Simon Goring, University of Wisconsin Andrea K. Thomer, University of Michigan Shane Loeffler, University of Minnesota Stephen C. Kuehn, Concord University
Anders J. Noren, University of Minnesota Nicholas P. McKay, Northern Arizona University Douglas Fils, Ocean Leadership Kerstin A. Lehnert, Columbia University

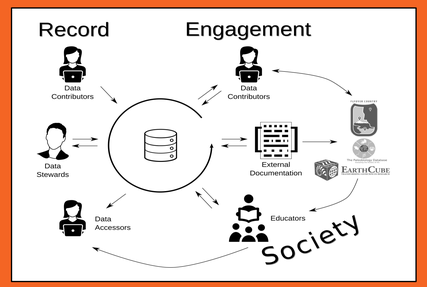
Throughput

http://throughputdb.com

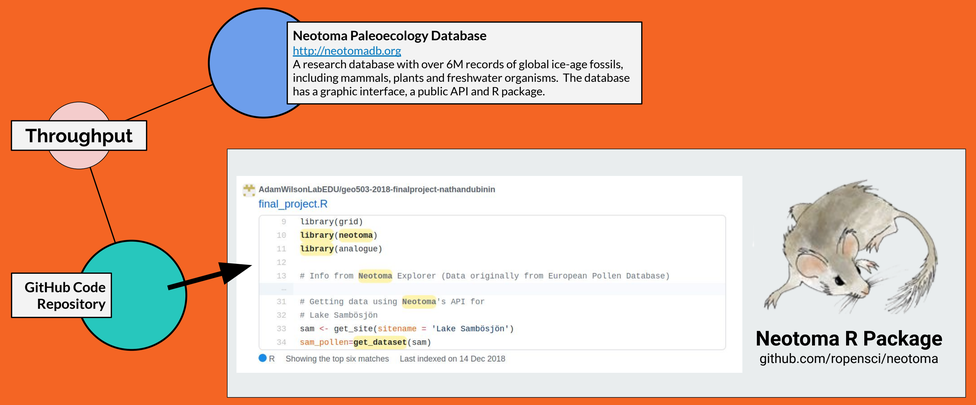
A tool to connect research data and code examples to improve learning opportunities and help build better documentation

Data is Central to Education

- Learning opportunities often rely heavily on data. Data may be generated by individuals learning within a program, may be provided by an instructor, or may be collected from external sources.
- Successful learning, particularly within sub-domains in the geosciences, can rely heavily on specialized knowledge of particular datasets, however this knowledge is unevenly distributed within academia.
- Data resources play a central role in providing data, using online platforms, digital catalogs and, often, documentation.
- Lack of proper documentation, difficulties in accessing data, and a lack of example workflows, particularly for complex datasets and analysis can result in barriers to equitable access to scientific knowledge and education.



Databases can be tools for Engagement. However engaging fully with a diverse user community can be complex for teams with little engagement experience.



New Technology Requires New Tools

- Statistical and scientific computing means that teaching examples often need new data sources.
- Student projects may be beyond the capacity of instructors to provide guidance.
- **Throughput (http://throughputdb.com) connects research databases with online code examples**. The code within these repositories can provide templates for early career researchers, or educators looking to build robust examples for instruction.
- Throughput will soon support the ability to annotate research data and code examples, to help improve documentation, and to act as an informal mechanism to pass on information among educators, researchers and learners.

Project

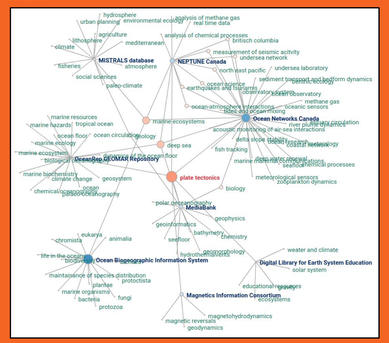
All project repositories (http://github.com/throughput-ec) are open and free to contribute within the bounds set in the Code of Conduct. All code is licensed under an MIT License, free to use and modify without credit.

Database snapshots available at:
https://figshare.com/collections/Throughput_Database_Snapshots/5075912

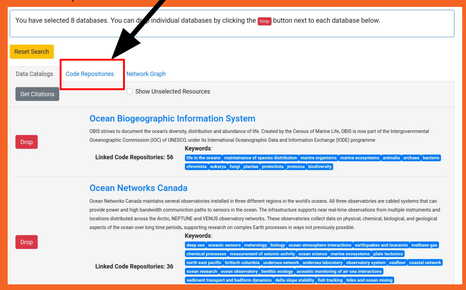
Goring SJ, Graham R, Loeffler S, Myrbo A, Oliver JS, Ormond C, & Williams JW. 2018. *The Neotoma Paleoecology Database: A Research Outreach Nexus*. Elements of Paleontology. Cambridge: Cambridge University Press. DOI: 10.1017/9781108681582

Discovery and Annotation

- At its heart Throughput is a set of interconnected objects. Databases, people, grants, code examples, text annotations and journal articles are all represented using unique identifiers.
- Each object can be connected to other objects, including Language elements, Audience information and Keywords. The resulting network is what powers the Throughput tool-set.
- We can search for databases that share related keywords ("find all databases that contain climate data"), or we can look for code resources that combine data from multiple data sources ("find a code repository that uses soil and hydrological data").
- Importantly, we can also tag individual elements within the database to help others, and make new connections ("This lesson plan uses data from NEPTUNE Canada").



Screenshot from http://throughputdb.com showing the data and code search capabilities.



Open Science, More Credit, More Collaboration

- Throughput provides a tool to obtain citations for data resources and code resoirces, as well as the individual annotations that make up Throughput itself.
- All development on Throughput is open source (http://github.com/throughput-ec) and collaboration is welcome.
- By exposing examples of data use and reuse by early career researchers we hope to normalize code sharing and credit for code
- Throughput also discovers and provides post hoc citations for links between code and publications using **xDeepDive** (http://geodeepdive.org) giving people the ability to showcase the impact of their work, even if it is not cited formally.

NSF
EarthCube
Transforming Geosciences Research