



# History and development of ABCDEFG: a data standard for geosciences

Mareike Petersen<sup>1</sup>, Falko Glöckler<sup>1</sup>, Wolfgang Kiessling<sup>2</sup>, Markus Döring<sup>3</sup>, David Fichtmüller<sup>4</sup>, Lertsutham Laphakorn<sup>1</sup>, Brian Baltruschat<sup>1</sup>, and Jana Hoffmann<sup>1</sup>

<sup>1</sup>Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Invalidenstr. 43, 10115 Berlin, Germany

<sup>2</sup>GeoZentrum Nordbayern, Friedrich-Alexander Universität Erlangen-Nürnberg, Loewenichstr. 28, 91054 Erlangen, Germany

<sup>3</sup>GBIF Secretariat, Universitetsparken 15, 2100 Copenhagen, Denmark

<sup>4</sup>Botanischer Garten und Botanisches Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Straße 6–8, 14195 Berlin, Germany

**Correspondence:** Mareike Petersen (mareike.petersen@mfn-berlin.de)

Received: 24 May 2017 – Revised: 25 July 2017 – Accepted: 18 August 2017 – Published: 17 January 2018

**Abstract.** Museums and their collections have specially customized databases in order to optimally gather and record their contents and associated metadata associated with their specimens. To share, exchange, and publish data, an appropriate data standard is essential. ABCD (Access to Biological Collection Data) is a standard for biological collection units, including living and preserved specimen, together with field observation data. Its extension, EFG (Extension for Geoscience), enables sharing and publishing data related to paleontological, mineralogical, and petrological objects. The standard is very granular and allows detailed descriptions, including information about the collection event itself, the holding institution, stratigraphy, chemical analysis, and host rock. The standard extension was developed in 2006 and has been used since then by different initiatives and applied for the publication of collection-related data in domain-specific and interdisciplinary portals.

## 1 Introduction

Natural history museums and universities harbor millions of collection items gathered in expeditions over hundreds of years. Most collections are highly diverse and comprise voucher specimens from different scientific disciplines such as botany, zoology, paleontology, geology, and anthropology. Independently of discipline and type of preservation, each collection item is accompanied by various metadata regarding not only its collecting event (time, region, and collector)

but also research data, aggregated over time (laboratory analysis, relations to other collection items, etc.). In addition, images or other multimedia files of the collection item or the gathering locality are connected to stored objects. Furthermore, collections which are born digital collections, e.g., animal sound archives, are likewise commonly associated with natural history museums. This variety of information makes each collection unique and leads to specific customizations in the databases for each museum or even for a single collection.

However, once mobilized, the data should not only subsist in institutional databases but also be shared with the scientific community and laymen alike using the internet. Leading scientific organizations have signed agreements to facilitate the exchange of scientific results in the sense of open access (e.g., Berlin Declaration on Open Access to knowledge in the sciences and humanities, 2003; Bouchout Declaration, 2014). Furthermore, large funding bodies have implemented principles for the sharing and dissemination of research data (e.g., Alliance of German Science Organizations, 2010; National Science Foundation, 2014) and various journals require accessibility of data after publication.

In order to effectively share data relating to collection units from structurally diverse databases, they need to be mapped to a data scheme which is universally intelligible. Data standards are designed to facilitate the exchange and publication of information via a range of different approaches and pattern. Their use provides semantics and structure in order to

avoid errors in interpreting the data, and enhances automated treatment, e.g., usage in web portals and services.

ABCD (*Access to Biological Collection Data*) is a data standard for biological collection units, including living and preserved specimens, and is also applicable to field observation data (Berendsohn, 2007). The standard allows a detailed and atomized mapping of collection specimen and field observation data in the biological science disciplines. However, fossil specimens are usually accompanied with stratigraphic and geological data, which are inadequately covered by ABCD. Since paleontological, mineralogical, and petrological items are also covered by natural history collections, the *Extension for Geoscience* (EFG; Kiessling et al., 2006) has been developed in order to complement ABCD towards a more complete standard for natural history (ABCDEFG).

Here we review the history and usage of ABCDEFG and highlight current developments, its uniqueness, and importance as a data standard in the domain of earth science.

