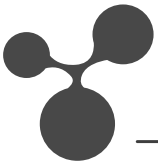
Technische Universität Dresden
Medienzentrum

Prof. Dr. Thomas Köhler
Dr. Nina Kahnwald
(Hrsg.)



GeNeMe '12

Gemeinschaften in Neuen Medien

an der
Technischen Universität Dresden

mit Unterstützung der

BPS Bildungsportal Sachsen GmbH
Campus M21
Communardo Software GmbH
Dresden International University
Gesellschaft der Freunde und Förderer der TU Dresden e.V.
Hochschule für Telekom Leipzig
IBM Deutschland
itsax - pludoni GmbH
Kontext E GmbH
Medienzentrum, TU Dresden
Webdesign Meier
SAP AG, SAP Research
T-Systems Multimedia Solutions GmbH

am 04. und 05. Oktober 2012 in Dresden

www.geneme.de
info@geneme.de

E.5 Identifiers in e-Science platforms for the ecological sciences

Karin Nadrowski, Daniel Seifarth, Sophia Ratcliffe, Christian Wirth, Lutz Maicher
Universität Leipzig, Spezielle Botanik und Funktionelle Biodiversität

Abstract.

In the emerging Web of Data, publishing stable and unique identifiers promises great potential in using the web as common platform to discover and enrich data in the ecologic sciences. With our collaborative e-Science platform "BEFdata", we generated and published unique identifiers for the data repository of the Biodiversity – Ecosystem Functioning Research Unit of the German Research Foundation (BEF-China; DFG: FOR 891). We linked part of the identifiers to two external data providers, thus creating a virtual common platform including several ecological repositories. We used the Global Biodiversity Facility (GBIF) as well the International Plant Name Index (IPNI) to enrich the data from our own field observations. We conclude in discussing other potential providers for identifiers for the ecological research domain. We demonstrate the ease of making use of existing decentralized and unsupervised identifiers for a data repository, which opens new avenues to collaborative data discovery for learning, teaching, and research in ecology

Keywords: *Life Science Identifier, Web of Data, ecology, e-Science, data management, web application, Ruby on Rails, BEF-China, Topic Maps, scientific species names, Ecological Metadata Language*

1 Introduction

Within the few last years the Web has evolved from a global information space of linked documents to one where both documents and data are linked [1]. By radically improving information retrieval, this emerging Web of Data will generate new types of applications and tools promoting better informed decisions and data synthesis for the public as well as for science [2]

In the field of ecological learning, teaching, and research, the growing resource of web-based data offers new opportunities and challenges for data discovery and syntheses. One of the most obvious and basic challenges is to find suitable data. This not only applies when looking for data but also when ecologists want to increase the

visibility of their own data. Being able to link data resources opens up new avenues in data exploration that were not easily accessible when one had to search the whole of the Web in order to find them [3].

Identifiers are crucial to naming and referencing of discrete entities in data. Well known examples of identifiers include ISBN numbers for books, or Digital Objects Identifiers [4] for content objects such as journal articles or data sets. Identifiers can also refer to more conceptual objects such as vocabulary terms, keywords, or metadata structures [5]. Identifiers enable data entities to be linked explicitly and implicitly [6]. Explicit linkage occurs whenever two entities are linked through their identifiers, i.e. by RDF triples [7]. Implicit linkage occurs when different, not obviously related datasets, use the same identifiers. Once these data sources are connected the information from both sources can be merged. As in the whole web, implicit linkage is decentralized and unsupervised and it is this absence of two-sided approval in implicit linkages that drives the fast growth of the global Web.

To make identifiers usable for humans, they must be enriched with hypertext so that humans can read the information related to the identifier and use the links to explore related data. To enable machine reading the data must additionally be actionable. For this it must be in a standard form and self resolving.

Here we demonstrate in a use-case how decentralized and unsupervised identifiers can aid data publication and exploitation in ecology. For this we use an e-Science platform[1] for collaborative ecological data management, which is used in an ecological repository[2] (described in more detail below). We demonstrate publishing identifiers for its data and metadata, and use identifiers in external providers to enrich our original data repository.

