



Brokered Alignment of Long-Tailed Observations (BALTO) Applications in Geoscience



University of Colorado Boulder



D. Sarah Stamps (dstamps@vt.edu, Virginia Tech), James Gallagher (OPeNDAP), Scott Peckham (University of Colorado), Anne Sheehan (University of Colorado), Nathan Potter (OPeNDAP), Maria Stoica (University of Colorado), Emmanuel Njinju (Virginia Tech), David Fulker (OPeNDAP), Kodi Neumiller (OPeNDAP), Zachary M. Easton (Virginia Tech), Robin R. White (Virginia Tech), Daniel R. Fuka (Virginia Tech)

INTRODUCTION

Driven by data-rich use cases that span geodesy, geodynamics, seismology, and ecohydrology, the BALTO project enables brokered access to diverse geoscience data, including data that have been collected/organized by individual scientists in novel or unusual forms, also known as “long-tail” datasets.

In BALTO, “brokering” means Web services that match diverse data-usage needs with heterogeneous types of source-data. This matching addresses form and semantics, which includes protocols, data structures, encodings, units of measure, variable names, and sampling meshes. The BALTO broker employs an extensible hub-and-spoke architecture: its hub will combine well-established, open-source, data-as-service software (from OPeNDAP) with the Geoscience Standard Names (GSN) to establish canonical representations for brokered datasets; each spoke—called an accessor—comprises (source-specific) data-access software along with metadata mappings that yield GSN-compliant variable names.

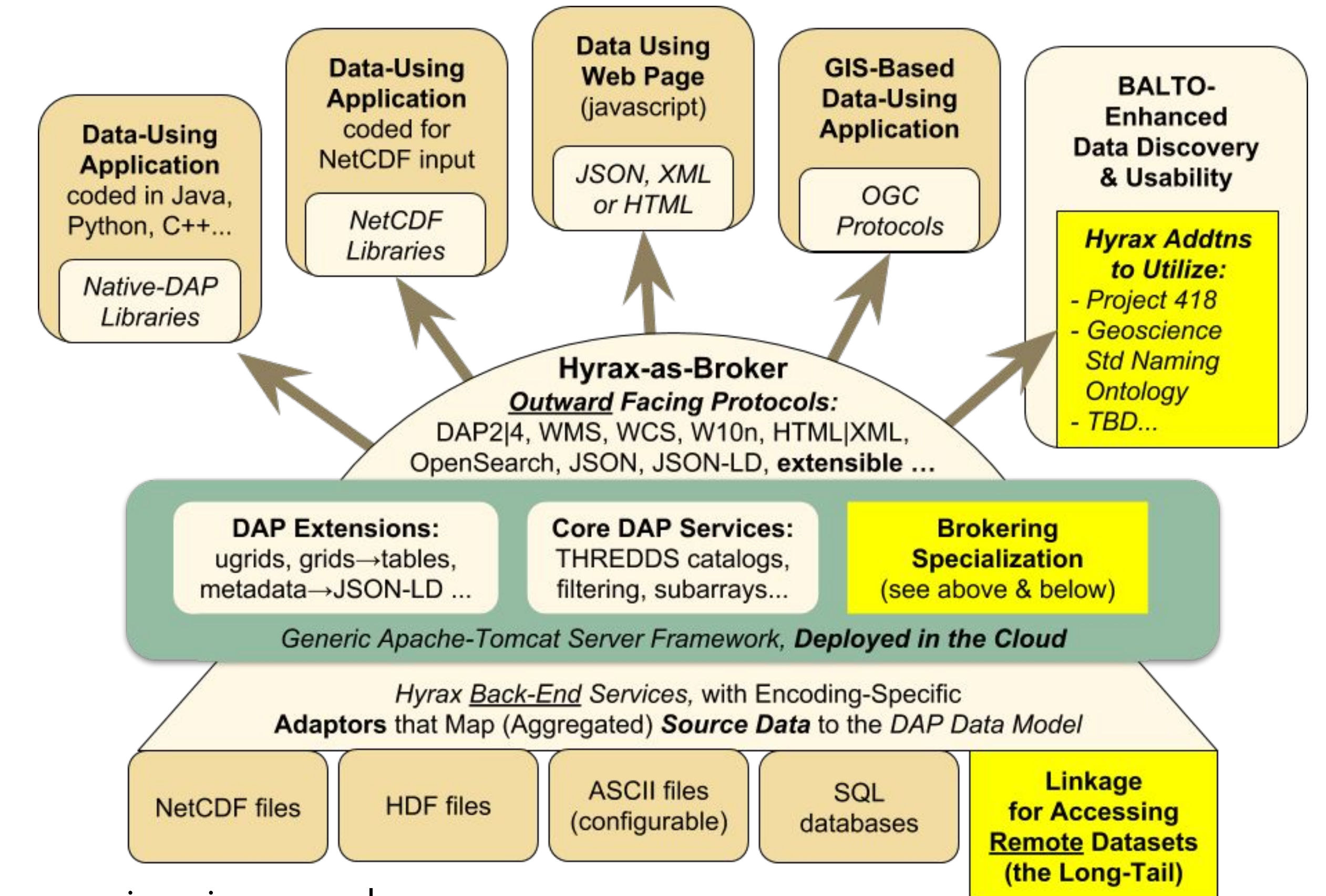
GEOCODES, GOOGLE DATASET SEARCH, AND JSON-LD