## 2 Historical background of the EFG schema

The development of ABCD started in 2000. Five years later, ABCD version 2.06 was ratified as a standard by the non-profit scientific and educational association *Biodiversity Information Standards* (TDWG), formerly known as the Taxonomic Databases Working Group ([www.tdwg.org](http://www.tdwg.org)). Development of the Extension for Geoscience (EFG) started in 2005. In the framework of the SYNTHESYS project ([www.synthesys.info](http://www.synthesys.info), funded by the European Commission) the *Geosciences Collection Access Service* (GeoCASE, [www.geocase.eu/project](http://www.geocase.eu/project)), led by Museum für Naturkunde Berlin, was developed to make standardized paleontological, mineralogical, and petrological collection data openly available through the internet. GeoCASE built upon the technology of the *Biological Collection Access Service* (BioCASE; developed by the BioCASE and BioCISE projects funded by the European Commission from 1996 to 2004; Berendsohn, 2000; Güntsch et al., 2007). As a first step towards the definition of a schema for the earth sciences, a team of 11 experts from several European institutions specified the requirements on typical data that describe paleontological and geological collection objects. Building on ABCD, the resulting schema extension was named EFG.

## 3 Composition of EFG

### 3.1 Extent of EFG

The ABCDEFG schema is very granular and allows a detailed description of data gathered in geoscience (see technical schema documentation at <http://www.geocase.eu/efg>). It comprises a broad range of properties, including

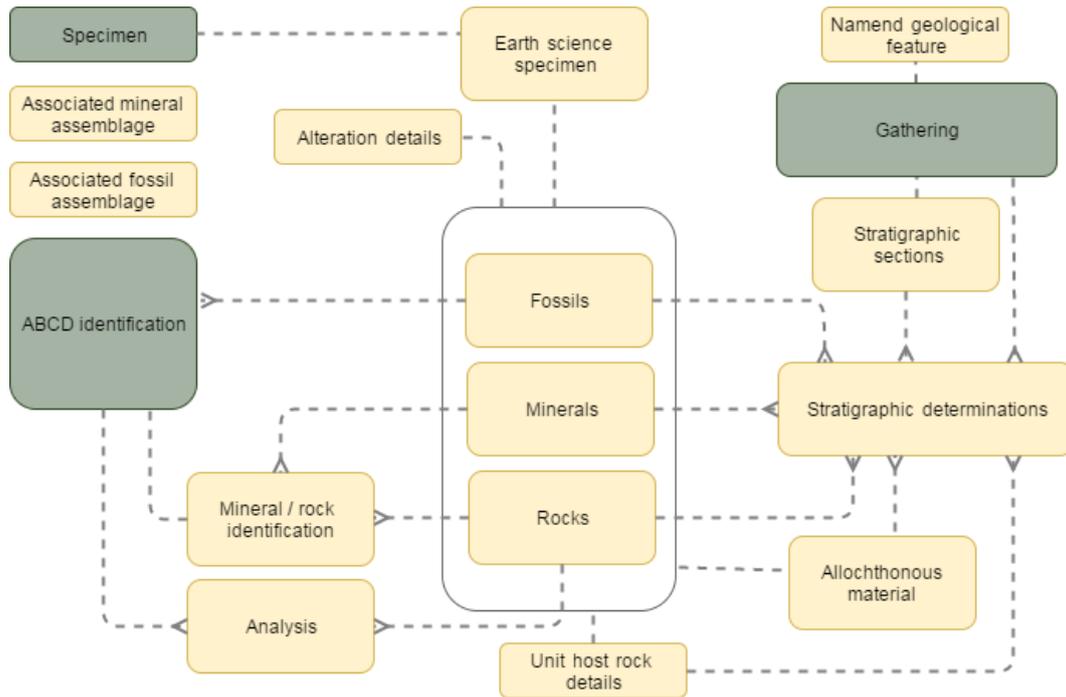
- geological and geomorphological observations;
- geological specimens and their preparation;
- paleontological, mineral, rock and sediment specimens; anomalous items (e.g., glacial erratics, transported assemblages);
- stratigraphic and absolute dates, measured stratigraphic sections, and borehole logs;
- identifications (extended to cover rock and mineral classifications, varietal names etc.) and analyses (techniques and results, e.g., chemical composition of minerals, petrological analysis of rock);
- integrated description of the host rock as part of a unit record (e.g., mineralization and lithological context of a fossil).