The remainder of this paper is structured as follows: in the next section we examine existing identifier schemata that are widely used in ecological research, with a focus on organism names. In the third section we introduce the BEFdata platform, a web-based data repository in the biodiversity ecosystem functioning domain. In section four we describe how valid life science identifiers resources are created in the BEFdata instance used in the BEF-China project. In section five a scenario for applying identifiers is described, by using them in the retrieval of species names from two ecological data resources. Additionally we present a list of further data providers, which could be used in a similar way for a diverse range of data types. In the last section we discuss further applications of the tools demonstrated in this paper.

1 BEFdata, https://github.com/befdata/befdata

2 http://china.befdata.biow.uni-leipzig.de

2 Identifiers for organism names

Almost any research or applied project in ecology deals with the scientific names of organisms. However, categorizing organisms into hierarchical taxonomic structures remains an active field of research and thus is subject to scientific progress and change. In addition, different research groups use different concepts for categorizing the organisms in questions, leading to different names for the same type of organism. Furthermore, scientific species names are long and since the overwhelming majority of ecological data is human-entered, misspellings and abbreviations are frequent, increasing the probability of data-mismatch [8], [9].

There are numerous initiatives that aim to normalize the use of species names. For example, the Global Names Index (GNI) [9], [10] currently lists 38 websites that index names of organisms. The GNI stores any name for an organism, including synonyms and misspellings, and offers to reconcile these names. It is an open source development and provides an API to search for names. The API returns a list of scientific species names, but it does not include further information on the species, as for example its taxonomic tree.

GBIF [11] is an international portal for biodiversity data with an emphasis of species occurrences in natural history collections. It is one of the initiators of GNI and also offers services related to names of organisms. GBIF offers web services that currently return 6 types of data: taxon data; occurrence record data; occurrence density data; dataset metadata; data publisher metadata; and data network metadata. The taxon data service returns a unique key for a given name reconciled against known synonyms. Using GBIF it is possible to retrieve a taxonomic tree for the species name.

The International Plant Name Index (IPNI) [12] is a database of plant names and associated references. IPNI defines Life Science Identifiers for the plant names and is also included in GNI. When searching for plant names GBIF returns the LSIDs from IPNI, which facilitates the retrieval of information held by IPNI.

The providers described above offer services to retrieve data, however additional data can be retrieved by crawling the according web pages. For example, the Life Science Identifer of IPNI can be retrieved from webpages of GBIF in this way. Other initiatives that work with species names offer a wealth of less structured data, examples of which are the regional floras [13]. One of these floras, the Flora of China (http://www.efloras.org), works with taxon identifiers that allow access to pages containing information on plant families, genera, as well as species. However, the pages are in plain text and would need text mining tools to retrieve structured information on the species. A given species name can be defined by one or more different identifiers by each of the web pages above (Table 1).

Table 1. Different identifiers are used to access a given species name (*Acer amplum*, a tree species occurring in China) in different data resources. IPNI is the International Plant Name Index and GNI the Global Names Index.

Data resource	Identifier
Flora of China	http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=200012920
IPNI	urn:lsid:ipni.org:names:56603-1
GNI	http://gni.globalnames.org/data_sources/19?search_term=acer+amplum&commit=Search returns: Acer amplum, Acer cappadocicum subsp. amplum

In the following section we demonstrate how an ecological data portal can utilize identifiers to publish its own resources using Life Science Identifiers and how external identifiers can be used to access information that enrich the data within the repository.

3 The BEFdata platform

The BEFdata portal [14][3] is an open source, web-based e-Science platform implemented in Ruby on Rails [15]. It stores raw data values coming from tabular data. It is currently used by three projects in the field of Biodiversity-Ecosystem Functioning research (BEF), the German and the Chinese projects of the BEF-China experiment[4] and FunDivEUROPE[5].

BEF-China, a research unit of the German Science Foundation (DFG, FOR 891), is the largest forest biodiversity experiment in the world and situated in subtropical China[6]. More than 150.000 trees have been planted in plots with different levels of diversity. Additionally, research is conducted in comparative study plots in natural forests [16]. The European projects of the FOR 891 research unit and their Chinese partner projects use a separate instance of the BEFdata platform. FunDivEUROPE[7] is a European-wide BEF research project, comprising of 24 partner institutions from 15 countries. The project includes a specifically designed network of comparative plots in mature forests across Europe, selected to cover a diversity gradient, whilst all other variables are kept as constant as possible. FunDivEUROPE additionally includes the existing tree diversity experiments, TreeDivNet[8].

