



EMS Annual Meeting Abstracts

Vol. 20, EMS2023-83, 2023, updated on 17 May 2024

<https://doi.org/10.5194/ems2023-83>

EMS Annual Meeting 2023

© Author(s) 2024. This work is distributed under the Creative Commons Attribution 4.0 License.



From Data to Decision - Using Open Data in the Cloud

Chris Stoner

Amazon Web Services, AWS Open Data Team, United States of America (cstner@amazon.com)

Cloud computing has opened the doors to large scale analysis of environmental data. AWS created the AWS Open Data program to help Governments, research institutions, and private companies share massive amounts of data publicly in the cloud. Sharing data in the cloud means that the data in the cloud can be utilized by many reseachers at the same time, without the time consuming download before research can start. Researchers no longer have to find storage for the data they want to analyze, rather they can use the data in place in the cloud. The AWS Open Data program hosts over 100 Petabytes of data for public use, with a broad range of compute and data analytics products, including Amazon EC2, Amazon Athena, AWS Lambda, Amazon SageMaker and SageMaker Studio Lab, and Amazon EMR to help the community use the data in the cloud at scale. Sharing data in the cloud lets data users spend more time on data analysis rather than data acquisition, and the Open Data program offers the Registry of Open Data to find and use these data for analysis. Join this session to hear first hand from organizations sharing data how and why they have put data into the Open Data Program, discover best practices for sharing data in the cloud, learn how to find publicly available datasets through the Registry of Open Data on AWS, find out how to use data in the cloud with templates and code samples that you can take home to try, and how you can share your own data through the AWS Open Data program.