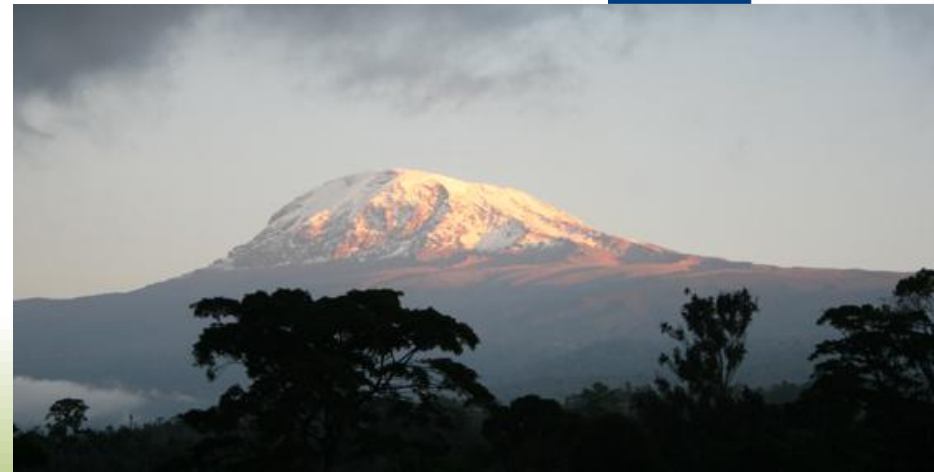


Central database design and data synthesis in the DFG Research Unit FOR 1246 – Kilimanjaro ecosystems under global change (KiLi)

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



Research Group
Kilimanjaro





Outline

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



Outline

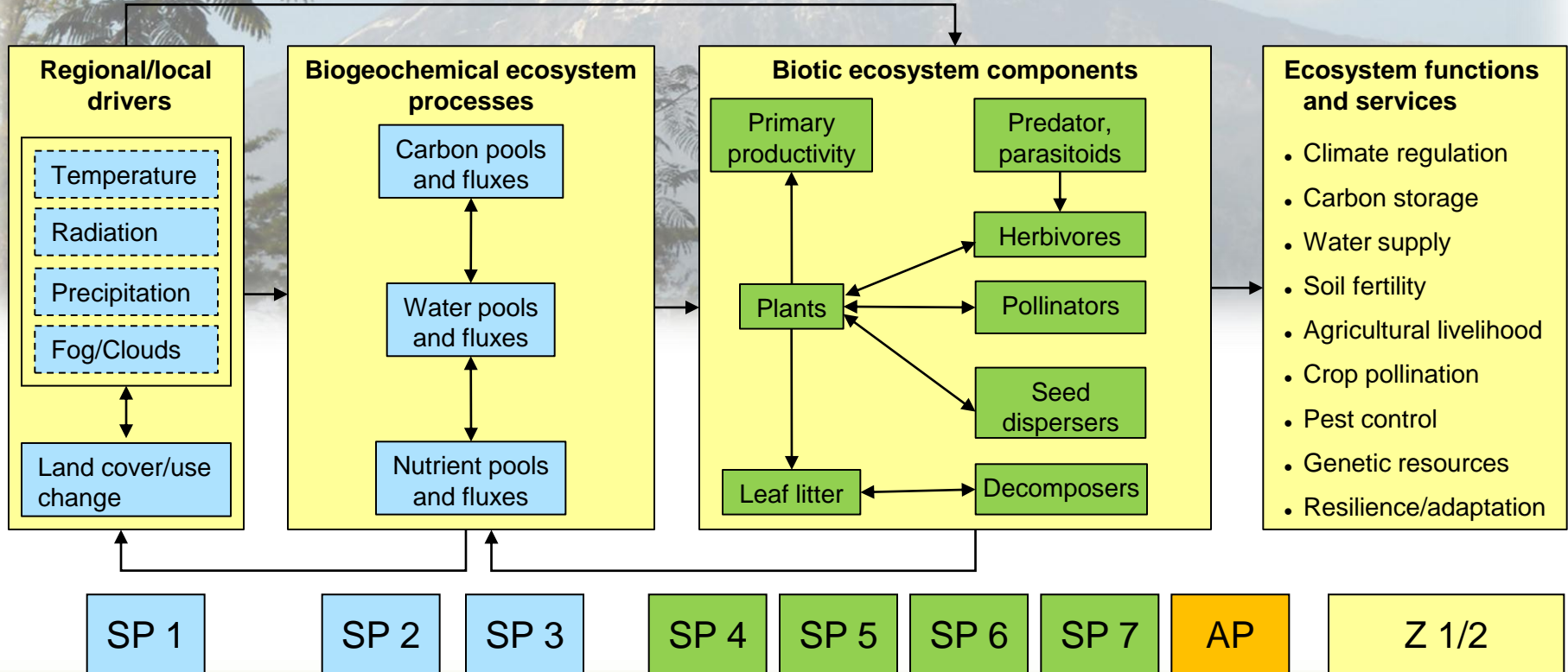
❖ KiLi project introduction

❖ KiLi central database design

- ✓ Hopes fulfilled
- ✓ Fears conquered
- ✓ Challenge accepted
- ✓ Future Outlook

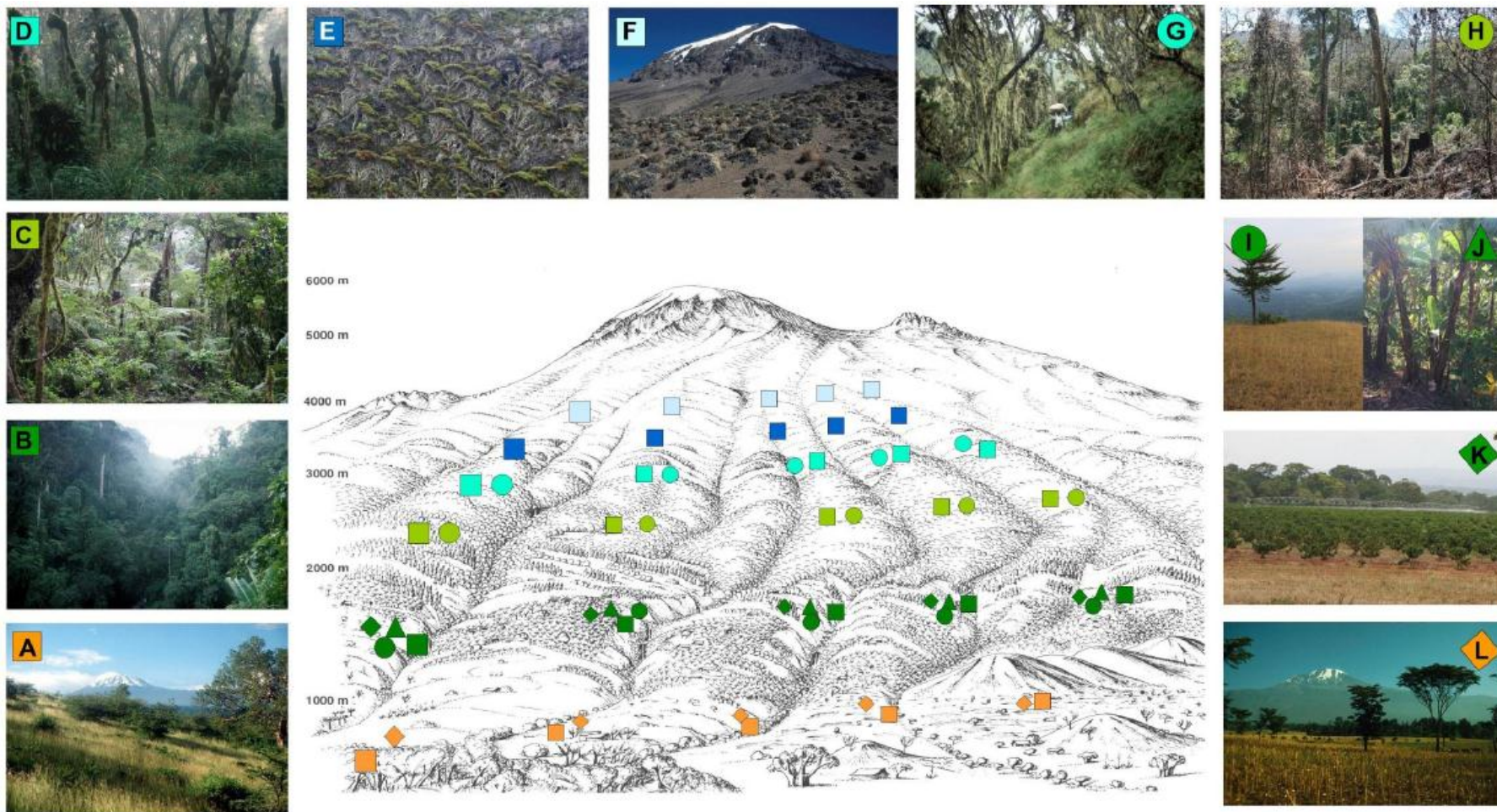
KiLi – Kilimanjaro ecosystems under global change