To enhance data discovery, GeoCODES and Google Dataset Search now harvest EarthCube dataset descriptions that are encoded with JSON-LD (Javascript Object Notation - Linked Data). BALTO’s (Hyrax-based) broker has long produced landing pages and dataset-request forms for its collections and datasets (including NASA, USGS, and others), and these HTML documents now contain descriptions encoded in JSON-LD. Hence the Google Dataset Search and GeoCODES can include a large number of DAP-accessible datasets, well beyond those from BALTO, per se. As a result, the integration of BALTO with GeoCODES may well enable searches of unprecedented sophistication because dataset descriptions from Hyrax typically are rich with metadata.



BALTO CYBERINFRASTRUCTURE FOR DISTRIBUTED BROKERING

Built on Hyrax, the BALTO broker yields middleware-style interoperability (order N+M complexity). Adapters map N types of source-data to the DAP Data Model, through which data in M formats may be retrieved via Web services.



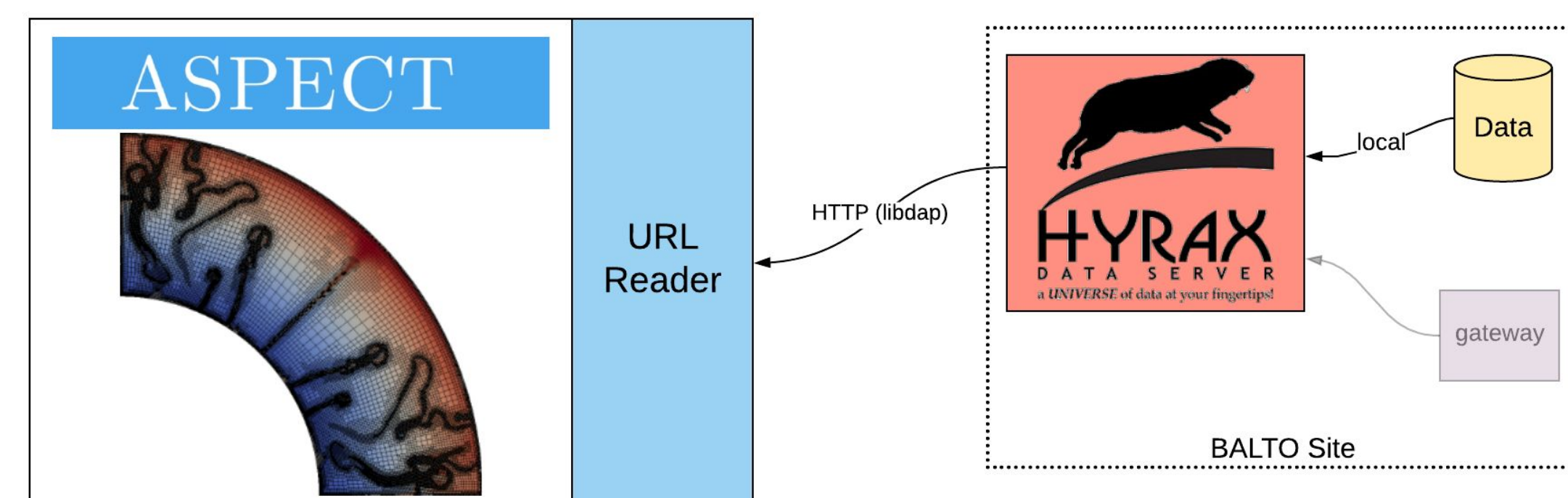
THREDDS Catalogs bolster these services in several ways:

