GAGE
National Science Foundation's
Geodetic Facility for the
Advancement of Geoscience

The Future of Published Papers: Improving Reproducibility Practices for USGS Publications

Alexandra M. Cohen[1,2] and Shannon Q. O'Neill[1,3]

UNAVCO

Background

Ideally, all published scientific works should be repeatable by another scientist reading the publication. Unfortunately, due largely to factors such as unfindable or unusable data, software, and samples, this is not always the case. As technology advances and scientists continue to tackle new questions, methods for maintaining reproducible publications must evolve to establish a basis for a more ideal and efficient scientific process.
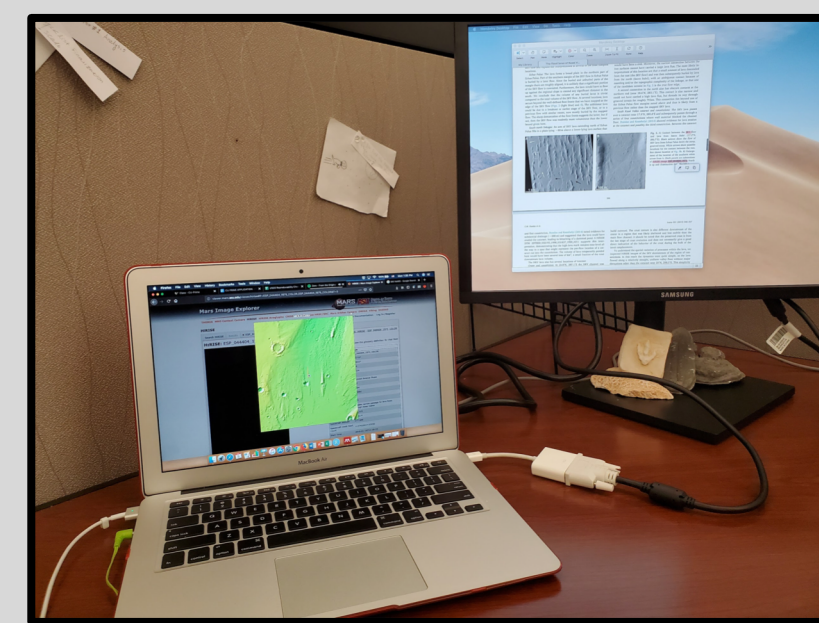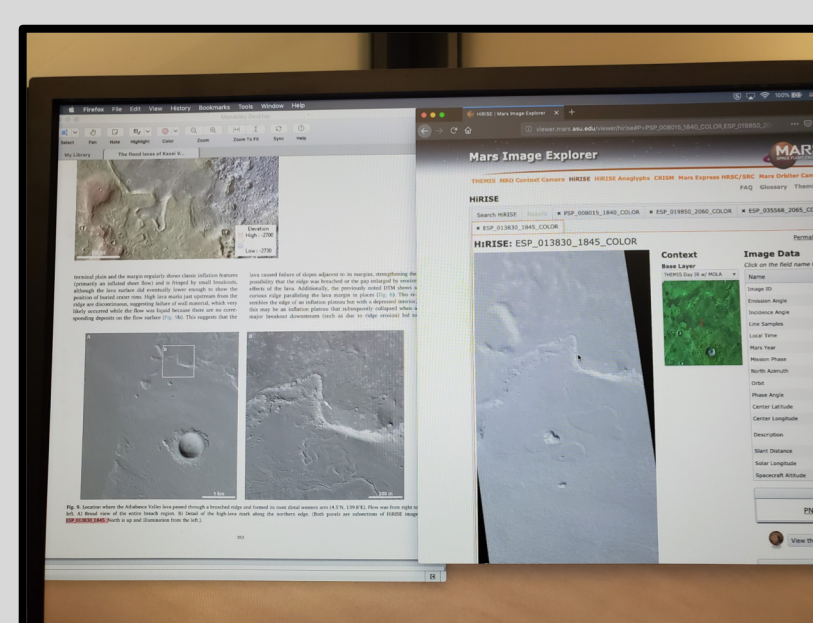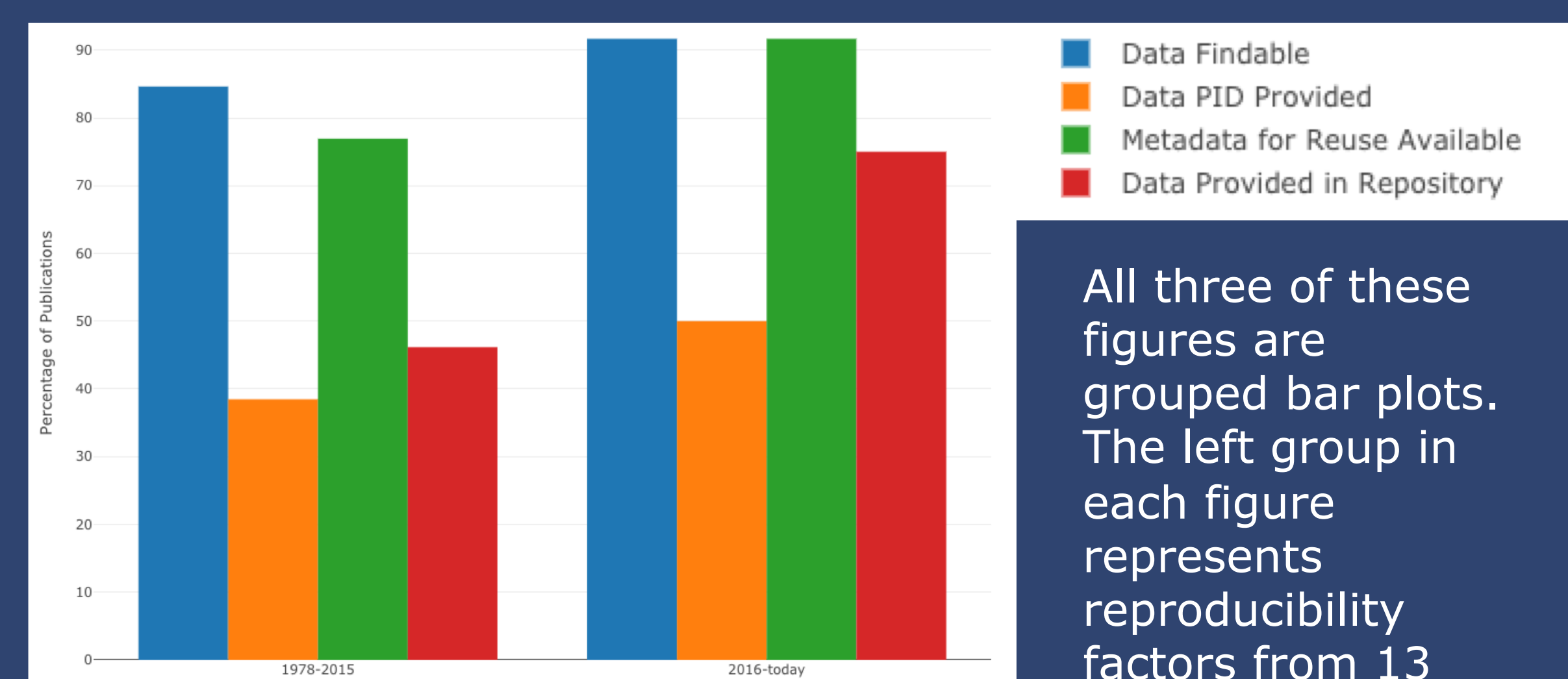
Study Focus