The BEFdata platform captures standard metadata on the scope of the data set,

3 https://github.com/befdata/befdata

4 http://china.befdata.biow.uni-leipzig.de, http://159.226.89.107

5 http://fundiv.befdata.biow.uni-leipzig.de

6 http://www.bef-china.de

7 http://www.fundiveurope.eu/

8 http://www.treedivnet.ugent.be/index.html

methods used to obtain the data, and originator and ownership information. For the public the metadata is available on web pages as well as through a RESTfull interface in Ecological Metadata Language (EML, [17]). BEF research uses a wide range of research objects, ranging from ecosystem components such as a vegetation layer, to organisms, and parts of organisms, and genetic sequences, to the description of ecosystem nutrient stocks and flows. The overwhelming majority of the data are human entered [18], [19]. There is a recent initiative to conceptualize the structure of such data in an observation data model [20], [21]. Observations are measurements or facts observed simultaneously on the same object.

The BEFdata platform provides processes for the upload, cleansing and distribution of research data. In a first step, uploaded data values are validated against naming conventions, referred to as "categories". For performance and privacy reasons it was decided not to publish unique and stable identifiers for each single raw data value, however, publishing categories is central to our intention to link the content of the data portal to external information sources. Other objects selected for publication include "users", "projects", "datasets", "datacolumns", "datagroups", "tags", "observations", "categories", each represented by an object model and database table (see Figure 1 for the data model). Data columns describe the columns in the spreadsheet containing the primary research data, while observations describe the rows.

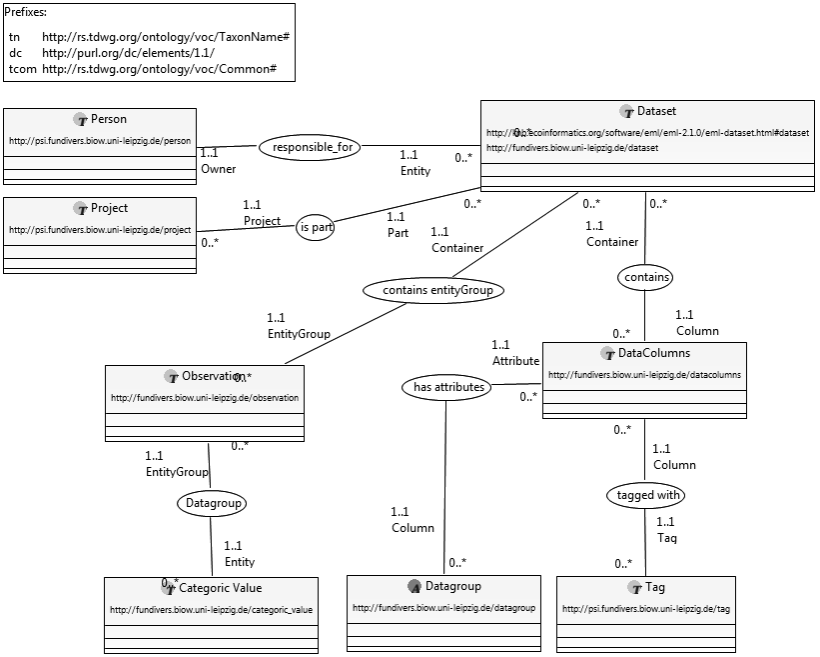


Fig. 1. The abstract data model of the BEFdata platform. The names and links of the entities are according to the observation ontology and describe scientific observations on research objects. Data in ecology is typically saved in tabular form with rows holding different measurements or facts about one object and columns holding data entries used by applying a specific methodology. We have developed tools to publish data and metadata from the BEFdata portal using unique and resolvable identifiers, including Life Science Identifiers [22].

