

**On the way to best practice in
Data Management:
Approaches of the UFZ and the LTER-
Europe network (Long Term
Ecosystem Research)**

Mark Frenzel (Biologist) & Thomas Schnicke (IT)

Best practice elements and goals

- ❑ How to? Defined **workflows**
- ❑ Technical support? Elaborated **software** solutions
- ❑ **Documentation** of data sets
- ❑ Aim? **Re-use** of data
- ❑ Data **publication** (relevant data): **DOI** (Digital Object Identifier)
for data sets

- ❑ Smiling data creators
- ❑ Smiling users

Background UFZ

Initiative started 3 years ago aiming at **DOI**: Science & IT;
development of **workflows, tools** and **attitudes!**

- **UFZ** (1200 people, >30 Departments)
 - Plenty of data
 - Long-term data (e.g. TERENO – Terrestrial Environmental Observatories – Helmholtz project)
 - Engaged in Long Term Ecosystem Research Germany (LTER-D) and Europe (LTER-Europe) networks
- **LTER-Europe** (>20 national networks, >250 sites)
 - Established in 2007
 - Supported by EU NoE AlterNet and other EU projects (EnvEurope, Expeer, eLTER)
 - Networking: Focus on **metadata collection**

The UFZ situation

- ❑ >20 years of data production (exponential growth rates!)
- ❑ **Heterogeneity** of data (person-generated, device-generated; measurement, observation, modelling, data bases)
- ❑ **Documentation** of data sets missing (meta data)
- ❑ **Quality control** (work flow & tools)?
- ❑ **Archiving**

Challenges of today and near future

- Open data
- Search and re-use of data (internal & external)

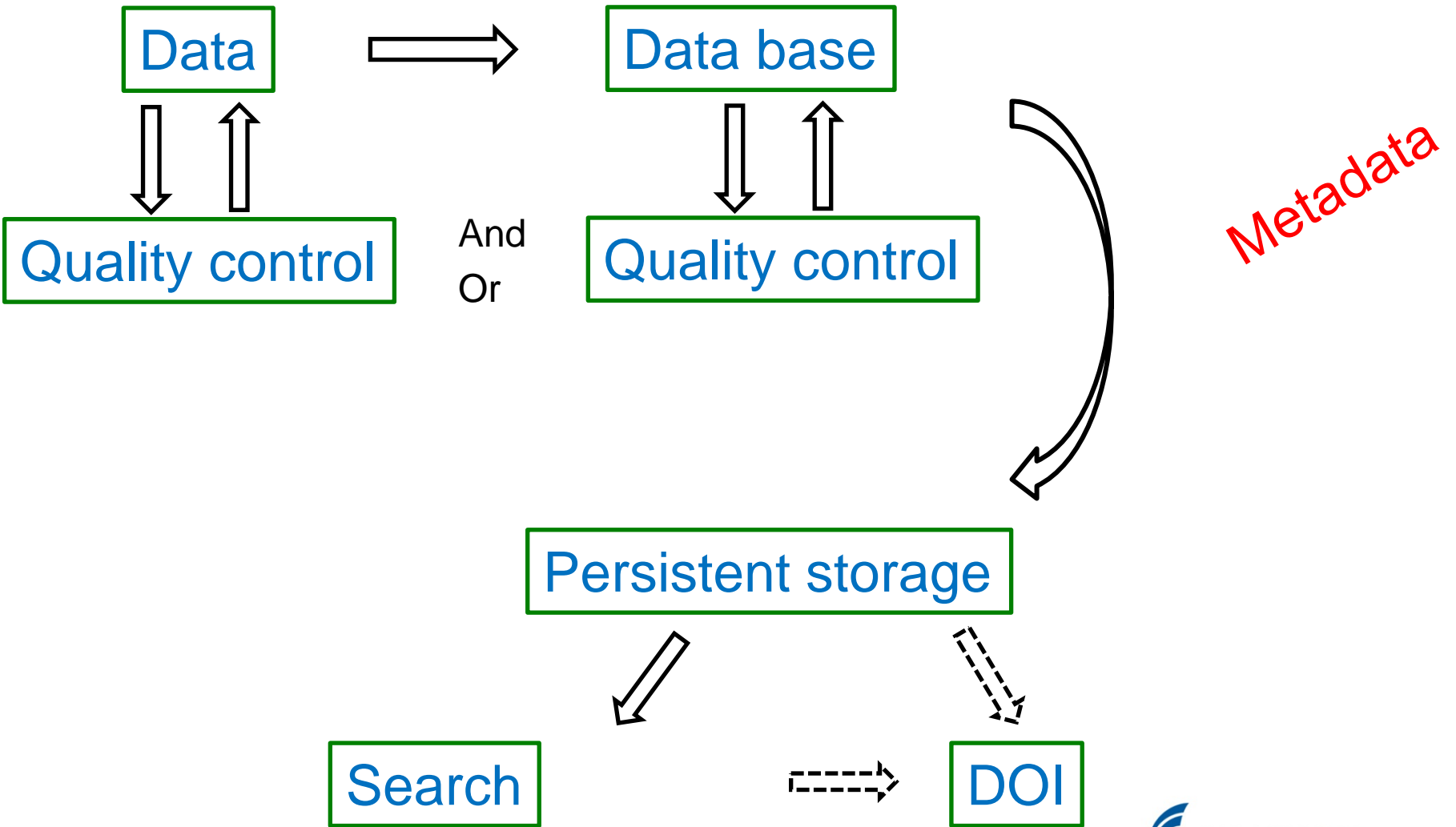
Missing reward system for comprehensive data management (technician, scientist, IT-people)