Global climate and land use change

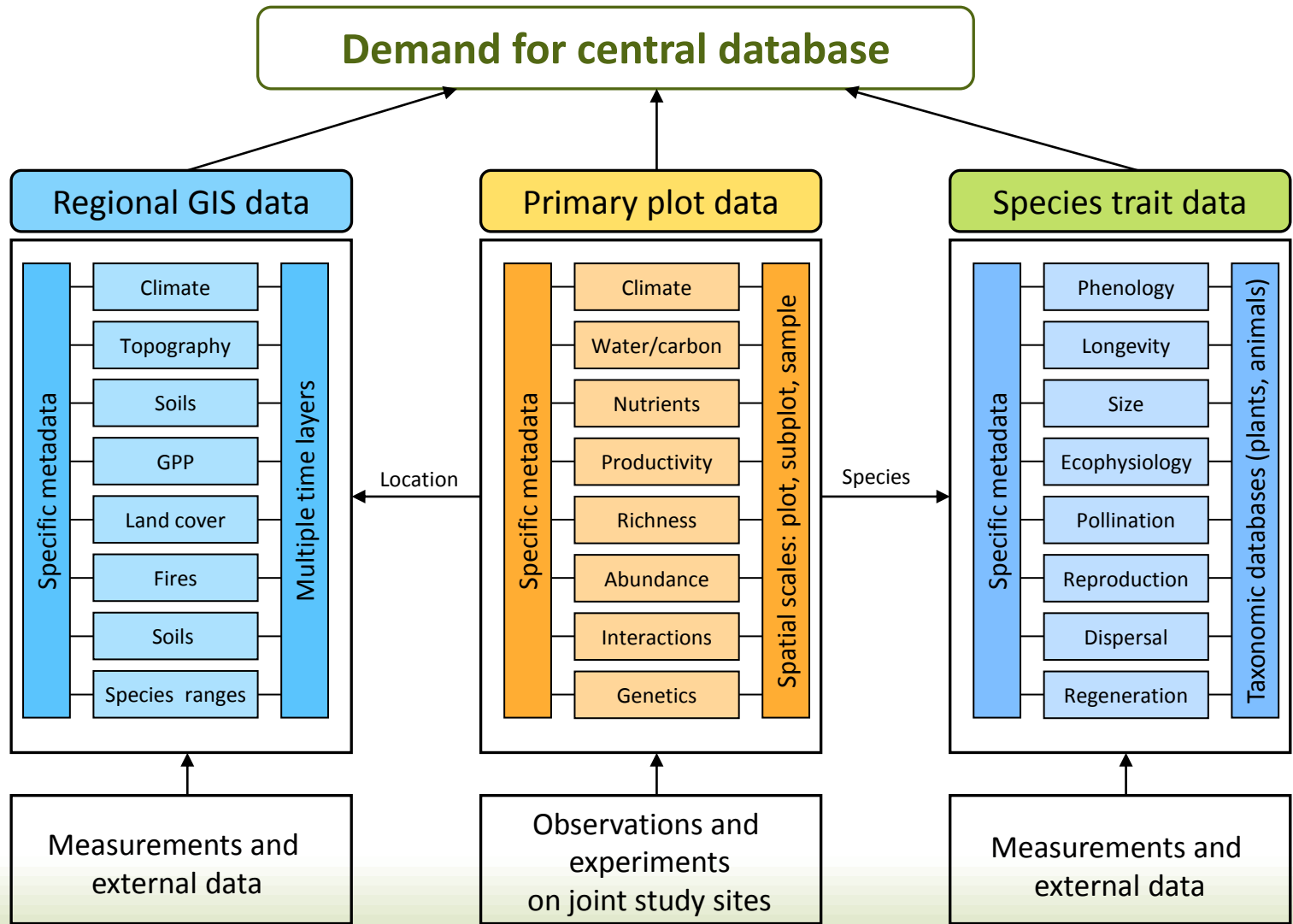


Data → Synthesis





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





Outline

❖ KiLi project introduction

❖ **KiLi central database design**

- ✓ Hopes fulfilled
- ✓ Fears conquered
- ✓ Challenges accepted
- ✓ Future Outlook



Hopes



Fears



Challenges



Data management online

Data sharing policy

Long term storage

Data synthesis

Data linking

Project activities

Communications

Time effort

Data control

Versioning

Flexibility

User friendly

Acknowledgement

Data linking within
different platforms

Permanent data storage



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

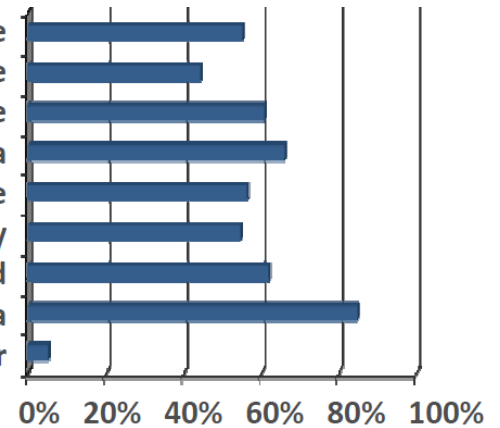
Data linking

Project activities

Communications

I can edit my data any time
 I can delete my data any time
 history of data use
 contact if someone wants to use my data
 guaranteed long-term-care of database
 guaranteed data quality
 user friendly data upload
 clear guidelines for reuse of data
 other

(Enke, 2011)