This project focuses on 25 publications associated with the USGS from the years of 1978-2019. Disciplines represented here include ecology, astrogeology, and geology. In 2016, the USGS enacted several policies aimed at increasing the reproducibility of their science. Through this project, we aim to identify where these policies have aided in the overall reproducibility of USGS research, and where improvement is still needed.

Analysis & Outcomes

Data Factors

Legend: Data Findable, Data PID Provided, Metadata for Reuse Available, Data Provided in Repository

All three of these figures are grouped bar plots. The left group in each figure represents reproducibility factors from 13 papers published between the years of 1978 and 2015. The right represents the same reproducibility factors for 12 papers published between 2016 and 2019.
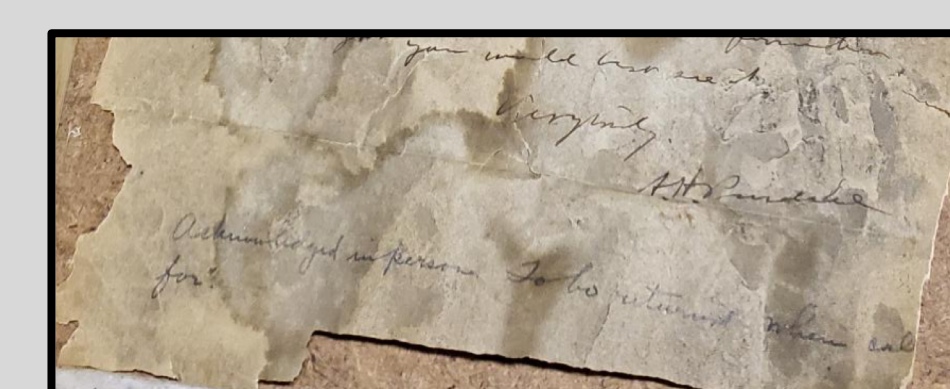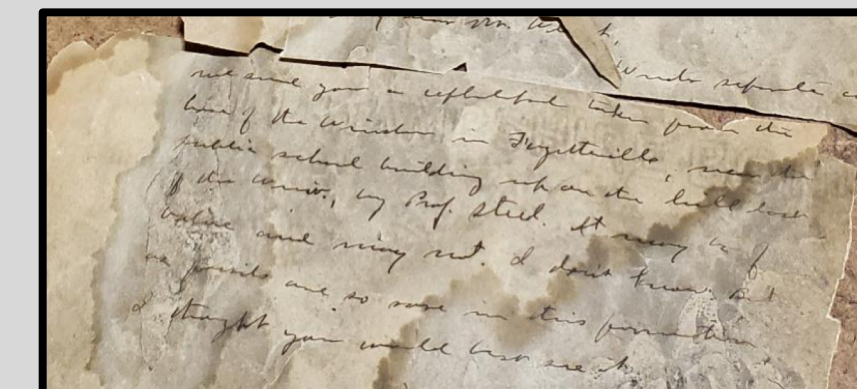
Software Factors