Looking forward

- ❑ Recognition of importance of data management (attitude)
 - Top down
 - Bottom up
- ❑ Magic tools (easy to handle)
- ❑ Motivated people (data producers)
- ❑ Fitting metadata schemes
- ❑ Linking data to people

Putting pieces together!

Elements of the data management chain



UFZ approach

Serving all kinds of data => Data Management Portal (DMP)

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH - UFZ

UFZ web applications
(internal administration area)

Welcome
(Your session expires in 15 minutes)
[\[extend session\]](#)

Data Management > Data Management Portal

all data projects ▾

Home Sample data Logger data Field management data Archive data Master data

Welcome to the Data Management Portal (DMP)

With this application you are able to **manage** scientific data from different areas:

- ✓ **Archive data** - File based archiving
- ✓ **Sample data** - Sampling and analysis data workflow
- ✓ **Logger data** - Automated transmission of measured values from sensors in the field
- ✓ **Field management data** - Planning and history of management on experimental fields

In case of any question please contact wkdv-datamanagement@ufz.de.

Different data – different requirements (I)

The case of UFZ logger data

Mostly sensors, e.g. weather stations (“Big“ Data) => TERENO

Responsibility taken by IT Department

- Database, interfaces
- Workflow, ...

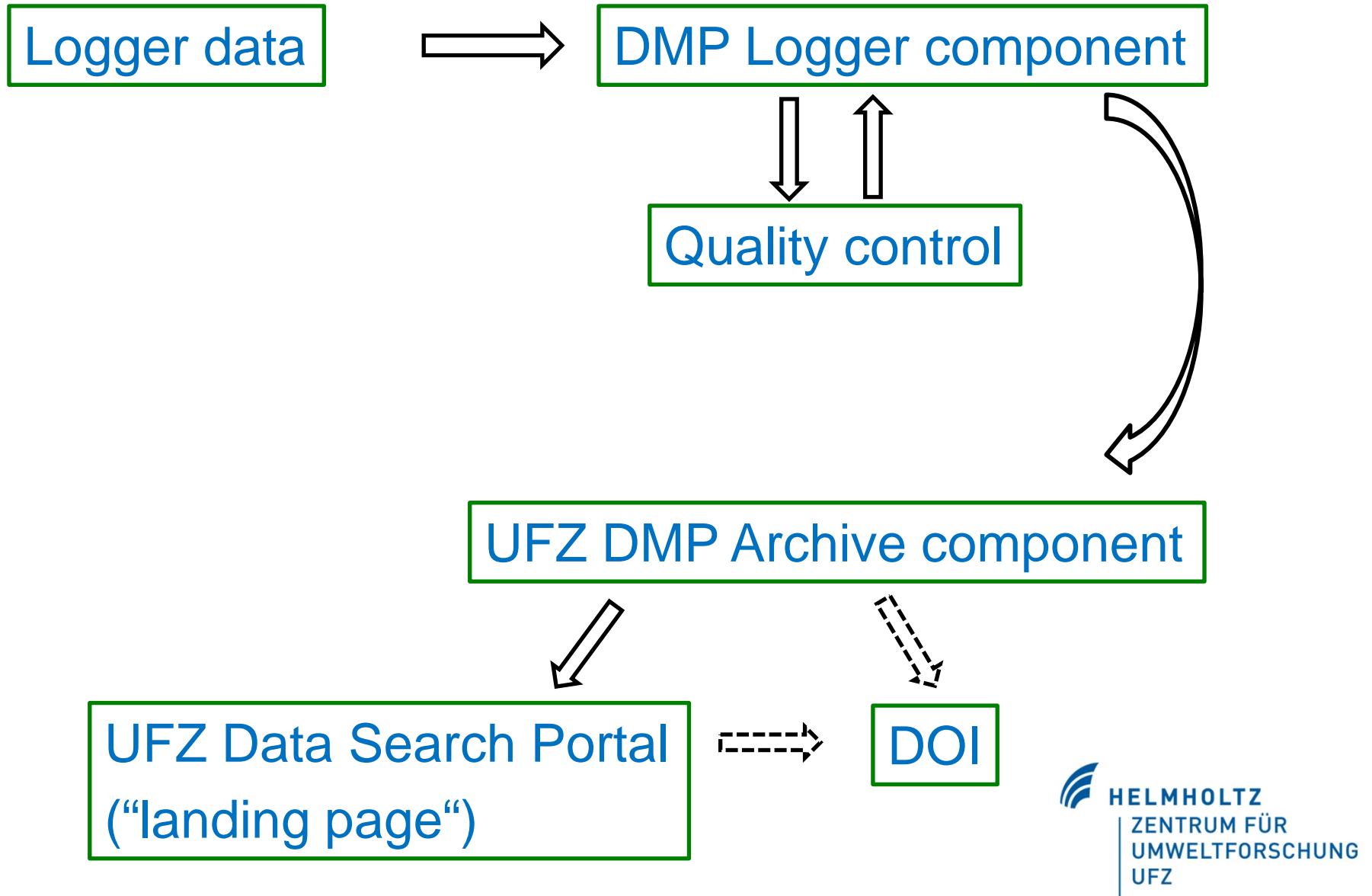
- Level 0 = Raw data
- Level 1 = Transformation to target unit (e.g. V in °C)
- Level 2 = Quality controlled (**checked** and **corrected** for missing data, outliers, trends)

UFZ quality control of logger data (=> DOI)

Priority condition set by head of UFZ

- ❑ Important for use and publication of data
- ❑ Evaluation of different software products
 - No proprietary software
 - Transparent cost model
 - Generic tool, allowing high degree of adaptation to user needs, user-defined algorithms
 - UFZ-decision: training on ORIGIN (data analysis and graphics software)

Workflow logger data UFZ



Different data – different requirements (II)

Example biodiversity data (mostly person-generated data)

IT often refuses to deal with this level of heterogeneity => Logic of ecologists related to data is quite different from IT people