As fossil objects share a considerable number of properties from both biology and geology, it is a great advantage to establish a geoscience schema as an extension of a biological data standard, rather than creating a completely separated standard for geosciences. Furthermore, another area of overlap are the data elements that refer to the collection or collection event itself. The EFG extension is not applicable without its backbone, the ABCD schema itself (Fig. 1). Elements describing the collection, the holding institution, contact persons, and any additional information on the dataset are placed within ABCD. Furthermore, details on the gathering event (location, date, collector, etc.) and links to object related multimedia material including their respective licensing or any related publication can be specified in the ABCD standard.

### 3.2 Structure and technical requirements

ABCDEFG is an XML-based schema and thus can be easily processed by software and is readable by humans. The schema is hierarchically structured and the descriptive elements are aggregated in thematic concepts. All elements are arranged according to their contextual relation to other elements and their cardinality (Holetschek et al., 2012). The full potential of ABCDEFG is unlocked in combination with the free and open-source BioCASE Provider Software (BPS; [http://www.biocase.org/products/provider\\_software/](http://www.biocase.org/products/provider_software/)). It consists of a generic XML wrapper software for relational databases and a custom XML-based protocol for communication, and it requires a schema definition (XSD) for each supported data standard.

The BPS XML wrapper is able to connect to a wide array of database management systems (e.g., MySQL, PostgreSQL, Microsoft SQL, Oracle, as well as Excel spreadsheets) and provides a user interface for mapping the database fields to the data standards' elements. This generic



**Figure 1.** Simplified overview of the schema extension EFG. Given are the relationships among EFG components and between EFG components and the ABCD schema. EFG elements are given in yellow and elements originating from ABCD in green.

approach of mapping the individual data models to the standard schema is the essential functionality for standardized data mobilization. Once an institution has set up the mapping for its database, the released content (including revisions, updates and new records) is exposed in the same manner through the BPS. Thus, only changes in the data model itself necessitate modifications in the mapping.

Although mapping of data is straightforward, the (potentially heterogeneous) content needs to be checked on a regular basis to verify that the results are appropriate within the data standard's context. This practice is supported by the BioCASE technology, as different datasets and sources provided in ABCD and ABCDEFG can easily be monitored using the BioCASE Monitor Service (Glöckler et al., 2013). By using the BioCASE data interface and protocol in the backend and a comprehensible user interface as a frontend, the monitoring tool facilitates checks of the structure, plausibility, and completeness of datasets. Furthermore, it verifies compliance of provided data for transformation into other target schemas and can be used for summaries and simple descriptive statistics across different datasets. Thus, data aggregators like the Geosciences Collection Access Service (GeoCASE) can make use of the tool in the supervision of the progress in data provision. A full implementation of the BioCASE Monitor Service for summarizing the data provision and checking the minimum requirements of data mapped to ABCDEFG can be found on the GeoCASE website ([http://geocase.eu/partners\\_and\\_providers](http://geocase.eu/partners_and_providers)).

## 4 Importance of ABCDEFG

### 4.1 Comparable standards

For several geoscientific disciplines, data recommendations exist and structures provide guidance on what kind of data should be recorded in what form during field work or which research information should be publicly available. The Evolution of Terrestrial Ecosystems Program (ETE; Damuth, 1997) was designed in the late 1990s and provided the basis for later developments of various data structures (Reed et al., 2015). At about the same time the Paleobiology Database (PBDB; <https://paleobiodb.org>) was founded. The PBDB is a public resource of collection-based occurrence and taxonomic data for fossils of all geological ages. The PBDB considerably inspired the development of the EFG schema. Other important data conventions for paleontology are the FAUNMAP and MIOMAP initiatives, focusing on fossil occurrences in North America from different time periods (Carrasco et al., 2005; Graham and Lundelius Jr., 2010).

