

Meeting Report – „Implementation of the ABCDEFG-schema for specimen-based earth science collection data”

Berlin, 1st – 3rd February 2007

Venue: Museum of Natural History
Humboldt University, Berlin
Invalidenstrasse 43
10115 Berlin

Attendants: Adrian Rissoné (Natural History Museum, London), Charles Copp (Bristol), David Lazarus (Museum of Natural History, Berlin), Evgeniy Meyke (University of Helsinki), Fedor Steeman (Natural History Museum, Copenhagen), Heike Mewis (Museum of Natural History, Berlin), Jörg Holetschek (Botanical Garden & Botanical Museum, Berlin), Ken Johnson (Natural History Museum, London), Lutz Hecht (Museum of Natural History, Berlin), Markus Doering (Botanical Garden & Botanical Museum, Berlin), Mietta Petronio (ETH Zurich), Milena Pika-Biolzi (ETH Zurich), Ralf-Thomas Schmitt (Museum of Natural History, Berlin), Robert Huber (University Bremen)

Abstract

The EFG-schema is supposed to become a standard protocol for the query of geoscientific databases, though several bureaucratic obstacles still have to be overcome. A charter to establish an EFG-task group within TDWG has to be written and the EFG schema has to be approved by the Executive of TDWG. The schema and the developed prototype of GeoCASE, including the databases of the Museum of Natural History in Berlin, will have to be presented at several meetings, and potential cooperations with other projects should be considered. Finally it is necessary to include EFG and the possibility to query geoscientific databases into GBIF.

Introduction

The aim of this meeting was to discuss the scope and potential of GeoCASE and to report the progress on the implementation of ABCDEFG in the MfN. Because of an accident, Wolfgang Kiessling was unable to participate, but nevertheless, the meeting was managed with a successful outcome.

Meeting Report

On Thursday, 1st February, the meeting started with a presentation of the Specify Biodiversity Collections management software by Rod Spears, which was not only a presentation for the participants of the workshop, but for all staff of the Museum of Natural History Museum (MfN). The MfN considers to use the Specify as standard software for its collection databases. Specify is an open source software that is not only useful as a collection database but it also simplifies the management of the collections.

The actual meeting on „Implementation of the ABCDEFG-schema for specimen based earth science collection data” started with a greeting by Lutz Hecht. Adrian Rissoné gave a short introduction into TDWG and his work for TDWG. He informed us that TDWG (formerly Taxonomic Database Working Group) got a new name and is now called “Biodiversity Information Standards”.

Charles Copp gave an overview of the ABCD and its extension (EFG), which is a fairly comprehensive schema. His presentation was followed by a short discussion of what information is missing in the extended schema (tectonics, sources of information about stratigraphic age, etc., including data from the Ocean Drilling Program), but everyone agreed that it is necessary to submit this version of the schema first before thinking about a further version with anymore information.

The next day started with presentations by Kenneth Johnson and Evgeniy Meyke who introduced the other participants into their work and experiences with collection databases. Evgeniy Meyke is currently developing a biodiversity research database software called “Taxis 4” and he wants to orientate on the extended ABCD schema for simplifying later implementation of the databases with the protocol. Heike Mewis followed with her report about mapping the EFG against the paleontological and mineralogical databases of the MfN. This prototype was called GeoCASE. Most of the information included in the EFG schema could not be used, because the information are not contained in the databases. The main problems that occurred were due to the structure of the databases itself, minor problems occurred due to the schema. Charles Copp agreed to change these problems within the schema. The audience agreed upon recommending standard database software such as Specify or Taxis 4 to institutions which are interested in mobilizing their data. This may simplify the procedure of mapping databases against the ABCDEFG schema. Both, Rod

Spears as well as Evgeniy Meyke are interested in following the structure of the EFG while programming their software.

When considering how to get EFG adopted by others and how to distribute it, the main idea was to include GeoCASE into GBIF. The problem is, that GBIF allows the search for species names only, which is problematic especially for querying mineralogical databases. There are no financial resources to build up a second portal of the size of GBIF. Markus Doering presented the structure of the GBIF and BioCASE networks, which illustrated that it is currently impossible to use either the GBIF or the BioCASE portal for geoscientific data. Everybody agreed that GBIF should be convinced to fully support the EFG by using the argument that GBIF should be a global provider for specimen based collection data.

Jörg Holetschek demonstrated a simple user interface (SimpleUI) developed within the BioCASE project. This could be adapted for GeoCASE to bridge the time until GeoCASE can use the GBIF portal. It is important to demonstrate the strength of EFG and GeoCASE via such a test portal. The MfN agreed to host this portal. The SimpleUI is a simple search portal that can be used to query data of ABCD providers. It does not have a local cache, but just queries the databases of all participating institutes. Charles Copp reported that Luxembourg is planning to be the next institute to join GeoCASE, and Milena Pika-Biolzi and Mietta Petronio notified that they are interested to mobilize the databases of the ETH in Zurich. Other European Specify users should also be asked to adopt the protocol.

In a final discussion we summarized the results and listed the tasks that have to follow this meeting.

Results

We successfully discussed where we are, where GeoCASE is, where the schema is, and where we want to go. There are no major issues with the schema, it needs some modifications only. A prototype of GeoCASE, including the paleontological and mineralogical databases of the MfN, is online, but needs to expand and to become more public.

It is important to move EFG/GeoCASE forward into TDWG. For this reason, a charter should be written, to establish a EFG task group within TDWG. This charter should be written by Wolfgang Kiessling with the help of Adrian Rissoné. Next to the conveyor, the task group needs core and non-core members. Core members play a significant role in progressing the schema. Core members of the EFG task group will be: Charles Copp, Adrian Rissoné, Wolfgang Kiessling and probably Ralf-Thomas Schmitt and/or Lutz Hecht. All other attendants will be non-core members.

The submission of the schema to TDWG will be performed by Charles Copp.

Heike Mewis was asked to write a summary of the current status of GeoCASE, including information about the schema, the test application, the scope and the state of encouraging others to participate.

It should be thought about a secure funding for the future of GeoCASE. If the SimpleUI will be hosted in the MfN, it is necessary to extend the project and write a proposal for Synthesys e.g. A roadmap for future financial support should be developed.

Adrian Rissoné will prepare a Wiki at the TDWG website.

Other organizations like Chronos, PBDB, IMA-Commission on Museums etc. should be informed about GeoCASE, to allow a future cooperation. During the next year, GeoCASE and the EFG should be presented at as many conferences as possible. Everybody who takes part on a geoscientific meeting should give a talk about this project.

In September 2007, the schema will be reviewed by experts. Therefore, names of possible reviewers that TDWG can use, can be suggested. David Lazarus suggested Pat Diver, who is a Chronos/oil industry data modeller/programmer.

Summary and Outlook

A simple portal will be set up in the MfN. Luxembourg and Zurich are the next who will join the GeoCASE network. A EFG task group will be established within TDWG and the schema will be submitted to TDWG, and reviewed in September 2007. It is necessary to distribute the idea of GeoCASE on the widest possible level. The primary goal is getting ABCDEFG a standard for geoscientific collection database query tools.