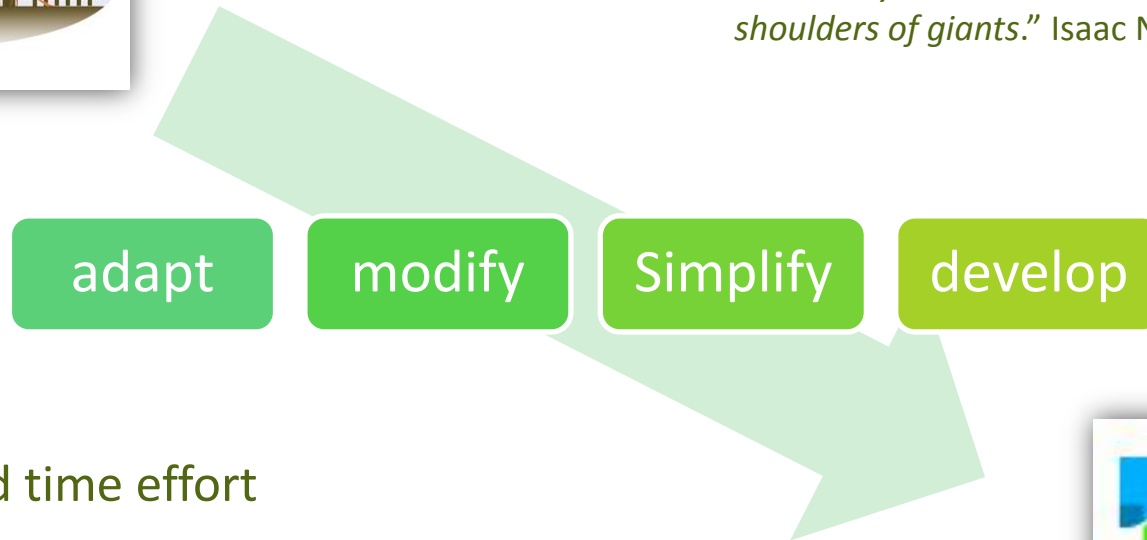




KiLi central database construction



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

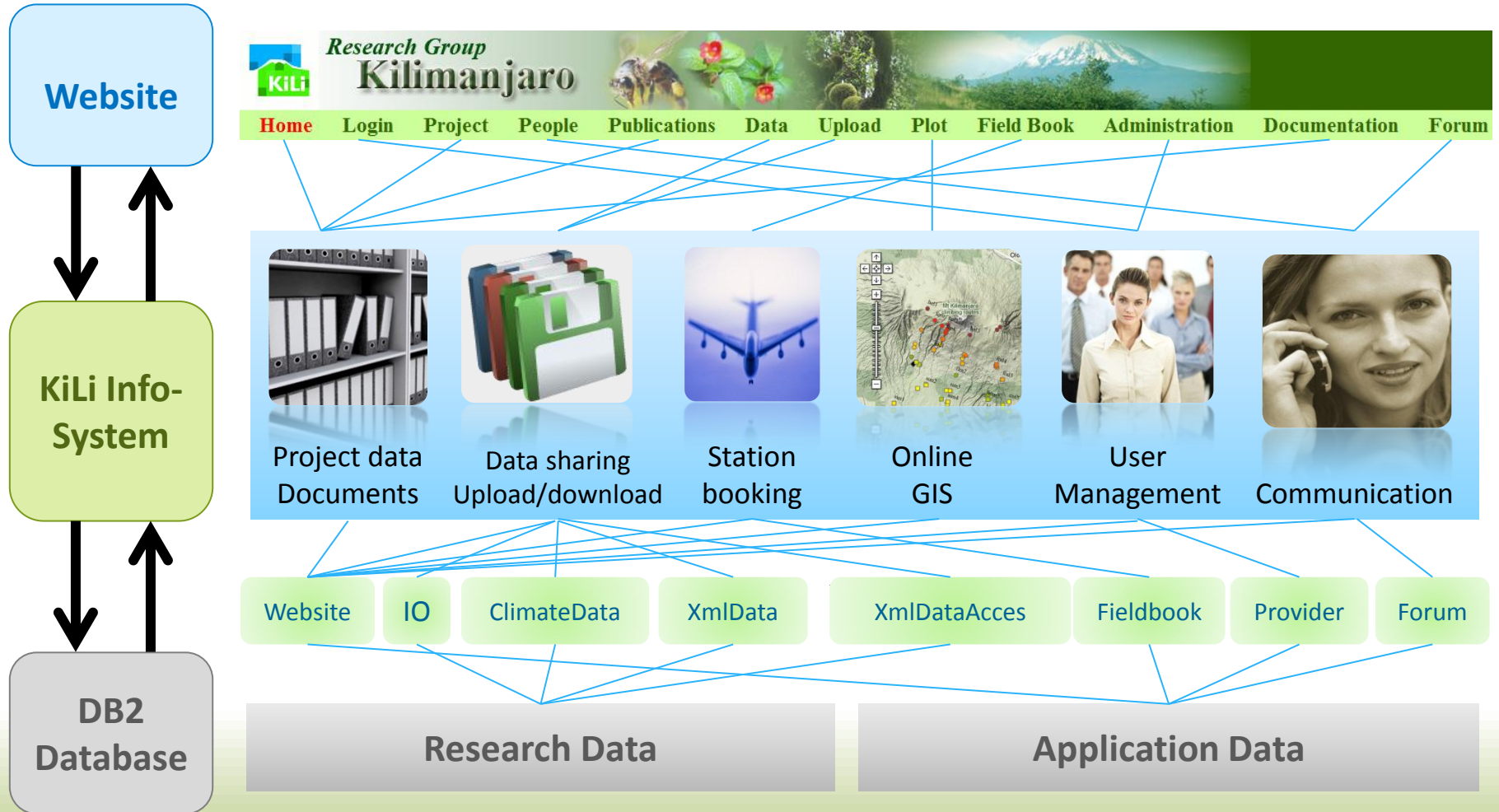


- ✓ Rapidly reduced time effort
- ✓ Improve platform compatibility and data linking
- ✓ Knowledge sharing and exchange