4 Generating Life Science Identifiers

In BEFdata, unique and stable identifiers have to be generated for each new entity. For this purpose, the schema of Life Science Identifiers is one possible choice [23] and is recommended by GBIF [24]. In this section we demonstrate the generation of LSIDs in an instance of BEFdata, used by the European research projects of BEF-China. Life Science Identifier (LSID) take the form urn:lsid:authority. org:namespace:object:(revision) [24]. The revision part of the LSID is optional and not implemented here as BEFdata currently does not track data provenance. The authority part has to be chosen with care as it must persist with change to the institute

name or project funding phases. This paper only demonstrates the generation and usage of LSIDs, so for this reason we currently do not provide a persistent domain name for the publication of the BEF-China objects but use the domain name of the BEF-China instance of the portal.

To generate the namespace part of the LSID we use the Ruby object names as the namespace. To generate the object part we use a MD5 algorithm [25] to construct hash sums from the object instances. Using hashes leads to opaque object identifiers, that do not contain any information about the referenced data object [24].

To achieve resolvability of the LSIDs within the BEFdata application, we use the Rails routes architecture. The method "subject_identifier" generates a LSID for each object in the data and can be used in the Rails application to add the LSID to the object view. It is important to emphasize that the generated LSIDs are unique, stable and resolvable, but not shared with any other application or data-set. To achieve this goal, the generated LSIDs should be reconciled with identifiers from other resources and applications. As we have shown in previous work [6] this approach of "semantic handshakes" allows both, full control about stable and unique identifiers at application level, and harmonization of identifiers in an ecosystem of interlinked applications.

5 Data augmentation through web services on species names

Since Ecology is an interdisciplinary field, there are a multitude of web services available. Many of these services have different foci and use, if at all, highly divergent APIs. The most efficient technology to use information from such online resources are Mediators, combined with wrappers [26], which have to be programmed separately for each service. These programs are here implemented as plugins that can be configured for a particular object with the application configuration file. This offers the flexibility to exchange plugins between instances of BEFdata and between similar Rails applications.

We use scientific species names as an example of harvesting information from external sources. We use two different authorities when working with scientific species names: GBIF and IPNI. Both offer a scientific species name search, for example, the species "*Acer amplum*" can be looked up in both services. Further information such as the relatedness of this species to other species can be collected. Our GBIF plugin has the following process flow: a wrapper issues a request to GBIF passing the species name; GBIF retuns a Taxon Concept Object Class, which includes an IPNI identifier for the species. The IPNI identifier can then be used to harvest further information on the object from IPNI.

Ecology is an integrative discipline, often requiring interactions with adjacent fields of research, such as nature conservation, biodiversity, evolution, genomics, geography, or climatology [2]. It can thus make use of several initiatives that have published resources in the web. In the following, we will present a list of services that could

be useful for the types of identifiers published using the tools presented in this paper (Table 2). It is often useful to distinguish between identifiers that refer to entities such as names of species, people, projects, or published scientific concepts at the one hand, and conceptual resources such as container types for entities and relationships between them [27]. The eight main entity types of the BEFdata platform (e.g.: Person, Dataset, Datagroup, Category) as well as the seven relationships between them represent conceptual resources. They are published as resolvable unique identifiers. Additionally, each of the entity types contains a finite number of instances, which represent the entities in the BEFdata portals. These are published as LSID.

Data published online can be human readable only, but it increases their usefulness to be also machine readable. For the purpose of presenting the online resources in Table 2 we make a distinction between four types, given in the order of how difficult it is to extract information: 1) resolvable and machine readable, 2) resolvable as structured web page, which can be accessed by crawling, 3) resolvable as unstructured web page. In the latter case an identifier exists, but the information retrieved is largely unstructured text that is only human-readable. As a last type we refer to 4) not resolvable resources (i.e. a definition in a Word document). In Table 2 these resource types are denoted as *, **, ***, and **** respectively.

Table 2 shows that there are initiatives that offer different types of resources. The US Long Term Ecological Research network (LTER) consists of independent research sites in the US. Data repositories are managed for each site independently in the LTER network. The Web Services Working group therefore aimed at providing services using the Representational State Transfer (REST; Fielding & Taylor, 2002) style available for all repositories independent of the implementation of the research site data repository [29]. Currently, they provide a unit registry[9] that not only provides a simple user interface, but also translates queries issued through the interface to REST requests. LTER has adopted a controlled vocabulary that is also available as web Service [30]. The controlled vocabulary was developed within the TemaTres service[10]. Web Services developed around this vocabulary can be used to discover terms[11] or download the list of prefered terms[12]. As another prominent example, GBIF offers several RESTful services to access data related to the ecological disciplines. Table 2 gives examples of further initiatives, without claiming to be complete.

