Seamless Long-Tail and Big Data Access via the EarthCube Brokering Cyberinfrastructure BALTO

D. Sarah Stamps[1], James Gallagher[2], Scott Peckham[3], Anne Sheehan[3], Nathan Potter[2], Kodi Neumiller[2], Emmanuel Njinju[1], Maria Stoica[3], Zachary Easton[1], Daniel Fuka[1], and David Fulker[2]

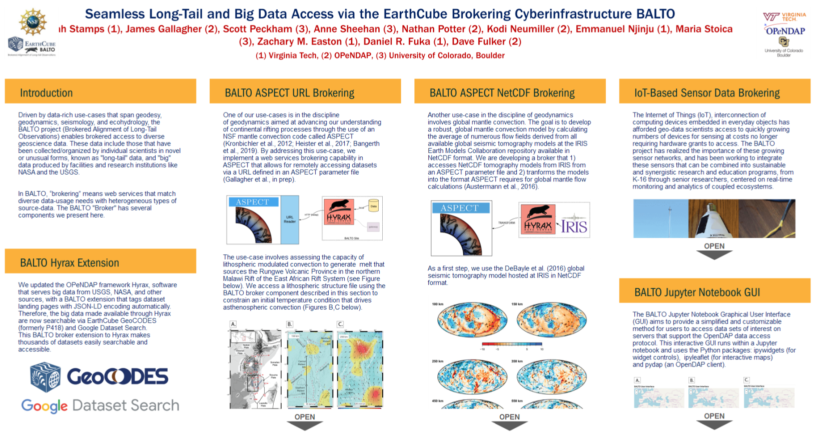
[1]Virginia Tech
[2]OPeNDAP
[3]University of Colorado Boulder

November 23, 2022

Abstract

The EarthCube BALTO broker (Brokered Alignment of Long-Tail Observations) provides streamlined access to both long-tail and big data using Web Services through several distinct mechanisms. First, we updated the OPeNDAP framework Hyrax, software that serves big data from USGS, NASA, and other sources, with a BALTO extension that tags dataset landing pages with JSON-LD encoding automatically. Therefore, the big data made available through Hyrax are now searchable via EarthCube GeoCODES (formerly P418) and Google Dataset Search. The BALTO broker extension to Hyrax makes thousands of datasets easily searchable and accessible. Second, we focused our efforts on a geodynamics use-case aimed at advancing our understanding of continental rifting processes through the use of an NSF mantle convection code called ASPECT. By addressing this use-case, we implemented a web services brokering capability in ASPECT that allows for remotely accessing datasets via a URL defined in an ASPECT parameter file. Third, through another use-case in ASPECT aimed at testing hypotheses involving global mantle flow, we developed a brokering mechanism for a "plug-in" that accesses NetCDF seismic tomography data from the NSF seismology facility IRIS, then transforms it into the format needed by ASPECT to run global mantle flow models constrained by seismic tomography. Fourth, we demonstrate methods to allow any scientist or citizen scientist to make their in-situ IoT based sensor data collection efforts available to the world. Finally, we are developing a Jupyter Notebook with a GUI that allows for users to search Hyrax servers for big datasets and long-tail data. These cyberinfrastructure developments comprise the entire EarthCube BALTO brokering capabilities.

Seamless Long-Tail and Big Data Access via the EarthCube Brokering Cyberinfrastructure BALTO



D. Sarah Stamps (1), James Gallagher (2), Scott Peckham (3), Anne Sheehan (3), Nathan Potter (2), Kodi Neumiller (2), Emmanuel Njinju (1), Maria Stoica (3), Zachary M. Easton (1), Daniel R. Fuka (1), Dave Fulker (2)

(1) Virginia Tech, (2) OPeNDAP, (3) University of Colorado, Boulder

VIRGINIA TECH

OPeNDAP

University of Colorado Boulder

PRESENTED AT:

EarthCube

2020 EarthCube Annual Meeting
Virtual – June 18, 2020

INTRODUCTION

Driven by data-rich use-cases that span geodesy, geodynamics, seismology, and ecohydrology, the BALTO project (Brokered Alignment of Long-Tail Observations) enables brokered access to diverse geoscience data. These data include those that have been collected/organized by individual scientists in novel or unusual forms, known as "long-tail" data, and "big" data produced by facilities and research institutions like NASA and the USGS.