All these data structures (ETE, PBDB, and FAUNMAP/MIOMAP) perfectly manage localities and faunal lists, but cannot be applied on specimen data in a straightforward fashion. Gilbert and Carlson (2011) established a data dictionary for paleoanthropology, including a specific table schema for specimens as the central unit of organization, and localities as clusters of specimens. This structure facilitates the data collection in the field, the curation of objects in institutional collections, and the publication of specimen related

information. Along with the International Geo Sample Number (IGSN) for the registering of samples to the System for Earth Sample Registration SESAR two metadata sets were defined. These metadata schemes comprehend core metadata used for the registration itself (registration metadata) as well as basic characteristics of physical samples and collections (descriptive metadata and latest version of both sets are available via <https://github.com/IGSN/metadata/wiki>). EarthChem (<http://www.earthchem.org/>), the NSF (National Science Foundation)-funded data facility for solid earth geoscience data, provides suggested and controlled vocabularies for geochemical, geochronological, and petrological data in their EarthChem Systems (<http://www.earthchem.org/resources/vocabularies>) in order to help standardizing these types of data in general. In addition, recommended properties for a detailed description of minerals and their structure can be derived from Mindat (<http://mindat.org>), the most comprehensive and world's largest public database for mineral information. However, none of the above-mentioned recommendations has yet reached the status of a ratified standard.

In biodiversity science and related disciplines, Darwin Core (DwC) and ABCD are the best-known and most commonly used standards. In parallel with ABCD, Darwin Core evolved over time and its development is mainly driven by the community (Wieczorek et al., 2012). Although both schemas aim at facilitating the mobilization of biodiversity data, they differ in their purpose. ABCD is a highly structured schema and was intended to be applicable to a great variety of data in various degrees of granularity, including their relations and cardinality. It is almost exclusively tied to the XML format. Darwin Core in contrast aims at providing a flexible vocabulary for sharing information about biological diversity and can be represented using various technologies including XML, RDF, or plain CSV files. It is used to describe the recorded occurrence of a species in space and time, and to link related information and evidence to this declaration. The Darwin Core standard was ratified in 2009 by TDWG and since then it is the most common standard for publishing datasets in the Global Biodiversity Information Facility (GBIF; <http://gbif.org>).

The usage of Darwin Core for geoscience is limited. Although specimens can be classified as “FossilSpecimen” (<http://rs.tdwg.org/dwc/terms/FossilSpecimen>) and basic information about the chrono-, litho-, and biostratigraphy can be accommodated in the *GeologicalContext* terms (<http://rs.tdwg.org/dwc/terms/GeologicalContext>), appropriate terms for geological collection items are still lacking in Darwin Core. Thus, important measurements and facts like the host rock of a specimen, the original stratigraphic association of allochthonous material and metasomatic, metamorphic, or diagenetic alterations of the rock cannot be shared and distributed using this data standard. The Darwin Core Paleontology Extension (DarwinCoPE) was proposed as a community standard in 2005, but even then it was described as

less detailed than the evolving ABCDEFG (Theodor, 2006). Furthermore, the PaleoCore initiative established a list of common terms for the publication of paleoanthropological datasets (<http://paleocore.org/standard/>) and combined elements from Darwin Core with some of the Dublin Core Metadata Initiative (DCMI, <http://dublincore.org/documents/dcmi-terms/>). However, the possibilities of accommodating comprehensive geoscientific data are still limited.

In summary ABCDEFG is the most comprehensive data standard for geoscientific collection data. Although there have been different approaches, none of them provide a comparable detailed and atomized schema for both paleontological and geological data.

#### 4.2 Use cases

ABCDEFG is already used for data aggregation and provision to a number of thematic data portals. The following examples illustrate different data portals and aggregators showing collection related information from the same data source endpoints mapped to the ABCDEFG standard.

The Global Biodiversity Information Facility (GBIF, <http://www.gbif.org>) is a web portal for the access to biodiversity data. It accepts several different data standards, including Darwin Core, ABCD, and ABCDEFG. As the paleobiological data provide important information about ancient biodiversity, fossils from natural history collections and in the published literature are important data sources for GBIF. Therefore, different institutions provide occurrence data of not only extant species but also of fossil specimens.