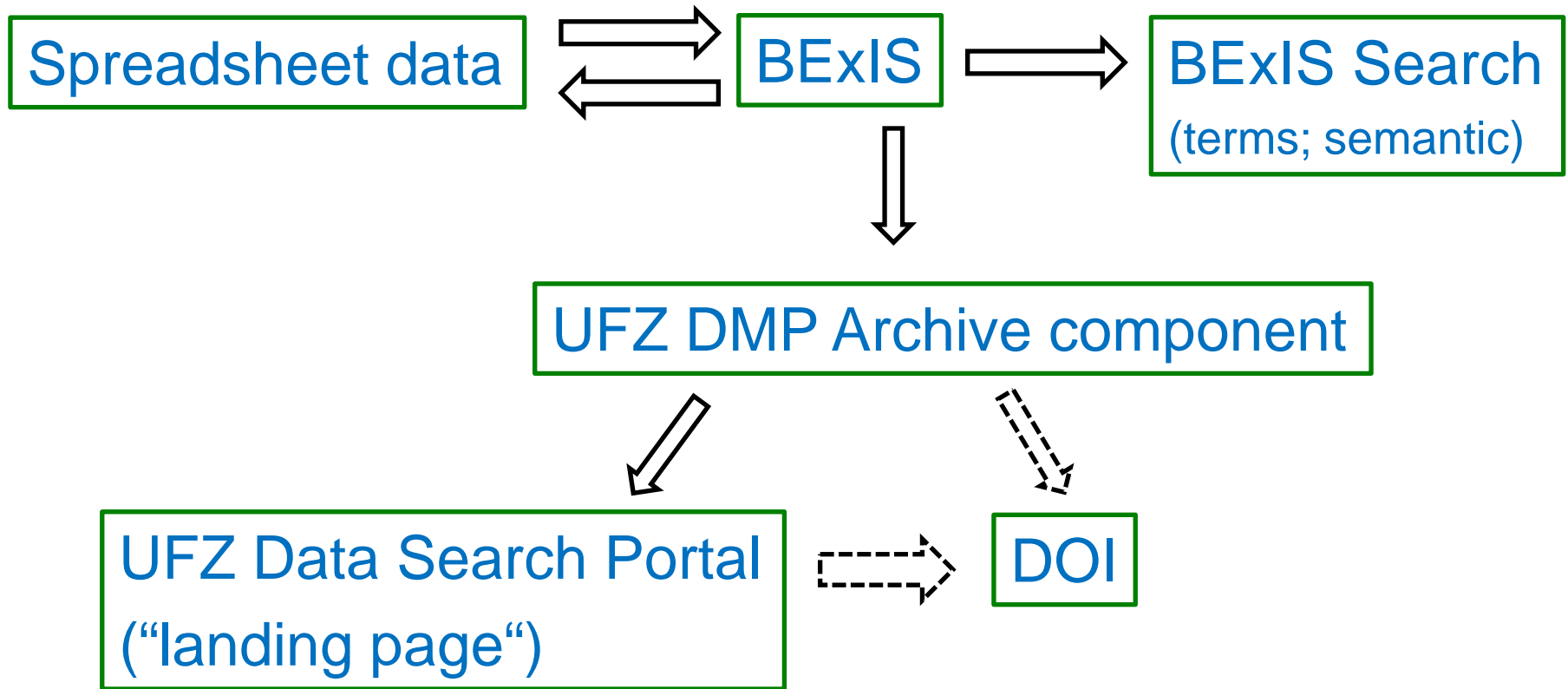
Spreadsheet-based data => Data base?!

- ❑ Plausibility tests
 - Software (e.g. occurrence of species A at location B plausible?)
 - Expert knowledge
- ❑ Technical consistency
 - Correct data types
 - Correct cell entries

