

Harvard Dataverse and the MOC

Mercè Crosas, Harvard University

Abstract

Harvard Dataverse, with 30,000 deposited datasets and 250,000 data files, is one of the largest general purpose repositories open to researchers world-wide for sharing, citing, and archiving data. It currently runs on AWS. This talk will introduce our proposed plan to run Harvard Dataverse on the MOC, while still maintaining a copy on AWS for redundancy. Furthermore, these efforts can eventually provide a generic multi-cloud option for any Dataverse repository, where a repository can be easily replicated and synchronized when needed between (private or public) clouds.

Speaker Bio



Mercè Crosas is the Chief Data Science and Technology Officer at the Institute for Quantitative Social Science (IQSS) at Harvard University. She has more than 10 years of experience leading the Dataverse project and more than 15 years of experience building data management and analysis systems in academia and biotechnology companies. She is part of numerous committees and working groups focused on data sharing, research data management, data citation, and data standards. Crosas is co-PI of the Dataverse Project, with IQSS director Gary King, and supervises the Consilience project for text analysis, the Data Science Services, and Data Curation team at IQSS. She collaborates with a wide range of data related projects, including the NIH Data Commons, Harvard Privacy Tools, DataTags, Data Provenance, the Structural Biology Grid Data , Cloud Dataverse with the Massachusetts Open Cloud, and the Boston Area Research Initiative, among others.

Crosas holds a Ph.D. in Astrophysics and a B.S. in Physics. Learn more at mercecrosas.com and [@mercecrosas](https://twitter.com/mercecrosas) on Twitter.