Examples:

Fossil (*Mammuthus primigenius*), <http://www.gbif.org/occurrence/1099276603>

Fossil (*Dorygnathus banthensis*), <http://www.gbif.org/occurrence/1099085192>

Within the SYNTHESYS project task *Geosciences Collection Access Service* (GeoCAsE) a data portal ([www.geocase.eu/access](http://www.geocase.eu/access)) was developed to make data from the GeoCAsE network freely available on the internet. The portal uses the BioCAsE technology to perform a distributed query on the original provider's databases. In contrast to paleontological specimens, data on geoscientific collection objects are not available via GBIF. GeoCAsE thus represents a complementary portal for non-biological data, while both portals overlap in data provision on fossil objects. GeoCAsE originally had its scope on geoscientific objects (fossils, minerals, and rocks) from European institutions. However, it is now open for worldwide earth science collections

Examples:

Fossil (*Mammuthus primigenius*), <http://geocase.eu/portal/?unitid=MB.Ma.13543>

Fossil (*Dorygnathus banthensis*), <http://geocase.eu/portal/?unitid=MB.R.3661>

Mineral (azurite), [http://geocase.eu/portal/?unitid=MFN\\_MIN\\_1999\\_2270](http://geocase.eu/portal/?unitid=MFN_MIN_1999_2270)

The German Federation for Biological Data (GFBio; <http://www.gfbio.org>) takes care of the management and standardization of biological research data during the entire data life cycle (DataOne, 2017) and uses ABCDEFG for fossil data in order to standardize heterogeneous datasets and to publish the data on the national GFBio portal. The GFBio data centers are responsible for the sustainable management, long-term accessibility, and archiving of their client's data. Therefore, they build upon domain-specific data standards like ABCDEFG that are actively used, well documented, and accepted by the community.

Example:

Fossil (*Jurassiphorus cailliaudanum*), <https://www.gfbio.org/data/search?q=MB.Ga.524>

Species (*Mammuthus primigenius*), <https://www.gfbio.org/data/search?q=Mammuthusprimigenius>

*European digital library Europeana*. The European digital library Europeana (<http://www.europeana.eu>) displays the cultural heritage of Europe. The EU-funded project OpenUp! and its successor project Europeana DSI provide more than 3 million images and other multimedia objects from the natural history domain (Berendsohn and Güntsch, 2012; <http://open-up.eu>). Project partners are delivering their data in the ABCDEFG schema using the BioCASE technology and making their multimedia files (mainly images) freely available, including those of fossils, minerals, and rocks. As Europeana focuses on media objects, specific filter requests exclude all those records without any multimedia files. But again, none of the data providers need to set up a separate schema mapping, parallel to those of the above-mentioned data portals. Furthermore, Europeana compiles data from cultural heritage collections that use their own data standards. In order to harmonize these with the data from the natural history domain, a subset of elements from the ABCDEFG schema was identified to meet the requirements of the Europeana data model (ESE/EDM; Zágorsšek et al., 2012). As ABCDEFG is XML-based, the transformation to the Europeana data model is straightforward.

Examples:

Fossil (*Dorygnathus banthensis*), [http://www.europeana.eu/portal/de/record/11622/\\_MFN\\_FOSSIL\\_VERTEBRATES\\_V\\_MFN\\_GERMANY\\_MB\\_R\\_3661.html](http://www.europeana.eu/portal/de/record/11622/_MFN_FOSSIL_VERTEBRATES_V_MFN_GERMANY_MB_R_3661.html)

Mineral (azurite), [http://www.europeana.eu/portal/de/record/11622/\\_MFN\\_MINERALOGY\\_MFN\\_GERMANY\\_MFN\\_MIN\\_1999\\_2270.html](http://www.europeana.eu/portal/de/record/11622/_MFN_MINERALOGY_MFN_GERMANY_MFN_MIN_1999_2270.html)

*Object landing pages of the Museum für Naturkunde Berlin*. Recently, the Consortium of European Taxonomic Facilities (CETAF; <http://cetaf.org/>) adopted the approach of using actionable, web-based uniform resource identifiers ([http-URIs](http://www.ietf.org/rfc/rfc3986.txt)) for physical objects (Güntsch et al., 2017). The advantage of this approach is the immediate linkage of the physical object with its digital representation. As the Hypertext Transfer Protocol ([http](http://www.ietf.org/rfc/rfc2616.txt)) is a well-known syntax, people tend to “click” on the URI and expect to find information about the physical object. Thus, the [http-URIs](http://www.ietf.org/rfc/rfc3986.txt) ideally resolve to web landing pages with a compilation of the objects' media and collection data. At the Museum für Naturkunde Berlin these landing pages make use of ABCDEFG structured data to use the standard's controlled and well-documented terms for exposing the information.