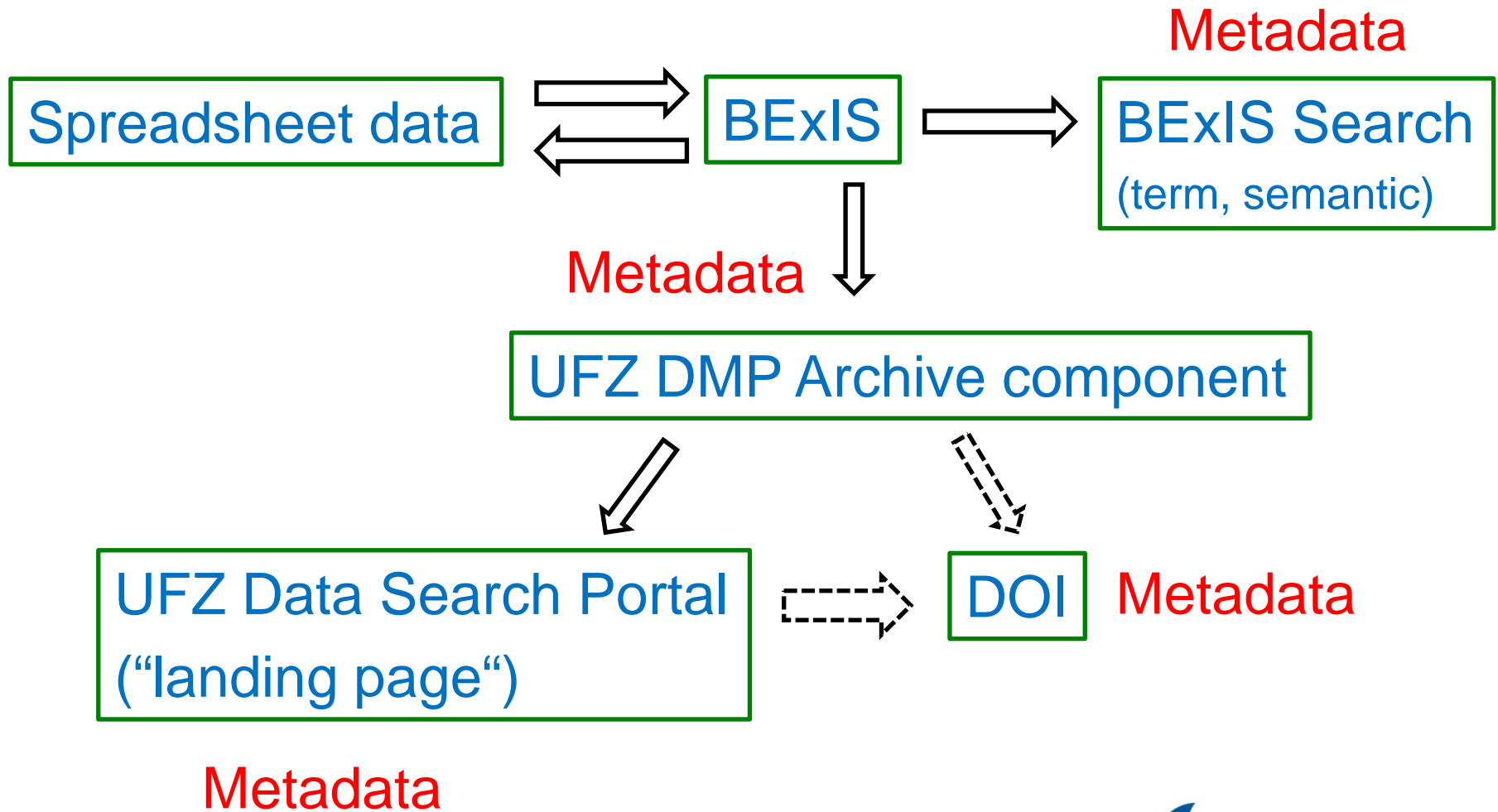
BExIS 2 at UFZ

- ❑ Preferred tool for biodiversity data (examples: TERENO long term bird & bee monitoring)
- ❑ Test case for UFZ DOI-ready data sets
- ❑ Running on virtual machine (since December 2014)
- ❑ Some issues encountered, but appreciated help from BExIS development team
- ❑ Promising future (funded development)

Workflow biodiversity data UFZ



Workflow biodiversity data UFZ



DMP UFZ: Metadata Archive component

Description of data:

| Dublin Core | |
|------------------------------------|---|
| Public ID of the data: | http://www.ufz.de/record/dmp/archive/715 |
| Title: | Luftbild Wanzleben 2003 |
| Subject: | Orthofoto s/w |
| Description: | Gebiet Wanzleben (ca. 5x5 km) Greenveins-Gebiet (2002-2005) TERENO-Gebiet (ab 2009) |
| Coverage: | Wanzleben 2003 |
| Type of data: | picture / image |
| Status of data: | raw data |
| Dataformat: | tif tfw |
| Creators: | Baeßler, Cornelia |
| Source/Origin: | - |
| References to related data: | - |
| Publisher: | Helmholtz-Zentrum für Umweltforschung |
| Rights holders: | Rechte unklar, vermutlich Bundesamt für Kartographie und Geodäsie |
| Contributors: | - |
| Citation: | - |
| Time or Period of creation: | 22.10.2003 |
| Language of the content: | German |
| Target group: | - |
| Change of ownership: | - |
| Teaching method: | - |
| Update method: | - |
| Update frequency: | - |
| Update policy: | - |
| Access permissions: | limited access |