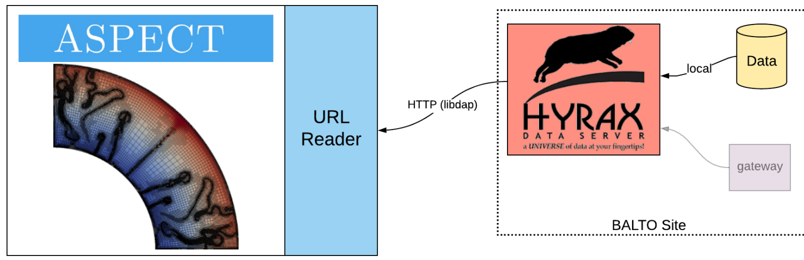
In BALTO, “brokering” means web services that match diverse data-usage needs with heterogeneous types of source-data. The BALTO "Broker" has several components we present here.

BALTO HYRAX EXTENSION

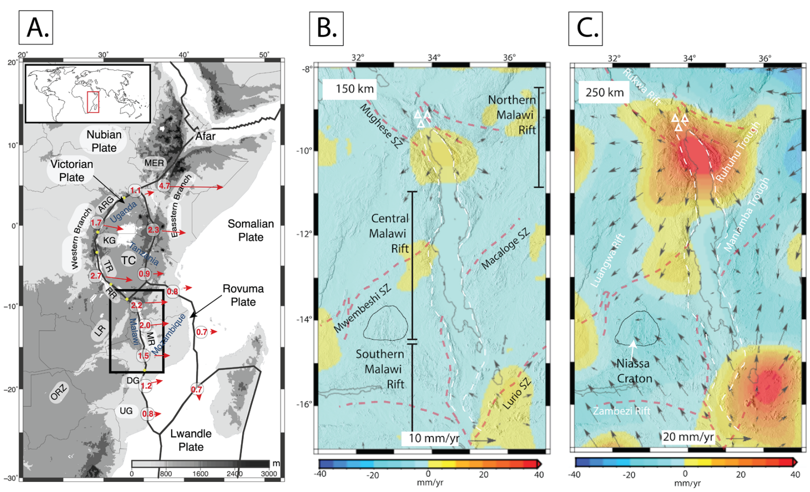
We updated the OPeNDAP framework Hyrax, software that serves big data from USGS, NASA, and other sources, with a BALTO extension that tags dataset landing pages with JSON-LD encoding automatically. Therefore, the big data made available through Hyrax are now searchable via EarthCube GeoCODES (formerly P418) and Google Dataset Search. This BALTO broker extension to Hyrax makes thousands of datasets easily searchable and accessible.

BALTO ASPECT URL BROKERING

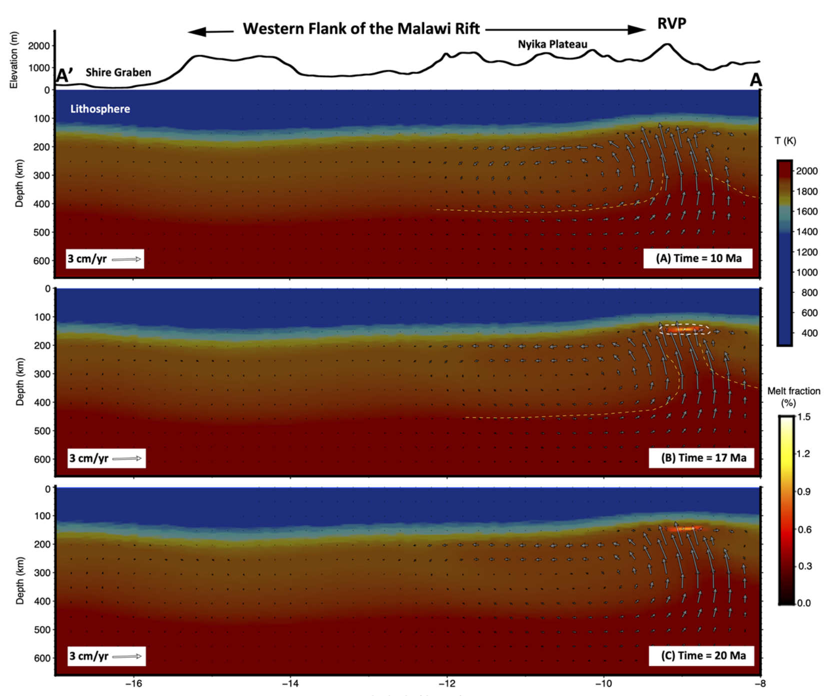
One of our use-cases is in the discipline of geodynamics aimed at advancing our understanding of continental rifting processes through the use of an NSF mantle convection code called ASPECT (Kronbichler et al., 2012; Heister et al., 2017; Bangerth et al., 2019). By addressing this use-case, we implement a web services brokering capability in ASPECT that allows for remotely accessing datasets via a URL defined in an ASPECT parameter file (Gallagher et al., in prep).



The use-case involves assessing the capacity of lithospheric modulated convection to generate melt that sources the Rungwe Volcanic Province in the northern Malawi Rift of the East African Rift System (see Figure below). We access a lithospheric structure file using the BALTO broker component described in this section to constrain an initial temperature condition that drives asthenospheric convection (Figures B,C below).