Examples:

Fossil (*Dorygnathus banthensis*), [http://coll.mfn-berlin.de/u/MB\\_R\\_3661](http://coll.mfn-berlin.de/u/MB_R_3661)

Mineral (azurite), [http://coll.mfn-berlin.de/u/MFN\\_MIN\\_1999\\_2270](http://coll.mfn-berlin.de/u/MFN_MIN_1999_2270)

Table 1 summarizes the number of institutions using the ABCDEFG schema and the number of objects published in different portals.

## 5 Conclusion and outlook

ABCDEFG has been established in the national and international community and its reach is constantly growing. Once the standard and the supporting software are implemented at an institution and its database systems are linked up, there are many application opportunities. Apart from facilitating highly distributed data exchange and publication, this also facilitates institutional processes such as designing a collection database by using the ABCDEFG terms and data management with documentation and archiving based on ABCDEFG structures and definitions.

The ABCD schema and its extension EFG are currently being revised. Within the project “ABCD 3.0 – A community platform for the development and documentation of the ABCD standard for natural history collections” (funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation), <http://abcd.biowikifarm.net/>) the entire schema is reviewed by the scientific community. The EFG extension is being expanded to cover specific demands of earth scientists and extraterrestrial collection items such as meteorites. The ABCDEFG schema was imported into the TDWG Terms Wiki ([https://terms.tdwg.org/wiki/ABCD\\_EFG](https://terms.tdwg.org/wiki/ABCD_EFG)), a development platform for data standards related to biodiversity. The complete documentation of all individual ABCDEFG terms is now available on a site, the TDWG Terms Wiki, which allows collaborative improvement (curation, annotation, discussion) of the terminology and schema.

**Table 1.** Use cases of the ABCDEFG. Given are the different portals, number of institutions using ABCDEFG, and the number of fossils, rocks, and minerals provided in ABCDEFG to different internet platforms. The same datasets might be published in several portals. Please note that most fossil data provided in ABCDEFG and published on GBIF are not directly originating from collection items but are extracted from the literature via the PBDB.

Portal	Number institutions		Number objects	
	Fossils	Rocks and minerals	Fossils	Rocks and minerals
GeoCASE	9	5	907 965	255 862
GBIF	9	n/a	762 059	n/a
GFBio	1	n/a	24 857	n/a
Europeana	7	3	28 398	5701

Furthermore, specifications of relationships among various standards (e.g., “has broader match”, “is part of”) or direct translations into terms of other vocabularies are possible. This facilitates the usage of the schema and enables further development by the scientific community.

In a further step, the XML-based structure is being changed into a more semantic form (Resource Description Framework, RDF), including terms of external ontologies. To facilitate the usage of ABCDEFG, specific application schemas will be developed. These schemas will include mandatory concepts (see Holetschek, 2016) as well as commonly used elements (Holetschek, 2015) and further relevant concepts identified by the community. Thereby the application schemas represent all information that scientists and curators would like to share with colleagues and the broader public. An application schema for minerals, meteorites, and drilling cores commonly found in geological collections will be provided.

The ABCDEFG schema has been submitted to TDWG for its ratification, the decision of which is pending. TDWG’s Paleobiology Interest Group (<https://github.com/tdwg/paleo>) will continue the effort of extending Darwin Core with paleontological elements where needed. This will be conducted in compliance with ABCDEFG.

*Data availability.* ABCDEFG XML Schema Definition (XSD) is available at <https://doi.org/10.7479/pwnr-sh74> (Kiessling et al., 2018).