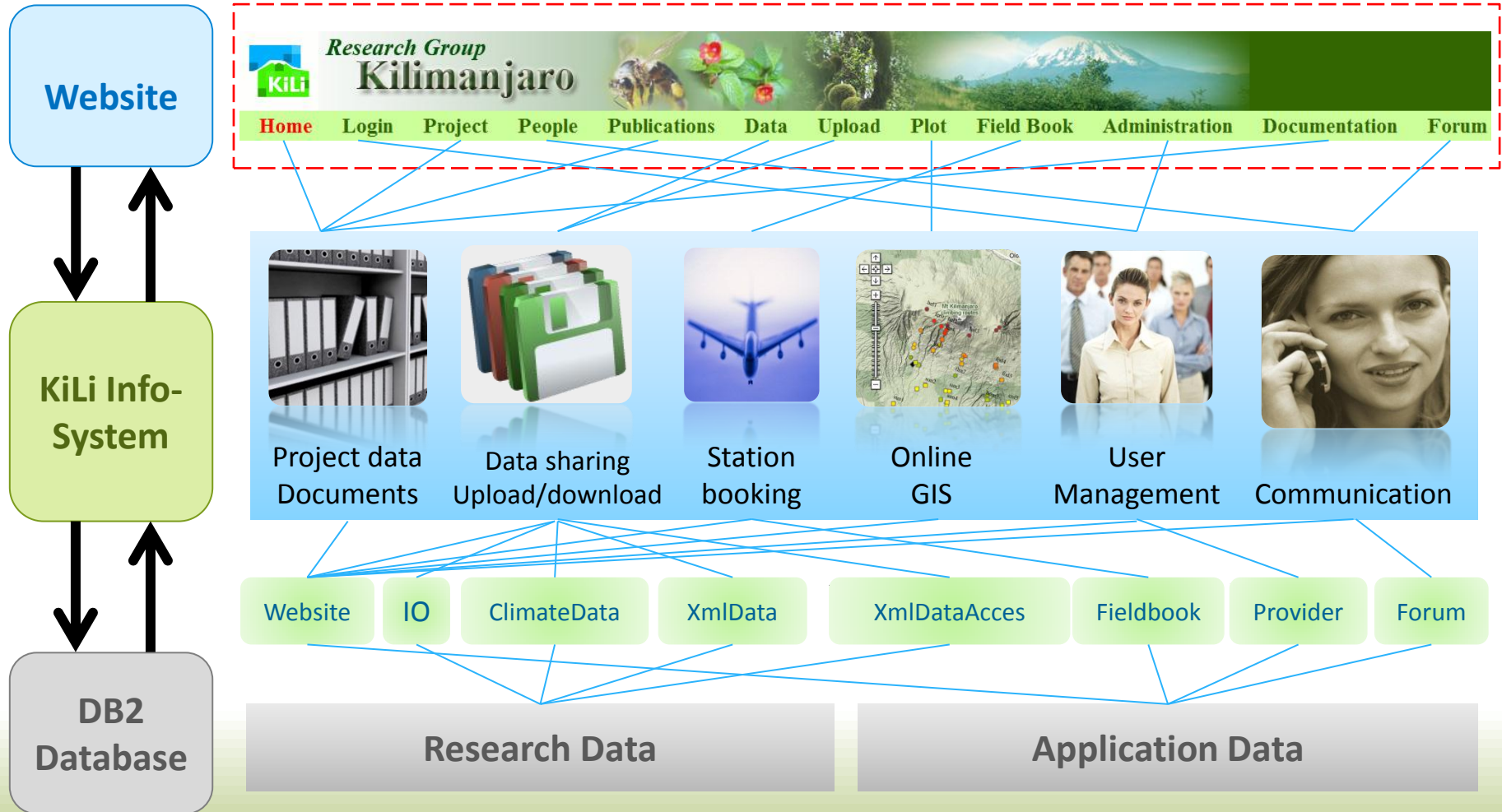




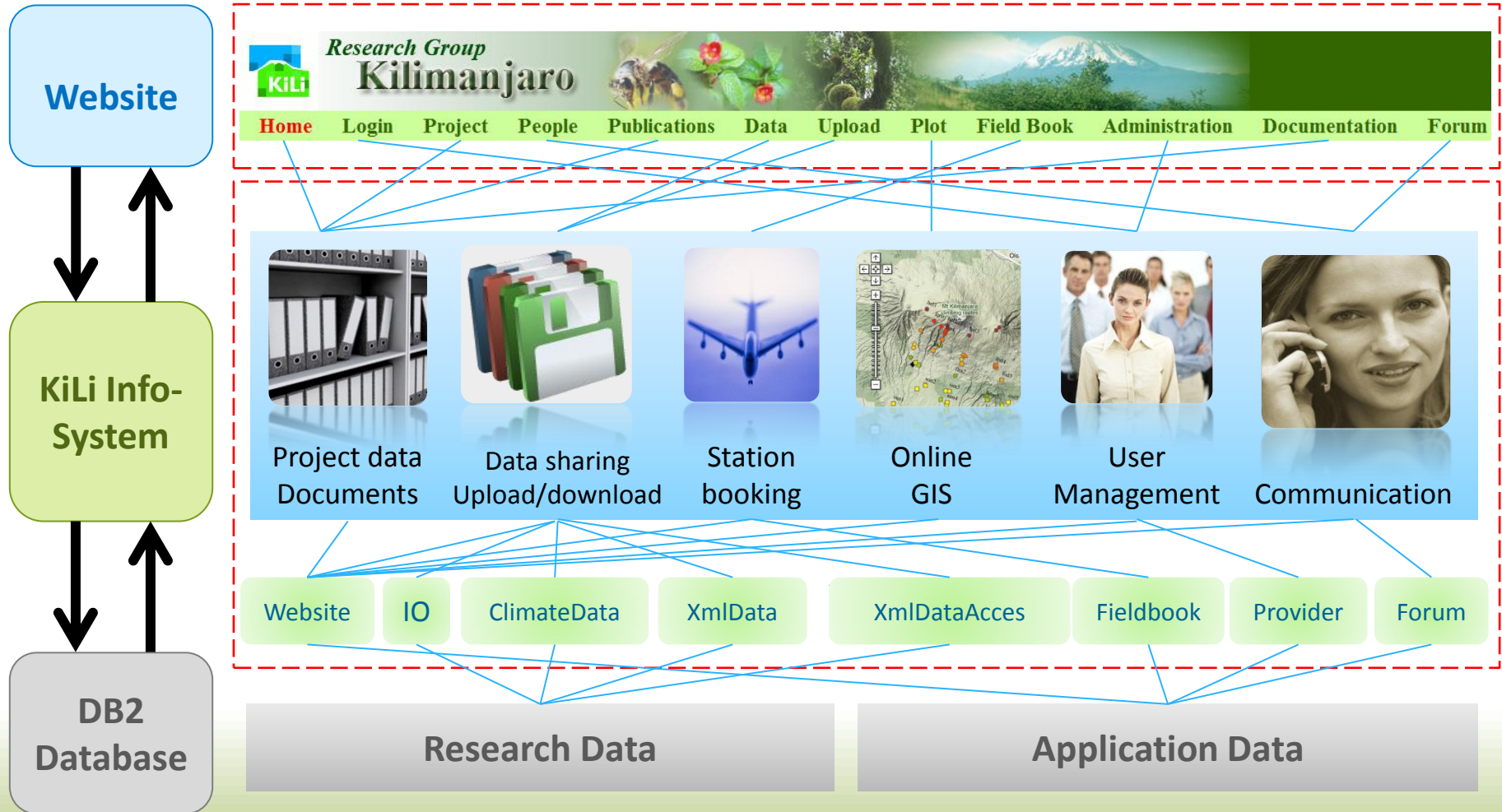
KiLi central database architecture



KiLi central database architecture



KiLi central database architecture



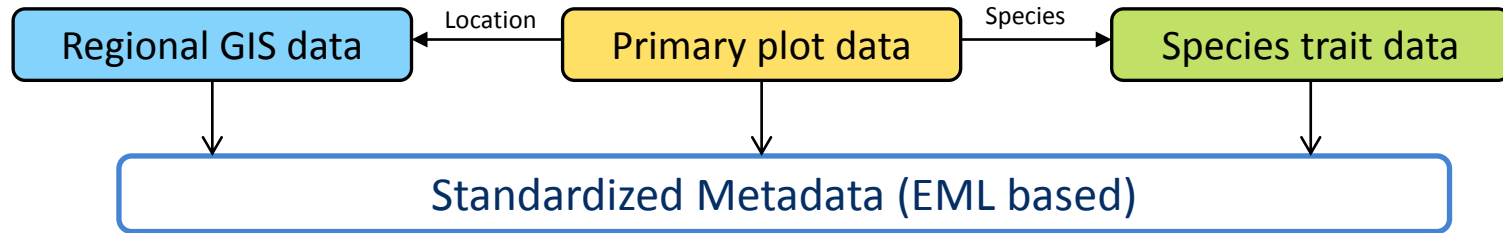


KiLi central database architecture

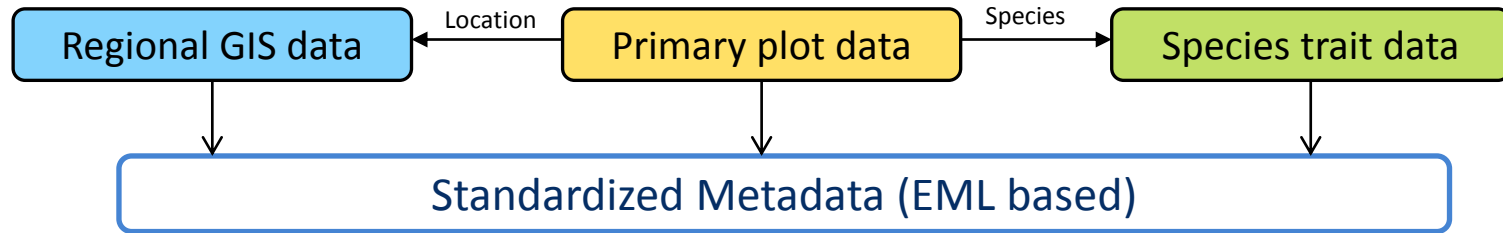




KiLi central database architecture



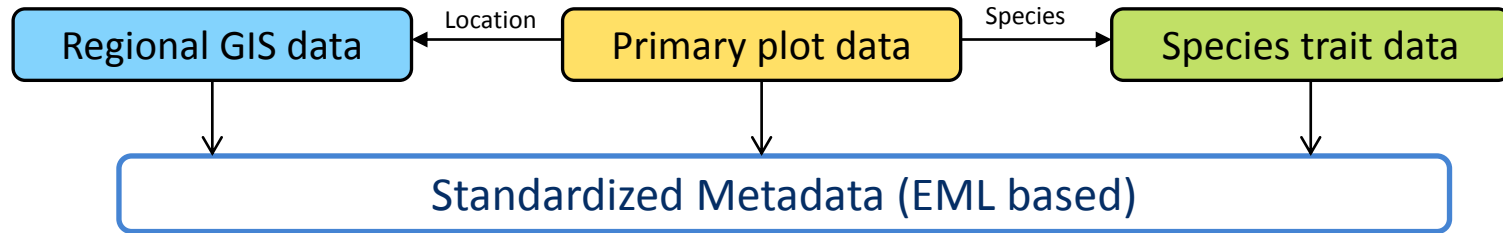
KiLi central database architecture



```
<?xml version="1.0" encoding="UTF-8"?>
<kili:metaProfile xsi:schemaLocation="http://www.biozentrum.uni-wuerzburg.de/schema.xsd" xmlns:kili="http://www.w3.org/2001/XMLSchema-instance"
  <kili:author kili:source="mixed">
    <kili:id kili:source="Intern">10524</kili:id>
    <kili:versionID kili:source="Intern">1.1.12</kili:versionID>
  <kili:lineage kili:source="Intern">
    <kili:title>SP7_Coffee Experiment: Fruitset and Quality</kili:title>
    <kili:owners>
    </kili:author>
    <kili:keywords>
    <kili:time>
  <kili:dataset>
    <kili:fileName kili:source="Intern">10524_170212_105536_coffee_experiment_metadata.csv</kili:fileName>
    <kili:datasetType>structuredData</kili:datasetType>
    <kili:qualityLevel>raw</kili:qualityLevel>
    <kili:dataStatus>incomplete</kili:dataStatus>
  <kili:dataStructure>
    <kili:dataFormat>
      <kili:fileFormat>ASCII</kili:fileFormat>
      <kili:delimiter>semicolon</kili:delimiter>
      <kili:decimalCharacter>.</kili:decimalCharacter>
    <kili:header>
      <kili:headerRowNO>1</kili:headerRowNO>
      <kili:unitRowNO>0</kili:unitRowNO>
      <kili:dataRowNO>2</kili:dataRowNO>
      <kili:primaryKey>
    </kili:header>
    </kili:dataFormat>
    <kili:matrix>
      <kili:variables>
        <kili:variable number="1">
          <kili:variableName number="1">PlotID</kili:variableName>
          <kili:typeOfVariable number="1">alphanumeric</kili:typeOfVariable>
          <kili:unit number="1"/>
          <kili:description number="1"/>
          <kili:methodology number="1"/>
        </kili:variable>
        <kili:variable number="2">
          <kili:variableName number="2">Cluster</kili:variableName>
          <kili:typeOfVariable number="2">integerNumber</kili:typeOfVariable>
          <kili:unit number="2"/>
          <kili:description number="2">cluster 1-3 consist of 4 trees (pollinator enclosure, verte
            <kili:methodology number="2"/>
          </kili:variable>
      </kili:variables>
    </kili:matrix>
  </kili:dataset>
</kili:metaProfile>
```



KiLi central database architecture



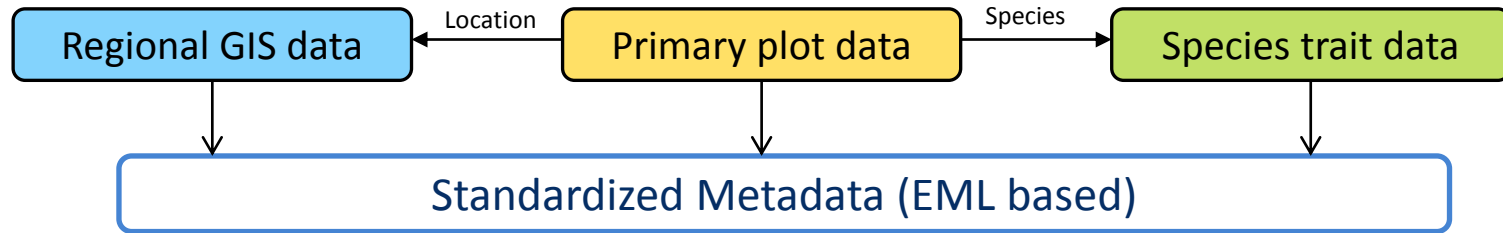
| author | | | |
|------------------|--|------|---|
| id | 10524 | | |
| versionID | 1.1.12 | | |
| lineage | datasetid | | |
| datasetid | 0 | | |
| version | 1.0.0 | | |
| title | SP7_Coffee Experiment: Fruitset and Quality | | |
| owners | | | |
| name1 | Alice Claßen | | |
| projectName1 | KLi SP7 | | |
| institute1 | University of Würzburg | | |
| email | alice.claassen@uni-wuerzburg.de | | |
| phone1 | +49 (0)931 31-82793 | | |
| mobile1 | +49 (0) 162 7142 108 | | |
| keywords | | | |
| abbreviation | | | |
| acronymPair1 | | | |
| acronym1 | Pol | | |
| meaning1 | Pollinator exclusion | | |
| acronymPair2 | | | |
| acronym2 | Vert | | |
| meaning2 | Vertebrate exclusion | | |
| keywords | | | |