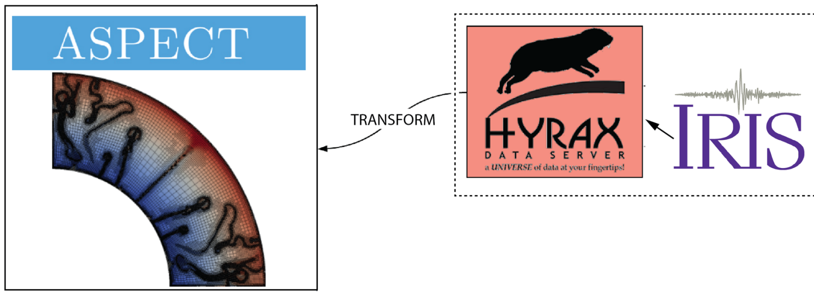
The geodynamic approach we use includes non-Newtonian, temperature-, pressure- and porosity-dependent creep laws of anhydrous peridotite for the sub-lithospheric convecting mantle that allows for calculating melt fractions. The model is run to steady-state (>15 Ma), after which melt is produced beneath the Rungwe Volcanic Province assuming relatively high mantle potential temperatures determined by Rooney et al. (2012).



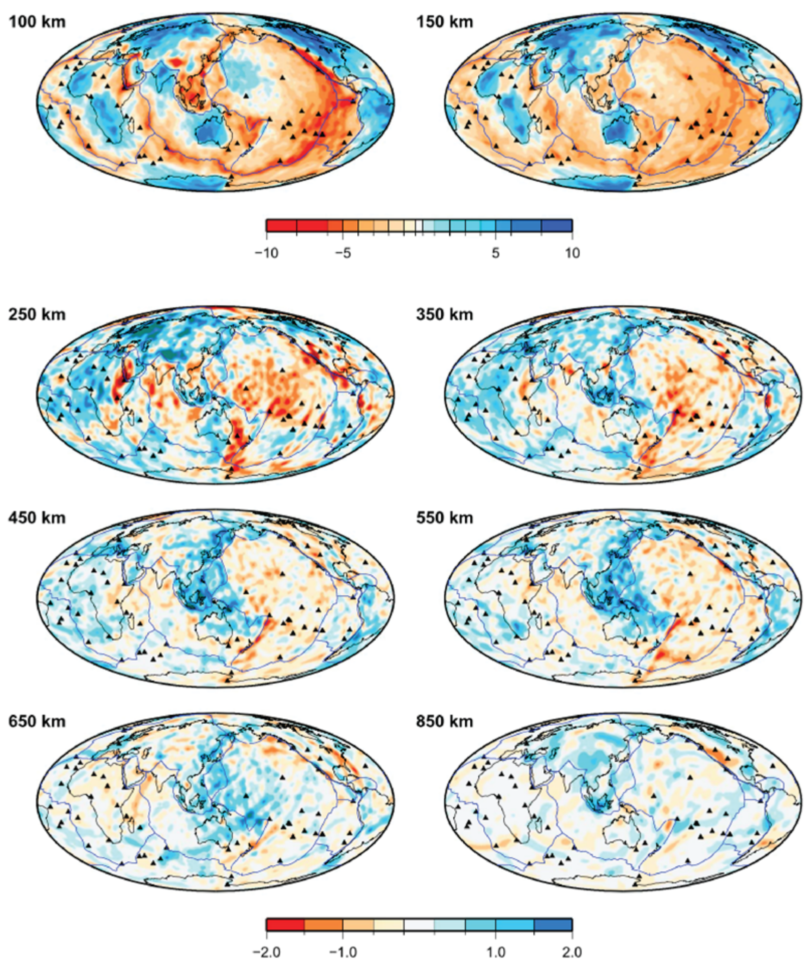
See Emmanual Njinju's and Kodi Neumiller's poster for more details.

BALTO ASPECT NETCDF BROKERING

Another use-case in the discipline of geodynamics involves global mantle convection. The goal is to develop a robust, global mantle convection model by calculating the average of numerous flow fields derived from all available global seismic tomography models at the IRIS Earth Models Collaboration repository available in NetCDF format. We are developing a broker that 1) accesses NetCDF tomography models from IRIS from an ASPECT parameter file and 2) tranforms the models into the format ASPECT requires for global mantle flow calculations (Austermann et al., 2016).



As a first step, we use the DeBayle et al. (2016) global seismic tomography model hosted at IRIS in NetCDF format.



Shear wave seismic velocity perturbations from the reference velocity in percent distribution at different depths from DeBayle et al., 2016 associated with model 3D2016-03Sv.

See Kodi Neumiller's poster for more details.

