Towards a machine-readable knowledge base of deep time: challenges, current progress, and future work

Xiaogang Ma (max@uidaho.edu), Chao Ma (chao@uidaho.edu)

University of Idaho, Moscow, ID, USA

Background

- The study of the co-evolving geosphere and biosphere needs various subjects of datasets to be integrated
- Geologic time can be used as a common reference to connect the data silos
- Heterogeneous geologic time standards are used in data records of different disciplines
- In the open data environment, We need a Web-based and machine-readable knowledge base of deep time to automate data access and synthesis in executable workflows

Data Sources

rruff.info

mindat.org

paleobiodb.org

onegeology.org

mrdata.usgs.gov

earthchem.org

rcsb.org

… …

Computer Technologies

- RDF and OWL building the conceptual framework of geologic time
- SKOS for encoding the global, national and regional geologic time standards
- SPARQL for querying the knowledge base
- Interfaces to R and Python

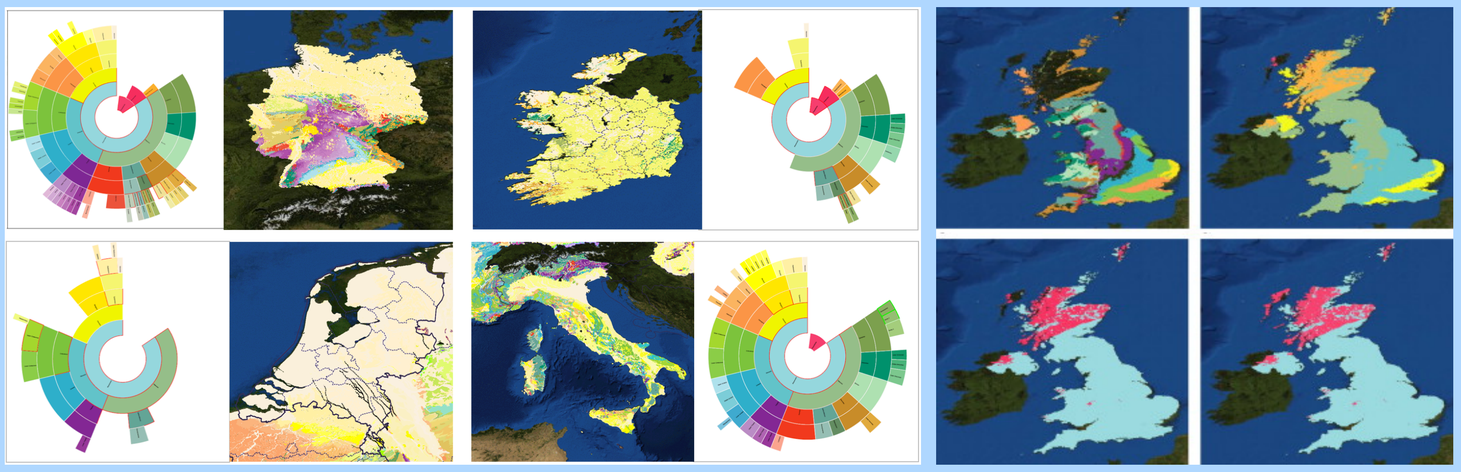
Plans for Future Work

Existing initial works show that we can further extend and leverage the knowledge base of deep time in executable workflows for data-intensive research

- We need methods to address the heterogeneity among local, regional and global geologic time standards and represent them in a machine-readable model
- We need to build services of the knowledge base that allow machines to access. A SPARQL endpoint can be build. Interfaces such as an R package and a Python Module can be built for querying the knowledge base from workflow platforms
- A website can be built for the visualization of the knowledge base, which allows human users to query and look for topics of interest in the knowledge base
- Other topics, such as information of major paleo-biotic, paleo-geographic and paleo-climatic events, can also be added to the knowledge base in the future

