SciDataCon 2025



Contribution ID: 103

Type: Presentation

Sustainable Open Data Infrastructure to Protect Government Data during Regime Changes: African Examples

Tuesday 14 October 2025 17:06 (11 minutes)

1. Introduction and Background Recent policies and laws in the US that threaten data access have highlighted the need for infrastructure to ensure important government data is not lost during regime changes. US federal government information has been in the public domain since the 1895 Printing Act which prohibited any copyright on federal government publications (United States Congress, 1895: 608). However, ideologically motivated political interference can threaten government data even in the most robust data ecosystems. Webb and Kurtz (2022: 68) reveal data removal by administrators at 12 federal agencies during 2017-2020, and (Nost et al., 2021: 1) calculate that about 20% the EPA's website content was removed between 2016 and 2020. In 2025 executive orders of the (The White House, 2025a; 2025b) that targeted federal programmes related to diversity, equity and environmental protection led to many departments and offices removing datasets related to gender, race and climate change {Choo, 2025 #1362@1;Mallapaty, 2025 #1358. A further executive order of the (The White House, 2025c) to pause foreign aid obligations and disbursements ended many US programmes to co-collect and disseminate data with LMICs (Offord, Cohen and Enserink, 2025) and shuttered agency data sites (Mallapaty, 2025).

Regime changes in African countries have also impacted access to government data. The government of Tanzania withdrew from the (Open Government Partnership, 2017) and reneged on their open government data (OGD) commitments in 2017 after a change of government. (Mutuku and Idriss, 2019: 557) note that early Nigerian and Ivorian OGD initiatives also did not survive changes of leadership. Supportive OGD infrastructures such as open data repositories can help data to endure through such crises (Jarvenpaa and Essen, 2023: 11).

2. Data Rescue at DataFirst

The team at DataFirst, an African research data service, has stepped in when African data is threatened with erasure. Apart from regime changes, threats to data originate from the absence of robust data infrastructure in many African countries, which hamper ongoing preservation of and access to both government data and research datasets (Willoughby, 2019: 239). DataFirst (2025) is an African research data service and internationally certified repository based at the University of Cape Town (UCT) The DataFirst team have saved data from South Africa's first income panel survey when the survey site was closed during institutional changes. Data sustainability can also face risks other than political interference, such as media deterioration and loss of contextual knowledge (Mayernik et al., 2020: 5). This was the case with the 2015-2016 rescue of data in hard copy from two South African surveys documenting outcomes for victims of early and later forced removals. These rich historical datasets are available on our data site.

DataFirst's US data rescue efforts focus on two valuable data series. First, USAID-funded DHS program data, which includes African demographic and health data collected since the 1980s (DHS Program, 2025). The discontinuation of the DHS program will have negative consequences for SDG progress monitoring by African governments, as DHS data has been used heavily for this purpose (Woolfrey, 2020). After the shuttering of US-AID, it was no longer possible to register projects with the DHS Program, but an API enabled bulk downloads of the data files by those with already registered projects. DataFirst downloaded all African datasets, and we are currently creating metadata records and organising data sharing permissions with African government agencies, and will apply for further permissions to share the rest if DHS Program access is not restored. The second data series targeted for data rescue is education data from USAID funded projects. Access to this data was lost with the shuttering of the USAID's Development Data Library. DataFirst's AFLEARN project, which focuses on foundational learning, offered project implementers in African countries the option to upload

their data via a Nextcloud file hosting service to a secure server at UCT. With agreements from implementers DataFirst will provide long-time hosting and secure and sustained access to these threatened datasets.Open Government Data as A Risk Reduction Strategy

The 2025 data rescue efforts exposed weaknesses in funding agencies'data management policies and procedures. USAID's data policy fell short of implementing Obama-era policy and legislation to make openness the default for government data, including data from federally funded scientific research (United States Agency for International Development, 2020: 10-12, 16-17) As a result, Africa-based USAID project implementers, who had in some cases effectively signed over their data to USAID did not have sharing permissions and were therefore hesitant to place the data in rescue repositories. A policy fully aligned with federal open data legislation could be a strong data sustainability mechanism. Data rescue obstacles relating to ownership demonstrate that building OGD infrastructure can be a risk reduction strategy to preserve data as a societal good (Dodds and Wells, 2019: 260-267).

Primary author: WOOLFREY, Lynn (University of Cape Town, DataFirst)

Presenter: WOOLFREY, Lynn (University of Cape Town, DataFirst)

Session Classification: Presentations Session 5: Rigorous, responsible and reproducible science in the era of FAIR data and AI / Infrastructures to Support Data-Intensive Research

Track Classification: SciDataCon2025 Specific Themes: Infrastructures to Support Data-Intensive Research - Local to Global