| keyword1 | Coffea arabica | | |
| keyword2 | coffee pollination | | |
| time | | | |
| timeRange | | | |
| timeFormat | dd MM, yyyy | | |
| startDate | 01.09.2011 | | |
| endDate | 31.12.2012 | | |
| dateEntry | 2012-02-16 | | |
| dateLastModified | 2012-07-25 | | |
| data | | | |
| fileName | 10524_170212_105536_coffee_experiment_metadata.csv | | |
| datasetType | structuredData | | |
| qualityLevel | raw | | |
| dataStatus | incomplete | | |
| dataStructure | | | |
| fileFormat | ASCII | | |
| delimiter | semicolon | | |
| decimalCharacter | . | | |
| header | | | |
| headerRowNO | 1 | | |
| unitRowNO | 0 | | |
| dataRowNO | 2 | | |
| variables | | | |
| variableName | typeOfVariable | unit | descript |
| 1 PlotID | alphanumeric | | cluster 1-3 consist of 4 trees (pollinator exclusion, vertebrate exclusion, open n trees with pollinator exclusion) |
| 2 Cluster | integerNumber | | |
| 3 Treatment | character(16) | | exclusion of pollinators, vertebrates, b |
| 4 Branch | character(16) | | B1 and B2: two different tags per tree and C0: two marked tags (without pollinators) |
| references | | | |
| database | | | |
| papers | | | |
| papers1 | | | |
| comment | | | |

XSLT

```

<?xml version="1.0" encoding="UTF-8"?>
<kill:metaProfile xsi:schemaLocation="http://www.biozentrum.uni-wuerzburg.de/schema.xsd" xmlns:kill="http://www.w3.org/2001/XMLSchema-instance">
  <kill:author kill:source="mixed">
    <kill:id kill:source="Intern">10524</kill:id>
    <kill:versionID kill:source="Intern">1.1.12</kill:versionID>
  </kill:author>
  <kill:lineage kill:source="Intern">
    <kill:title>SP7_Coffee Experiment: Fruitset and Quality</kill:title>
    <kill:owners>
      <kill:author>
      <kill:keywords>
      <kill:time>
    </kill:owners>
  </kill:lineage>
  <kill:fileFormat kill:source="Intern">10524_170212_105536_coffee_experiment_metadata.csv</kill:fileFormat>
  <kill:datasetType>structuredData</kill:datasetType>
  <kill:qualityLevel>raw</kill:qualityLevel>
  <kill:dataStatus>incomplete</kill:dataStatus>
  <kill:dataStructure>
    <kill:dataFormat>
      <kill:fileFormat>ASCII</kill:fileFormat>
      <kill:delimiter>semicolon</kill:delimiter>
      <kill:decimalCharacter>.</kill:decimalCharacter>
    </kill:dataFormat>
    <kill:header>
      <kill:headerRowNO>1</kill:headerRowNO>
      <kill:unitRowNO>0</kill:unitRowNO>
      <kill:dataRowNO>2</kill:dataRowNO>
      <kill:primaryKey>
    </kill:header>
    <kill:dataFormat>
      <kill:matrix>
    </kill:dataFormat>
  </kill:dataStructure>
  <kill:variables>
    <kill:variable number="1">
      <kill:variableName number="1">PlotID</kill:variableName>
      <kill:typeOfVariable number="1">alphanumeric</kill:typeOfVariable>
      <kill:unit number="1"/>
      <kill:description number="1"/>
      <kill:methodology number="1"/>
    </kill:variable>
    <kill:variable number="2">
      <kill:variableName number="2">Cluster</kill:variableName>
      <kill:typeOfVariable number="2">integerNumber</kill:typeOfVariable>
      <kill:unit number="2"/>
      <kill:description number="2">cluster 1-3 consist of 4 trees (pollinator exclusion, vertebrate exclusion)</kill:description>
      <kill:methodology number="2"/>
    </kill:variable>
  </kill:variables>
</kill:metaProfile>
  