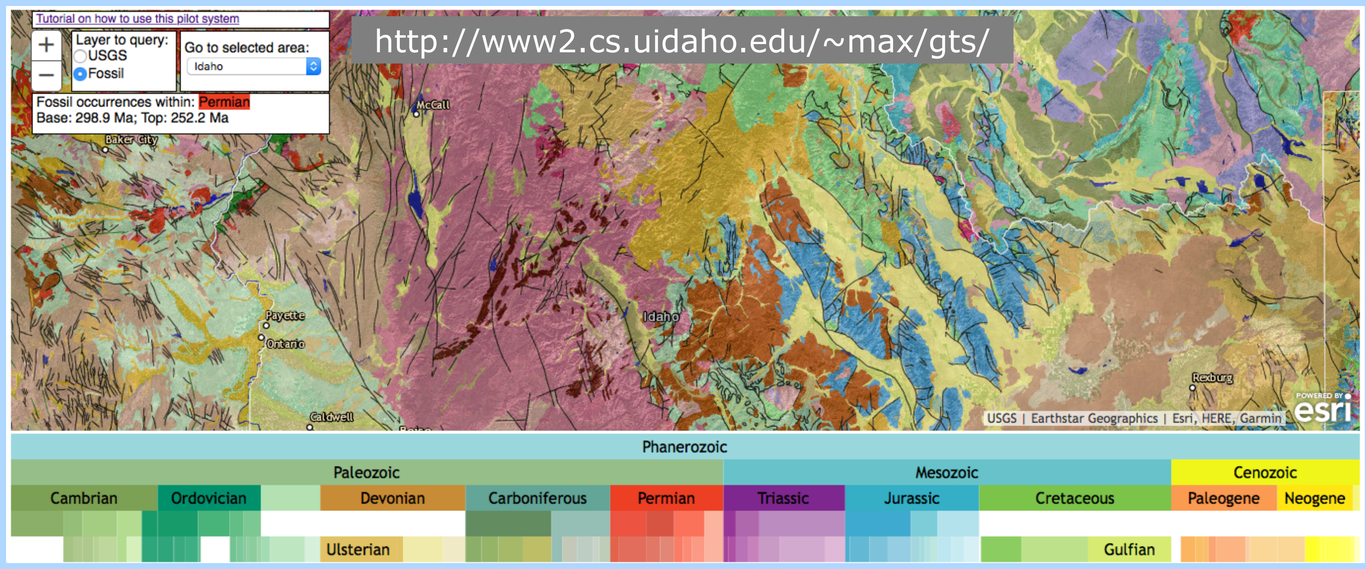
Initial Work 1: Using a visualization of geologic time to interact with WMS geologic map layers

- Using the metadata from WMS map layers, we can show the patterns of time concepts in each geologic map
- Using the ordinal-hierarchical structure of geologic time terms in the knowledge base, we can generalize features in geologic map layers according to the time records



Initial Work 2: Exploring regional geologic time and paleontological information of USA

- Build and visualize an ontology for the local geologic time scale of North America (adapted source code of visualization from PaleoBioDB)
- Ontology-driven retrieval and display of fossil occurrences and geologic maps for exploratory analysis



Initial Work 3: Represent and capture concept and attribute versioning in geological time scale

- There are continuous changes in the international geologic time scale, e.g. adding and removal of concepts, or changes in the attribute of concepts
- A semantic model was developed to represent those versioning changes. A case study was done to capture all the changes in the International Chronostratigraphic Chart from 2004 to 2018

We adapted the existing vocabularies developed by Dr. Simon Cox and the IUGS-CGI group to improve interoperability.

DOI of related publications: 10.1016/j.cageo.2018.03.004; 10.1016/j.cageo.2011.07.018; 10.1007/s12145-017-0304-8; 10.1016/j.cageo.2011.02.011 10.1007/s12145-013-0110-x;

Acknowledgement: This presentation is based upon work supported by the National Science Foundation under Grant No. 1835717.