Dublin core elements
Flexibility foreseen

Metadata LTER-Europe: DEIMS

DEIMS = Drupal Ecological Information Management System
(<http://data.lter-europe.net/deims/>)

AVAILABLE RESOURCES

RESEARCH SITE

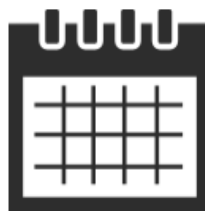


Find out about the European and global network of ecosystem research, monitoring and experimentation sites.

Icon made by Freepik from
www.flaticon.com is licensed under
CC BY 3.0

[read more ...](#)
[discover ...](#)

DATASET



Find out about the available datasets from the European and global ecosystem research, monitoring and experimentation site network.

Icon made by Freepik from
www.flaticon.com is licensed under
CC BY 3.0

[read more ...](#)
[discover ...](#)

PERSON



Find out about the network of European and global researchers dealing with ecosystem research, monitoring and experimentation.

Icon made by Freepik from
www.flaticon.com is licensed under
CC BY 3.0

[read more ...](#)
[discover ...](#)

PUBLICATION



Find out about publication and reports published or contributed by the European and global network of ecosystem research, monitoring and experimentation site network.

Icon made by Freepik from
www.flaticon.com is licensed under
CC BY 3.0

[read more ...](#)
[discover ...](#)

Metadata – which rule to follow?

Elements already implemented in DEIMS

- ❑ INSPIRE directive
- ❑ Dublin core
- ❑ Strategic thinking (e.g. in terms of site meta data)
- ❑ Consider needs for catalogue search and re-use as well
- ❑ Do not overstrain metadata providers...!

Important issues

Make life easier for metadata providers (multitude of datasets),
e.g. by

- ❑ Enable copying of metadata
- ❑ Thesauri (controlled vocabulary)

DEIMS: Metadata for datasets

Documentation: <http://data.lter-europe.net/deims/dataset/documentation>

1. TITLE
2. IDENTIFIER
3. CREATOR AND CONTACT POINTS
4. METADATA PROVIDER
5. METADATA DATE
6. PUBLICATION DATE
7. LANGUAGE
8. ABSTRACT
9. KEYWORD SET
10. ACCESS AND USE CONSTRAINTS
11. INTELLECTUAL RIGHTS
12. ONLINE DISTRIBUTION
13. GEOGRAPHIC BOUNDING COORDINATES
14. GEOGRAPHIC BOUNDING ALTITUDES OR DEPTHS
15. TEMPORAL EXTENT
16. TAXONOMIC COVERAGE
17. METHODS DESCRIPTION
18. INSTRUMENTATION DESCRIPTION
19. SAMPLING DESCRIPTION
20. LEGAL OBLIGATION REPORTING

Coming back to BExIS – wishlist and outlook

- ❑ High flexibility in metadata creation and editing
- ❑ Extend supported data format to ODF
- ❑ Allowing harvesting of metadata
- ❑ Link to thesauri (controlled vocabularies, e.g. EnvThes => LTER-Europe)
- ❑ Connection to Linked data (<http://linkeddata.org/>) - Connect Distributed Data across the Web
- ❑ What are others doing? E.g. US NEON (National Ecological Observatory Network) data portal (<http://data.neoninc.org/home>)

Bring all the pieces together, make use of what is already there!

Final slide: DOI-Perspective

- ❑ UFZ: approaching DOI publication agent status
- ❑ DataCite / Helmholtz **Open Science** Workshop
„Datenpublikationen mit DOIs - persistent und zitierbar“ (12 Jan 2016, UFZ Leipzig)