*Competing interests.* The authors declare that they have no conflict of interest. The authors and the handling editor are affiliated to the same institution, but this fact did not give any preference to the choice of the journal. The journal was chosen in order to raise awareness of the data standard in the geoscientific and paleontological community.

*Acknowledgements.* The authors thank various project collaborators for their input during the development, improvement and usage of the ABCDEFG schema. The following projects and ini-

tiatives and the people involved therein sustainably influenced the current status of the schema: GeoCASE (<http://www.geocase.eu/project>), BioCASE (<http://www.biocase.org/index.shtml>), GBIF (<http://www.gbif.org>), and OpenUp! (<http://open-up.eu/en>). We are thankful to the late Charles Copp for his work and his considerable contribution to the development of EFG.

Funding was provided via GeoCASE (European Commission, SYNTHESYS I, Activity D) for initial developmental phase and funding of workshops, GBIF-D (Federal Ministry of Education and Research, Bundesministerium für Bildung und Forschung, BMBF) for provider support and maintenance, and ABCD 3.0 (DFG) for current developments and revisions. All this support is gratefully acknowledged.

Edited by: Johannes Müller

Reviewed by: Donald Hobern and Walter Berendsohn

## References

- Alliance of German Science Organizations: <http://www.allianzinitiative.de/en/core-activities/research-data/principles.html> (last access: 12 May 2017), 2010.
- Berendsohn, W. G. (Ed.): Resource Identification for a Biological Collection Information Service in Europe (BioCISE), Botanic Garden and Botanical Museum Berlin-Dahlem, available at: [http://www.biocase.org/doc/Berendsohn\\_W\\_G\\_al\\_2000\\_-\\_Concepts\\_for\\_a\\_European\\_portal\\_to\\_biological\\_collections.pdf](http://www.biocase.org/doc/Berendsohn_W_G_al_2000_-_Concepts_for_a_European_portal_to_biological_collections.pdf) (last access: 13 July 2017), 2000.
- Berendsohn, W. G. (Ed.): Access to biological collection data. ABCD Schema 2.06 – Ratified TDWG Standard, available at: <http://www.bgbm.org/TDWG/CODATA/Schema/default.htm> (last access: 9 January 2017), 2007.
- Berendsohn, W. G. and Güntsch, A.: OpenUp! Creating a cross-domain pipeline for natural history data, in: No specimen left behind: mass digitization of natural history collections, edited by: Blagoderov, V. and Smith, V. S., ZooKeys, 209, 47–54, <https://doi.org/10.3897/zookeys.209.3179>, 2012.
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, available at: <https://openaccess.mpg.de/Berliner-Erklaerung> (last access: 12 May 2017), 2003.
- Bouchout Declaration: <http://www.bouchoutdeclaration.org/declaration> (last access: 12 May 2017), 2014.
- Carrasco, M. A., Kraatz, B. P., Davis, E. B., and Barnosky, A. D.: Miocene Mammal Mapping Project (MIOMAP), Uni-