1. Data users retrieve them to learn what’s available.
2. Data providers employ them to configure their servers and create *virtual* views of their holdings.
3. New advances permit them to reference remote data (from services that produce THREDDS catalogs) to permit distributed brokering & indirect retrievals.

Recent Hyrax advances also permit P418 harvesting of JSON-LD metadata about collections & datasets.

COMPUTATIONAL MODELING WITH CIG CODE ASPECT

Computational modeling often requires input files for constraining initial conditions. We have developed the capability to access, seamlessly, csv files through parametric files of the NSF CIG code ASPECT (Advanced Solver for Problems in Mantle Convection; Kronbichler et al., 2012; Heister et al., 2017; Bangert et al., 2018).



Our use-case is a lithosphere-asthenosphere system mantle convection model focused on the Malawi Rift of the East African Rift System. This model uses a lithospheric structure to constrain the initial temperature condition of the convection model (see poster by Emmanuel Njinju). For this model, the parameter file is set-up such that the user accesses, seamlessly through brokering, a csv file on the BALTO server.

In this implementation, ASPECT dereferences a URL using libdap and retrieves CSV ASCII data. It stores this data as a series of vectors. ASPECT then goes through these vectors and appends the values onto a stringstream and uses its existing sub-systems to read the data just as it would a local file.

Bangert, W.; Dannberg, J.; Gassmüller, R.; Heister, T. (2018), ASPECT v2.0.1 [software], doi: [10.5281/zenodo.1297145](https://doi.org/10.5281/zenodo.1297145), url: <https://doi.org/10.5281/zenodo.1297145>
Kronbichler, M.; Heister, T.; Bangert, W. (2012), High accuracy mantle convection simulation through modern numerical methods, Geophysical Journal International, 191 (1), 12-29, doi: [10.1111/j.1365-246X.2012.05609.x](https://doi.org/10.1111/j.1365-246X.2012.05609.x), url: <http://gji.oxfordjournals.org/cgi/doi/10.1111/j.1365-246X.2012.05609.x>
Heister, T.; Dannberg, J.; Gassmüller, R.; Bangert, W. (2017), High accuracy mantle convection simulation through modern numerical methods - II: realistic models and problems, Geophysical Journal International, 210 (2), 833-851, doi: [10.1093/gji/ggx195](https://doi.org/10.1093/gji/ggx195), url: <https://academic.oup.com/gji/article-lookup/doi/10.1093/gji/ggx195>

BROKERING INTEGRATED SENSOR NETWORKS



The Internet of Things (IoT), interconnection of computing devices embedded in everyday objects, has given 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 communication protocol (LoRa) to connect sensors to EarthCube Architectures.

\$25 Multi-Science Station
WX Solar Soil
Battery Power Reg. Arduino LoRa 10DF Motion GPS

STEM

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