9 http://unit.lternet.edu/unitregistry

10 http://vocab.lternet.edu

11 http://vocab.lternet.edu/webservice/keywordlist.php

12 http://vocab.lternet.edu/webservice/preferredterms.php

Table 2. Data as well as metadata stored in BEFdata instances and made accessible using the tools described in this paper could further be related to the following resources, without claiming to be complete. We chose resources to be related to the ecological disciplines. There are providers that offer several services at once, two of which we have highlighted in a separate column: The US Long Term Ecological Researchnetwork (LTER), and the Global Biodiversity Information Facility (GBIF). The stars corresponds to the increasing effort needed to retrieve data: * – machine readable; ** – structured web page; * – unstructured web page; **** - versioned offline.**

BEFdata resources (Topics)	LTER	GBIF	other providers
Logical resources			
Standards for data formatting		**1	**2, ****3
Entities			
Projects		*4	
Datasets	**5	*6	**7, **8
Data			
- species names		*9	*10
- gene accession numbers			*11
- plant traits			**12
- habitat types			**13
- geographic locations		*14	
- units of measurements	*15		*16
Keywords	*17		****3, **12, *18, **19

1) http://www.tdwg.org/standards – GBIF uses standards developed by the Biodiviersity Information Standards (TDWG).
2) http://knb.ecoinformatics.org/software/eml/eml-2.1.0/index.html - Ecological Metadata Language
3) Extensible Observation Ontology [20]
4) http://data.gbif.org/ws/rest/provider, http://data.gbif.org/ws/rest/network– RESTfull service for data providers and networks, including their datasets
5) http://metacat.lternet.edu/das/lter/browse.jsp

6) http://data.gbif.org/ws/rest/resource
7) http://datadryad.org- Dryad, a repository of datasets published in journals as well as further repositories.
8) http://www.pangaea.de– Data publisher for Earth and Environmental Science
9) http://data.gbif.org/ws/rest/taxon – RESTfull service giving back the taxonomic tree associated to species as well as data providers for specimen of that species.
10) http://gni.globalnames.org– Global Names Index [9]
11) GeneBank® [31]
12) http://www.try-db.org– Data can be requested after submitting plant trait data to the TRY initiative [32]
13) http://eunis.eea.europa.eu/habitats.jsp– Habitat search interface of the European Environment Agency
14) http://data.gbif.org/ws/rest/occurrence– RESTfull service for occurences of individuals, including latitude and longitude. http://data.gbif.org/ws/rest/density– returns the count of occurences for one degree cell of latitude and longitude.
15) http://unit.lternet.edu/unitregistry
16) http://www.bipm.org/en/si– The International System of Units
17) http://vocab.lternet.edu/webservice/keywordlist.php, http://vocab.lternet.edu/webservice/keywordlist.php
18) http://id.loc.gov/authorities/subjects.html - Library of Congress subject headings
19) http://biodiversity-chm.eea.europa.eu/thesaurus– Thesaurus of the Convention of Biological Diversity from the website of the European Environmental Agency

6 Outlook

In this paper we have shown how decentralized and unsupervised identifiers can be used not only to publish resources from an ecological repository, but also how external identifiers can be used to enrich data within the repository. In our example, scientific names for plant species were published as Life Science Identifiers (LSID) and subsequently enriched by two web services offering information on these species. By doing so we could extract a taxonomic tree for each species as well as reconcile synonyms. However, scientific plant species names only represent a small part of the LSIDs published. Indeed, scientific plant species names are only instances of one of our object types (“Datagroup”, see Figure 1). The two instances concerned were called “Tree species reference list” and “Synonyms of tree species”. However, it is well possible that other groups may decide to utilize other web services for species names. In addition, other Datagroup instances may profit from web services

