

EGU24-12962, updated on 09 Nov 2024
<https://doi.org/10.5194/egusphere-egu24-12962>
EGU General Assembly 2024
© Author(s) 2024. This work is distributed under
the Creative Commons Attribution 4.0 License.



The *Marble* climate informatics platform: data discovery and data access

Deepak Chandan, **Misha Schwartz**, and Steve Easterbrook
University of Toronto, Computer Science, Toronto, Canada (dchandan@atmosp.physics.utoronto.ca)

Advances in remote sensing and computing infrastructure, and demands of modern climate research are driving the production of new climate datasets at a breathtaking pace. It is increasingly felt by researchers, that the growing volume of climate datasets is challenging to store, analyze or generally "shepherd" through their analysis pipelines. Quite often, the ability to do this is limited to those with access to government or institutional facilities in wealthier nations, raising important questions around equitable access to climate data.

The Data Analytics for Canadian Climate Services (DACCS) project has built a cloud based network of federated nodes, called Marble, that allows anyone seeking to extract insights from the large volumes of climate data to undertake their study without concerning themselves with the logistics of acquiring, cleaning and storing data. The aspiration for building this network is to provide a low-barrier entry not only to those working in core climate change research, but also to those involved in climate mitigation, resilience and adaptation work and to policy makers, non-profits, educators and students. Marble is one of the platforms selected to contribute to the 'Open Science Platform' component of the OGC's OSPD initiative.

The user-facing aspect of the platform is comprised of two components: (i) the Jupyter compute environment and (ii) the data server and catalogue. Here, we focus on the latter and present details of the infrastructure, developed on top of proven open-source software and standards (e.g. STAC), that allows for discovery and access of climate datasets stored anywhere on the network by anyone on the network. We will also discuss the publication capability of the platform that allows a user to host their own data on the network and make it quickly available to others.