```

KiLi central database architecture



| author | | | |
|------------------|--|------|---|
| id | 10524 | | |
| versionID | 1.1.12 | | |
| lineage | datasetd | | |
| datasetd | 0 | | |
| version | 1.0.0 | | |
| title | SP7_Coffee Experiment: Fruitset and Quality | | |
| owners | | | |
| name1 | Alice Claßen | | |
| projectName1 | KLi SP7 | | |
| institute1 | University of Würzburg | | |
| email | alice.claassen@uni-wuerzburg.de | | |
| phone1 | +49 (0)931 31-82793 | | |
| mobile1 | +49 (0) 162 7142 108 | | |
| keywords | | | |
| abbreviations | | | |
| acronymPair1 | Roll | | |
| acronym2 | Pollinator exclusion | | |
| meaning1 | | | |
| acronymPair2 | Vert | | |
| acronym2 | Vertebrate exclusion | | |
| meaning2 | | | |
| keywords | | | |
| keyword1 | Coffea arabica | | |
| keyword2 | coffee pollination | | |
| time | | | |
| timeRange | | | |
| timeFormat | dd MM, yyyy | | |
| startDate | 01.09.2011 | | |
| endDate | 31.12.2012 | | |
| dateEntry | 2012-02-16 | | |
| dateLastModified | 2012-07-25 | | |
| data | | | |
| fileName | 10524_170212_105536_coffee_experiment_metadata.csv | | |
| datasetType | structuredData | | |
| qualityLevel | raw | | |
| dataStatus | incomplete | | |
| dataStructure | | | |
| fileFormat | ASCII | | |
| delimiter | semicolon | | |
| decimalCharacter | . | | |
| header | | | |
| headerRowNO | 1 | | |
| unitRowNO | 0 | | |
| dataRowNO | 2 | | |
| variables | | | |
| variableName | typeOfVariable | unit | descript |
| 1 PlotID | alphanumeric | | cluster 1-3 consist of 4 trees (pollinator exclusion, vertebrate exclusion, open n trees with pollinator exclusion) |
| 2 Cluster | integerNumber | | |
| 3 Treatment | character(16) | | exclusion of pollinators, vertebrates, b |
| 4 Branch | character(16) | | B1 and B2: two different tags per tree and C0: two marked tags (without poll |
| references | | | |
| database | | | |
| papers | | | |
| papers1 | | | |
| comment | | | |

Recycled

XSLT

```

<?xml version="1.0" encoding="UTF-8"?>
<kill:metaProfile xmlns:kill="http://www.biozentrum.uni-wuerzburg.de/schema.xsd" xmlns:kill="http://www.biozentrum.uni-wuerzburg.de/schema.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.w3.org/2001/XMLSchema-instance" kill:authorSource="mixed">
  <kill:authorSource="Intern">10524</kill:id>
  <kill:versionID kill:source="Intern">1.1.12</kill:versionID>
  <kill:lineage kill:source="Intern">
    <kill:title>SP7_Coffee Experiment: Fruitset and Quality</kill:title>
    + <kill:owners>
    + <kill:author>
    + <kill:keywords>
    + <kill:time>
    - <kill:data>
    - <kill:fileFormat>
    - <kill:delimiter>
    - <kill:decimalCharacter>
    - <kill:header>
    - <kill:headerRowNO>
    - <kill:unitRowNO>
    - <kill:dataRowNO>
    - <kill:primaryKey>
    - <kill:header>
    - <kill:dataFormat>
    + <kill:matrix>
    - <kill:variables>
    - <kill:variableName number="1">
    - <kill:variableName number="1">PlotID</kill:variableName>
    - <kill:typeOfVariable number="1">alphanumeric</kill:typeOfVariable>
    - <kill:unit number="1"/>
    - <kill:description number="1"/>
    - <kill:methodology number="1"/>
    </kill:variable>
    - <kill:variableName number="2">
    - <kill:variableName number="2">Cluster</kill:variableName>
    - <kill:typeOfVariable number="2">integerNumber</kill:typeOfVariable>
    - <kill:unit number="2"/>
    - <kill:description number="2">cluster 1-3 consist of 4 trees (pollinator exclusion, verte
    - <kill:methodology number="2"/>
    - <kill:methodology number="2"/>
    </kill:variable>
  
```

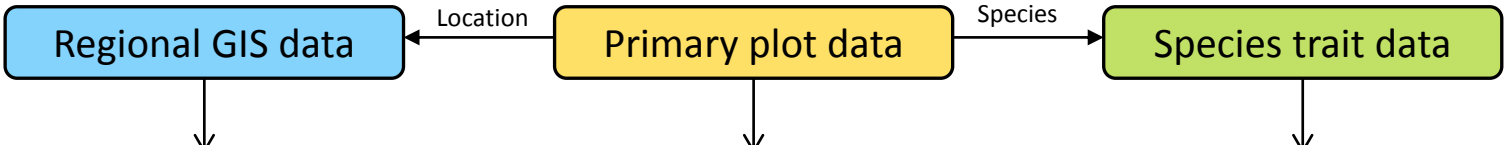


unstructured **Mass amount** **flexible syntax**

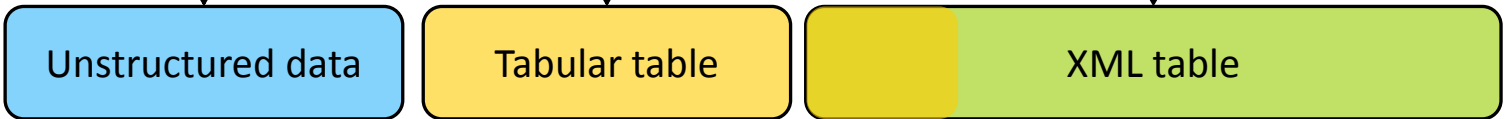
Remote sensing images,
vector geospatial data

Sensor data, experiment data

Processed data



Standardized Metadata (EML based)



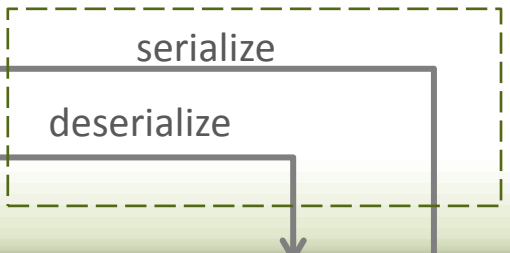
Geoserver

| DATASETID | METADATA | ARCHIVED | VALIDATED |
|-----------|----------|----------|-----------|
| 10360 | XML | ... | N |
| 10380 | XML | ... | N |
| 10363 | XML | ... | N |
| 10346 | XML | ... | N |

```

<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>