that do not deal with species names. For this reason we configured the instantiation of the tools for mediators and wrapper in form of plugins at a high level of the rails application. This makes the tools presented here exchangeable and extensible.
We consider this flexibility as a strength of the architecture, since it means that enriching a BEFdata application is not dependent on the domain logic. For example, the BEF-China and the FunDivEUROPE research group use different instances of the BEFdata platform and make use of scientific plant species names as raw data values. Scientific plant species names are validated against naming conventions within each instance. They may or may not decide to use the GBIF or IPNI Web service to extend their data. However, it is also possible that BEF-China decides to use identifiers given in the Flora of China instead, or to use several at once. The BEFdata instances could additionally use each others Datagroups through web services.
Although the BEFdata portal currently does not offer to manage data provenance, publishing identifiers for datasets, data columns and observations opens up new ways of exchanging and extending scientific analysis using workflows [33], [34]. For example, the Kepler software [34] is a visual workflow editor for research processes, from from downloading data that to using the R statistical software. Publishing data together with these workflows will make the process of scientific analysis more transparent and easier to learn.
In learning and teaching ecology BEFdata platforms can help to illustrate best practices in data management, including storage of data in a standard metadata format (EML). BEFdata platforms additionally allow to link to existing governmental or non-governmental agencies that provide controlled vocabularies, including libraries, museums, or citizen science projects. Many of these organizations offer web services to utilize their resources (see Table 2 for examples). Thus students can explore existing efforts to unify naming conventions and directly use them in a BEFdata platform for their own data.
In this paper we have shown how resolvable identifiers can help to improve data quality progress by linking own data to online available data from other repositories. In addition, identifiers can help improve scientific progress by making data use and analysis more transparent.

Acknowledgments. The BEFdata portal was developed within the BEF-China project by the German Science Foundation (DFG, FOR 981, sub-project “Data management”) of C.W. K.N. was supported by the same project. S. R. was supported by the EU project FunDivEUROPE (265171, Work package 1, Task I.4 “Data management, data quality assessment and control”) of C. W. We are grateful to the Topic Map Lap (BMBF, 03F0109, 03F01092) for providing the computational infrastructure of publishing some of our structured data.

References

[1] C. Bizer, T. Heath, and T. Berners-Lee, “Linked Data - The Story So Far,” International Journal on Semantic Web and Information Systems, vol. 5, no. 3, pp. 1-22, Jan. 2009.

[2] O. J. Reichman, M. B. Jones, and M. P. Schildhauer, “Challenges and opportunities of open data in ecology.,” Science (New York, N.Y.), vol. 331, no. 6018, pp. 703-5, Feb. 2011.

[3] T. Berners-Lee, “Linked Data,” 2009. [Online]. Available: http://www.w3.org/DesignIssues/LinkedData. [Accessed: 04-Feb-2012].

[4] DOI, “The DOI System,” 2012. [Online]. Available: http://www.doi.org/. [Accessed: 04-Feb-2012].

[5] G. Dunsire, D. Hillmann, J. Phipps, and K. Coyle, “A Reconsideration of Mapping in a Semantic World,” in International Conference on Dublin Core and Metadata Applications 2011, 2011, pp. 26-36.

[6] L. Maicher, “The Impact of Semantic Handshakes,” in Leveraging the Semantics of Topic Maps, L. Maicher, A. Sigel, and L. Garshol, Eds. Berlin / Heidelberg: Springer, 2007, pp. 140-151.

[7] F. Belleau, M.-A. Nolin, N. Tourigny, P. Rigault, and J. Morissette, “Bio2RDF: towards a mashup to build bioinformatics knowledge systems.,” Journal of biomedical informatics, vol. 41, no. 5, pp. 706-16, Oct. 2008.

[8] R. Page, “Using Google Refine and taxonomic databases (EOL, NCBI, uBio, WORMS) to clean messy data,” iPhylo blog, 2012. [Online]. Available: http://iphylo.blogspot.com/2012/02/using-google-refine-and-taxonomic.html. [Accessed: 07-Feb-2012].

[9] D. J. Patterson, J. Cooper, P. M. Kirk, R. L. Pyle, and D. P. Remsen, “Names are key to the big new biology.,” Trends in ecology & evolution, vol. 25, no. 12, pp. 686-91, Dec. 2010.

[10] GBIF and EOL, “Global Names Index,” 2012. .

[11] GBIF, “Global Biodiversity Information Facility,” 2012. [Online]. Available: http://www.gbif.org. [Accessed: 27-Jan-2012].

