EGU2020-12386
https://doi.org/10.5194/egusphere-egu2020-12386
EGU General Assembly 2020


# Data dissemination best practices and challenges identified through NOAA's Big Data Project

**Meredith Richardson**[1], Ed Kearns[1], and Jonathan O'Neil[2]
[1]Office of the Chief Information Officer (OCIO), NOAA, United States of America
[2]BDP Director, OCIO, NOAA, United States of America

Through satellites, ships, radars, and weather models, the National Oceanic and Atmospheric Administration (NOAA) generates and handles tens of terabytes of data per day. Many of NOAA's key datasets have been made available to the public through partnerships with Google, Microsoft, Amazon Web Services, and more as part of the Big Data Project (BDP). This movement of data to the Cloud has enabled access for researchers from all over the world to vast amounts of NOAA data, initiating a new form of federal data management as well as exposing key challenges for the future of open-access data. NOAA researchers have run into challenges of providing "analysis-ready" datasets to which researchers from varying fields can easily access, manipulate, and use for different purposes. This issue arises as there is no agreed-upon format or method of transforming traditional datasets for the cloud across research communities, with each scientific field or start up expressing differing data formatting needs (cloud-optimized, cloud-native, etc.). Some possible solutions involve changing data formats into those widely-used throughout the visualization community, such as Cloud-Optimized GeoTIFF. Initial findings have led NOAA to facilitate roundtable discussions with researchers, public and private stakeholders, and other key members of the data community, to encourage the development of best practices for the use of public data on commercial cloud platforms. Overall, by uploading NOAA data to the Cloud, the BDP has led to the recognition and ongoing development of new best practices for data authentication and dissemination and the identification of key areas for targeting collaboration and data use across scientific communities.