```



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

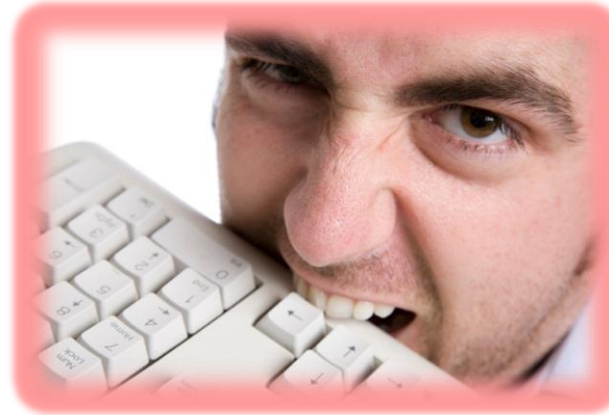
- ❖ KiLi project introduction
- ❖ KiLi central database design
 - ✓ Hopes fulfilled
 - ✓ **Fears conquered**
 - ✓ Challenge accepted
 - ✓ Future Outlook



Hopes

Fears

Challenges



Time effort

Data control

Versioning

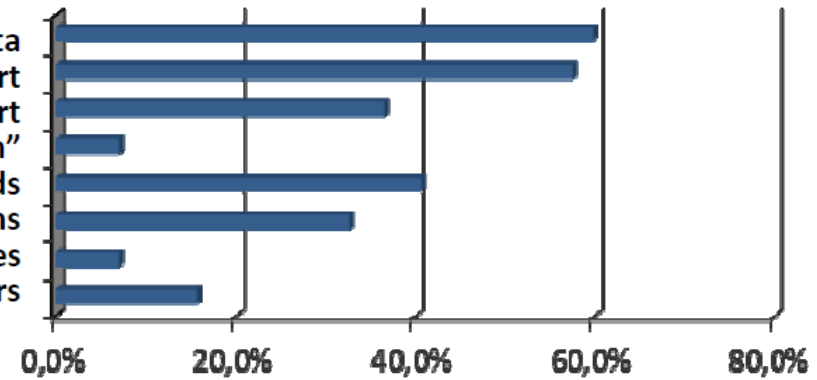
Flexibility

User friendly

Acknowledgement

loss of control over my data
 large time effort
 no acknowledgement for my effort
 "my data aren't good enough"
 missing data standards
 someone draws wrong conclusions
 criticism on my data or analyses
 others

(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

Data sharing policy

Long term storage

Data synthesis

Data linking

Project activities

Communications

Time effort

Data control

Versioning

Flexibility

User friendly

Acknowledgement

Data linking within
different platforms

Permanent data storage

❖ Data linking within different platforms

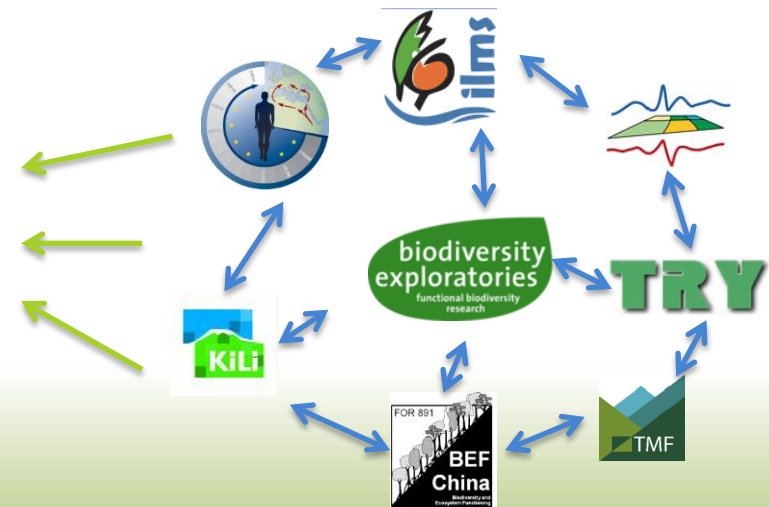
- Metadata standard →
- Semantic search →
- Online access →

- ✓ EML
- ✓ Cooperation between projects
- ✓ Ontology
- ✓ Link of web interface
- ✓ **Central homepage for data integration**