[12] IPNI, “International Plant Names Index,” 2012. [Online]. Available: http://www.ipni.org/. [Accessed: 04-Feb-2012].

[13] eFloras, “eFloras.org,” 2012. [Online]. Available: http://www.efloras.org/. [Accessed: 04-Feb-2012].

[14] K. Nadrowski et al., “Harmonizing, annotating, and sharing data in biodiversity-ecosystem functioning research,” Methods in Ecology and Evolution, vol. submitted, no. MEE–12-03-107, 2012.

[15] S. Ruby, D. Thomas, and D. Heinemeier Hansson, Agile Web Development with Rails. Raleigh, North Carolina: The Pragmatic Bookshelf, 2011.

[16] H. Bruelheide et al., “Community assembly during secondary forest succession in a Chinese subtropical forest,” Ecological Monographs, vol. 81, no. 1, pp. 25-41, Feb. 2011.

[17] E. H. Fegraus, S. Andelman, M. B. Jones, and M. Schildhauer, “Maximizing the value of ecological data with dtructured detadata: an introduction to Ecological Metadata Language (EML) and principles for metadata creation,” Bulletin of the Ecological Society of America, vol. 86, pp. 158-168, 2005.

[18] T. Lotz, J. Nieschulze, J. Bendix, M. Dobbermann, and B. König-Ries, “Diverse or uniform? - Intercomparison of two major German project databases for interdisciplinary collaborative functional biodiversity research,” Ecological Informatics, Jan. 2012.

[19] B. Michener, “DataOne: changing community practice and transforming the environmental sciences through access to data and tools,” in Data repositories in environmental sciences – concepts, definitions, technical solutions and user requirements, 2011, vol. 2.

[20] J. S. Madin, S. Bowers, M. P. Schildhauer, and M. B. Jones, “Advancing ecological research with ontologies,” Trends in ecology & evolution, vol. 23, no. 3, pp. 159-168, 2007.

[21] M. Schildhauer et al., “Using observational data models to enhance data interoperability for integrative biodiversity and ecological research,” in Data repositories in environmental sciences – concepts, definitions, technical solutions and user requirements, 2011.

[22] D. Seifarth, “Identitäten und Identifikatoren in ökologischen Daten mit Topic Maps,” Institut für Informatik, 2012.

[23] Dan Smith and B. Szekely, “LSID best practices,” 2005. [Online]. Available: http://www.ibm.com/developerworks/opensource/library/os-lsidbp/. [Accessed: 14-Feb-2012].

[24] GBIF, “A Beginner’s Guide to Persistent Identifiers, Version 1.0,” Global Biodiversity Information Facility (GBIF), Copenhagen, 2011.

[25] R. L. Rivest, The md5 message-digest algorithm. Network Working Group, 1992.

[26] F. Naumann and U. Leser, Informationsintegration. dpunkt, 2007.

[27] P. Bouquet, D. Giacomuzzi, and H. Stoermer, “OKKAM : Enabling a Web of Entities,” in Entity-Centric Approaches to Information and Knowledge Management on the Web, 2007.

[28] R. T. Fielding and R. N. Taylor, “Principled design of the modern Web architecture,” ACM Transactions on Internet Technology, vol. 2, no. 2, pp. 115-150, May 2002.

[29] J. H. Porter and M. Kortz, “Web services in the U . S . Long-Term Ecological Research Network : Now and in the future,” in Environmental Information Management Conference 2011, 2011, pp. 111-116.
[30] J. Porter et al., “A controlled vocabulary for LTER data keywords,” in Environmental Information Management Conference 2011, 2011, pp. 168-169.
[31] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers, “GenBank,” Nucleic acids research, vol. 39, no. Database issue, pp. D32-7, Jan. 2011.
[32] J. Kattge et al., “TRY - a global database of plant traits,” Global Change Biology, vol. 17, no. 9, pp. 2905–2935, Apr. 2011.
[33] W. K. Michener and M. B. Jones, “Ecoinformatics: supporting ecology as a data-intensive science,” Trends in Ecology & Evolution, Jan. 2012.
[34] C. Gries and J. H. Porter, “Moving from custom scripts with extensive instructions to a workflow system : use of the Kepler workflow engine in environmental information management,” in Environmental Information Management Conference 2011, 2011, pp. 70-75.