IOT-BASED SENSOR DATA BROKERING

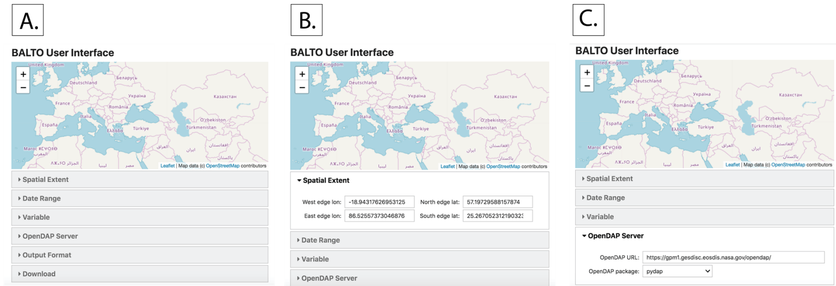
The Internet of Things (IoT), interconnection of computing devices embedded in everyday objects has afforded geo-data scientists access to quickly growing numbers of devices for sensing at costs no longer requiring hardware grants to access. The BALTO project has realized the importance of these growing sensor networks, and has been working to integrate these sensors that can be combined into sustainable and synergistic research and education programs, from K-16 through senior researchers, centered on real-time monitoring and analytics of coupled ecosystems.

BALTO takes advantage of the OpenSource Long-Range wireless communication protocol (LoRa) to connect sensors to EarthCube Architectures. capabilities.

BALTO has partnerships with representatives from 4H and and FFA to integrate these $25 multi-science stations into their learning programs.

BALTO JUPYTER NOTEBOOK GUI

The BALTO Jupyter Notebook Graphical User Interface (GUI) aims to provide a simplified and customizable method for users to access data sets of interest on servers that support the OpenDAP data access protocol. This interactive GUI runs within a Jupyter notebook and uses the Python packages: ipywidgets (for widget controls), ipyleaflet (for interactive maps) and pydap (an OpenDAP client).



See Scott Peckham's notebook presentation for more details.

ABSTRACT

The EarthCube BALTO broker (Brokered Alignment of Long-Tail Observations) provides streamlined access to both long-tail and big data using Web Services through several distinct mechanisms. First, we updated the OPeNDAP framework Hyrax, software that serves big data from USGS, NASA, and other sources, with a BALTO extension that tags dataset landing pages with JSON-LD encoding automatically. Therefore, the big data made available through Hyrax are now searchable via EarthCube GeoCODES (formerly P418) and Google Dataset Search. The BALTO broker extension to Hyrax makes thousands of datasets easily searchable and accessible. Second, we focused our efforts on a geodynamics use-case aimed at advancing our understanding of continental rifting processes through the use of an NSF mantle convection code called ASPECT. By addressing this use-case, we implemented a web services brokering capability in ASPECT that allows for remotely accessing datasets via a URL defined in an ASPECT parameter file. Third, through another use-case in ASPECT aimed at testing hypotheses involving global mantle flow, we developed a brokering mechanism for a “plug-in” that accesses NetCDF seismic tomography data from the NSF seismology facility IRIS, then transforms it into the format needed by ASPECT to run global mantle flow models constrained by seismic tomography. Fourth, we demonstrate methods to allow any scientist or citizen scientist to make their in-situ IoT based sensor data collection efforts available to the world. Finally, we are developing a Jupyter Notebook with a GUI that allows for users to search Hyrax servers for big datasets and long-tail data. These cyberinfrastructure developments comprise the entire EarthCube BALTO brokering capabilities.

REFERENCES

Austermann, J., Mitrovica, J. X., Huybers, P., & Rovere, A. (2017). Detection of a dynamic topography signal in last interglacial sea-level records. Science Advances, 3(7), e1700457.

Debayle, E., Dubuffet, F., & Durand, S. (2016). An automatically updated S‐wave model of the upper mantle and the depth extent of azimuthal anisotropy. Geophysical Research Letters, 43(2), 674-682.

Wolfgang Bangerth, Juliane Dannberg, Rene Gassmoeller, and Timo Heister. 2018, May 10. ASPECT v2.0.0. Zenodo. https://doi.org/10.5281/zenodo.1244587

Heister, T., Dannberg, J., Gassmöller, R., & Bangerth, W. (2017). High accuracy mantle convection simulation through modern numerical methods–II: realistic models and problems. Geophysical Journal International, 210(2), 833-851.

Kronbichler, M., Heister, T., & Bangerth, W. (2012). High accuracy mantle convection simulation through modern numerical methods. Geophysical Journal International, 191(1), 12-29.

Rooney, T. O., Herzberg, C., & Bastow, I. D. (2012). Elevated mantle temperature beneath East Africa. Geology, 40(1), 27-30.