- versity of California Museum of Paleontology, available at: <http://www.ucmp.berkeley.edu/miomap/> (last access: September 2017), 2005.
- Damuth, J. D.: ETE database manual, Evolution of Terrestrial Ecosystems Consortium, Washington D.C.: Dept. of Paleobiology, Smithsonian Institution, available at: [http://naturalhistory.si.edu/ete/ETE\\_Database\\_Manual\\_Introduction.html](http://naturalhistory.si.edu/ete/ETE_Database_Manual_Introduction.html) (last access: September 2017), 1997.
- DataOne: Data life cycle, available at: <https://www.dataone.org/data-life-cycle>, last access: 19 May 2017.
- Dublin Core Metadata Initiative: DCMI Metadata Terms, available at: <http://dublincore.org/documents/dcmi-terms/>, last access: 21 September 2017.
- Gilbert, H. and Carlson, J.: Data models and global data integration in paleoanthropology: a plea for specimen-based data collection and management, in: Pleistocene databases: acquisition, storing, sharing, edited by: Macchiarelli R. and Weniger, G. C., Mettmann: Neanderthal Museum, 111–121, 2011.
- Glöckler, F., Hoffmann, J., and Theeten, F.: The BioCASE Monitor Service – A tool for monitoring progress and quality of data provision through distributed data networks, *Biodiversity Data Journal*, 1, e968, <https://doi.org/10.3897/BDJ.1.e968>, 2013.
- Graham, R. W. and Lundelius Jr., E. L.: FAUNMAP II: New data for North America with a temporal extension for the Blancan, Irvingtonian and early Rancholabrean, FAUNMAP II Database, available at: <http://www.ucmp.berkeley.edu/faunmap/index.html> (last access: September 2017), 2010.
- Güntsch, A., Mergen, P., and Berendsohn, W. G.: The BioCASE Project – a Biological Collections Access Service for Europe, *Ferrantia*, 51, 103–108, 2007.
- Güntsch, A., Hyam, R., Hagedorn, G., Chagnoux, S., Röpert, D., Casino, A., Droegge, G., Glöckler, F., Gödderz, K., Groom, Q., Hoffmann, J., Holleman, A., Kempa, M., Koivula, H., Marhold, K., Nicolson, N., Smith, V. S., and Triebel, D.: Actionable, long-term stable and semantic web compatible identifiers for access to biological collection objects, *Database*, 1–9, <https://doi.org/10.1093/database/bax003>, 2017.
- Holetschek, J., Dröge, G., Güntsch, A., and Berendsohn, W. G.: The ABCD of primary biodiversity data access, *Plant Biosystems*, 146, 771–779, 2012.
- Holetschek, J. (Ed.): BioCASE Concept survey, Biological Collection Access Service for Europe, available at: [http://www.biocase.org/whats\\_biocase/concept\\_survey.cgi](http://www.biocase.org/whats_biocase/concept_survey.cgi) (last access: 9 January 2017), 2015.
- Holetschek, J. (Ed.): Commonly used ABCD 2.06 concepts, Documentation Wiki of the BioCASE Provider Software, available at: <http://wiki.bgbm.org/bps/index.php/CommonABCD2Concepts> (last access: 9 January 2017), 2016.
- Kiessling, W., Copp, C., Rissoné, A., Döring, M., and Mewis, H.: The EFG extension to the ABCD schema, edited by: Belbin, L., Rissoné, A., and Weitzman, A., Proceedings of TDWG: Abstracts of the 2006 Annual Conference of Biodiversity Information Standards (TDWG), St. Louis, Missouri, USA, 15–22 October 2006, 6–7, 2006.
- Kiessling, W., Copp, C., Rissoné, A., Döring, M., and Mewis, H.: The Extension for Geosciences (EFG) of the Access to Biological Collection Data (ABCD) schema, available at: <https://doi.org/10.7479/pwnr-sh74>, last access: 17 January 2018.
- National Science Foundation: <https://www.nsf.gov/bfa/dias/policy/dmp.jsp> (last access 12 May 2017), 2014.
- Reed, D., Barr, W. A., McPherron, S. P., Bobe, R., Geraads, D., Wynn, J. G., and Alemseged, Z.: Digital data collection in paleoanthropology. *Evolutionary Anthropology: Issues, News, and Reviews*, 24, 238–249, 2015.
- Theodor, J.: DarwinCoPE, a Proposed Paleontological Extension to DarwinCore 2, edited by: Belbin, L., Rissoné, A., and Weitzman, A., Proceedings of TDWG: Abstracts of the 2006 Annual Conference of Biodiversity Information Standards (TDWG), St. Louis, Missouri, USA, 15–22 October 2006, 49–50, 2006.
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., and Vieglaiss, D.: Darwin Core: an evolving community-developed biodiversity data standard, *PlosOne*, 7, e29715, <https://doi.org/10.1371/journal.pone.0029715>, 2012.
- Zágoršek, K., Sejkora, J., and Hoffmann, J.: C4.5.0 - Model for the integration of content from the areas of palaeontology and mineralogy, OpenUp! Project Deliverable, available at: [http://open-up.eu/sites/open-up.eu/files/C450\\_Model\\_for\\_the\\_integration\\_of\\_content\\_palaeontology\\_mineralogy\\_NM.pdf](http://open-up.eu/sites/open-up.eu/files/C450_Model_for_the_integration_of_content_palaeontology_mineralogy_NM.pdf), 2012.