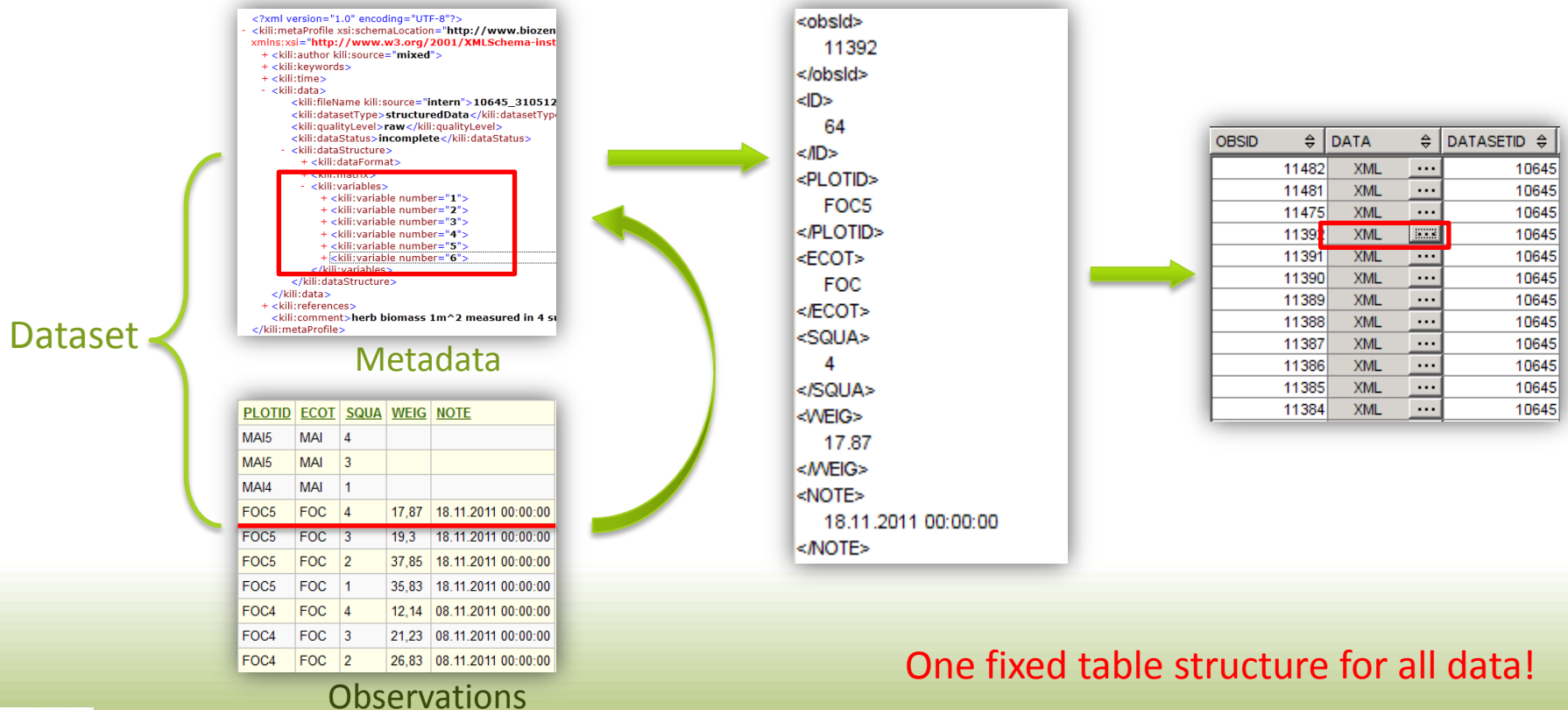
DataONE

LIFEWATCH



❖ Data linking within different platforms

- Transformation of data structure ✓ XML table



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

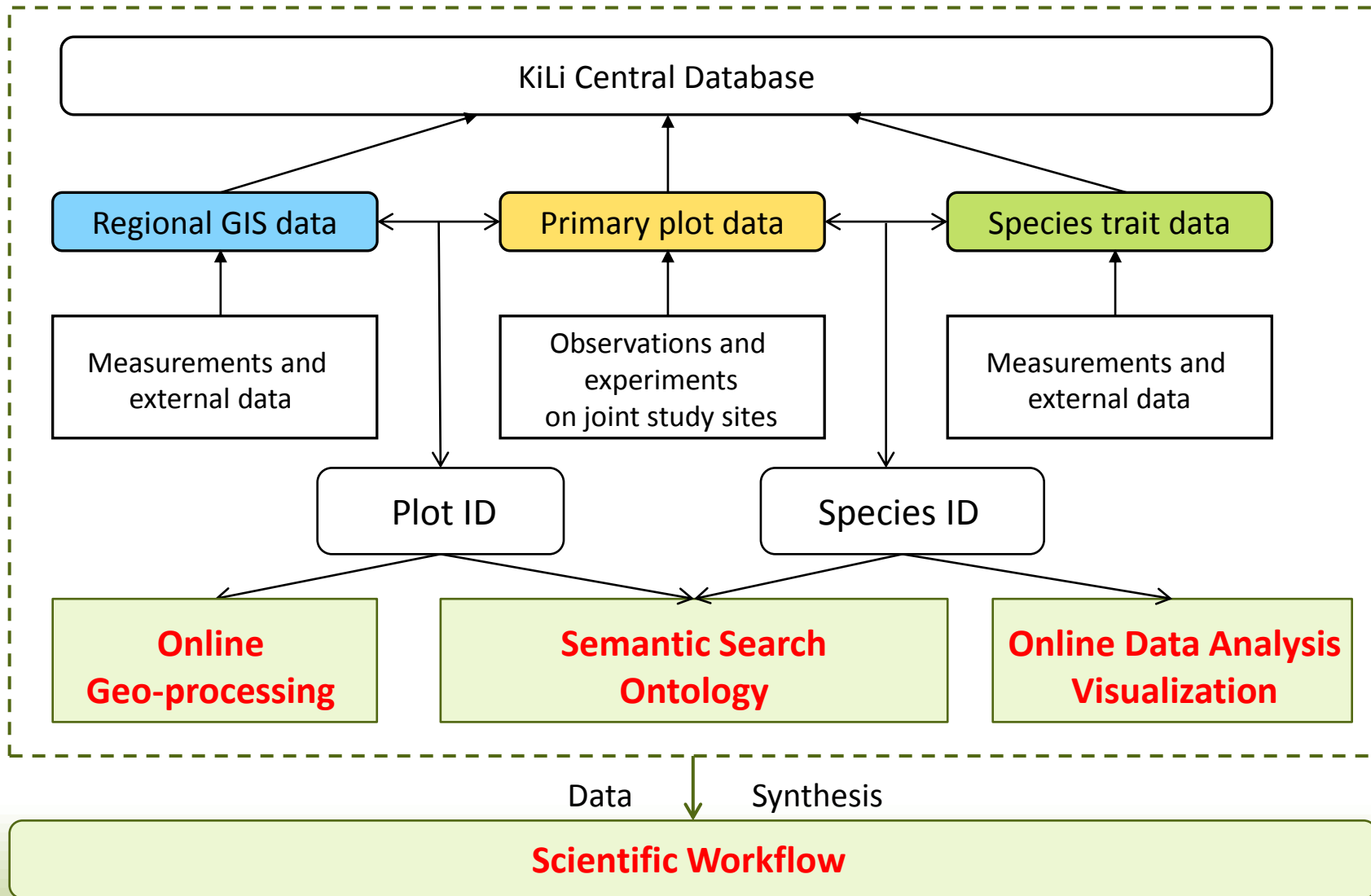


Ecological center



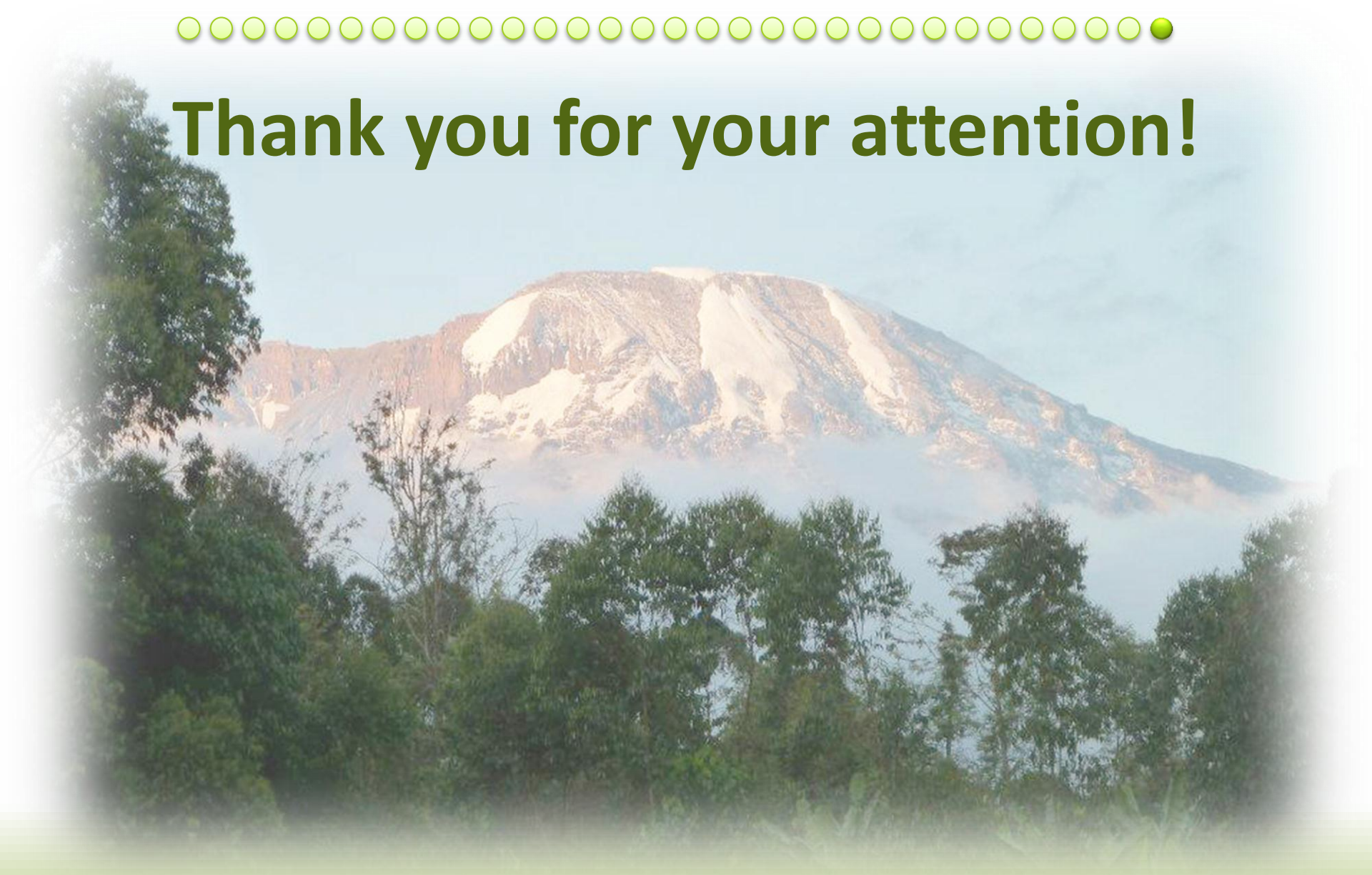
Outline

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





Thank you for your attention!



www.kilimanjaro.biozentrum.uni-wuerzburg.de

GFÖ 2012

Data publishing, linking, using ecological data online