Legend: Software Findable, Software PID Provided, Software in a repository
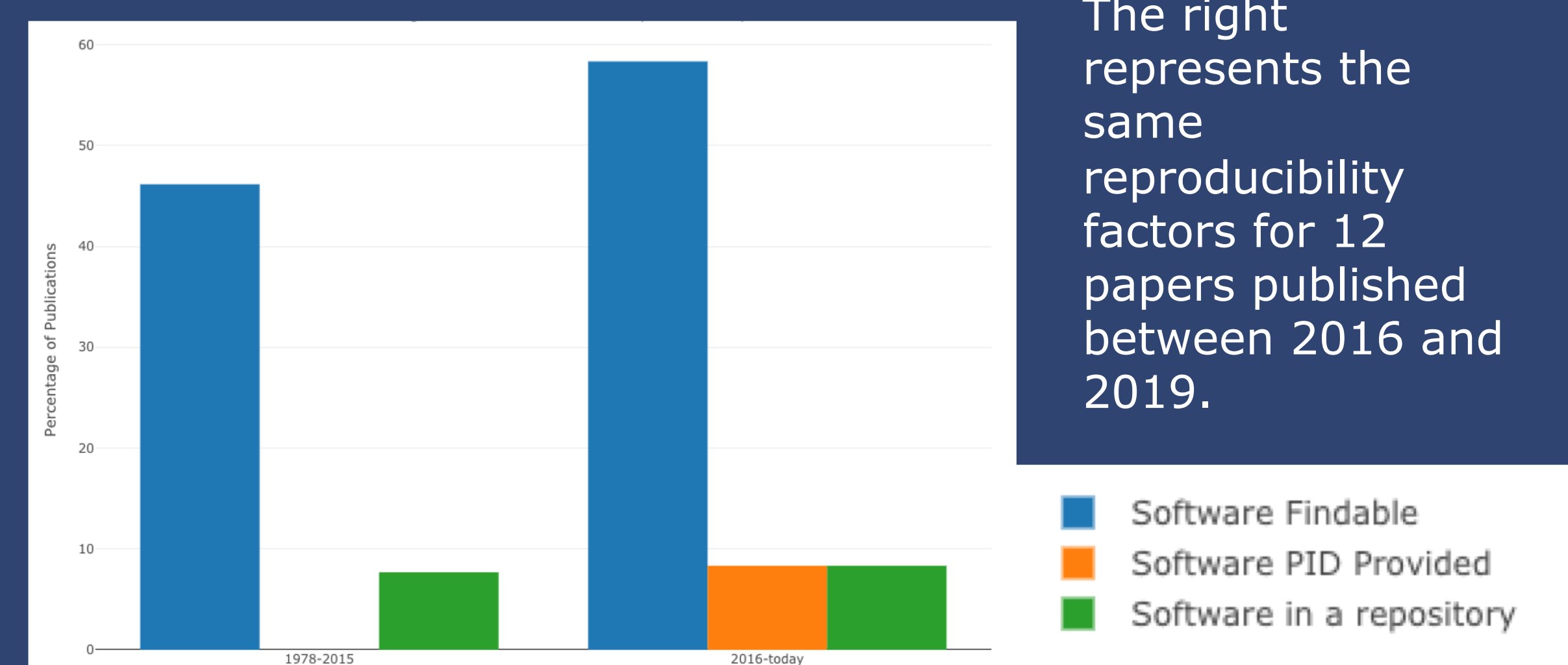
Other Factors
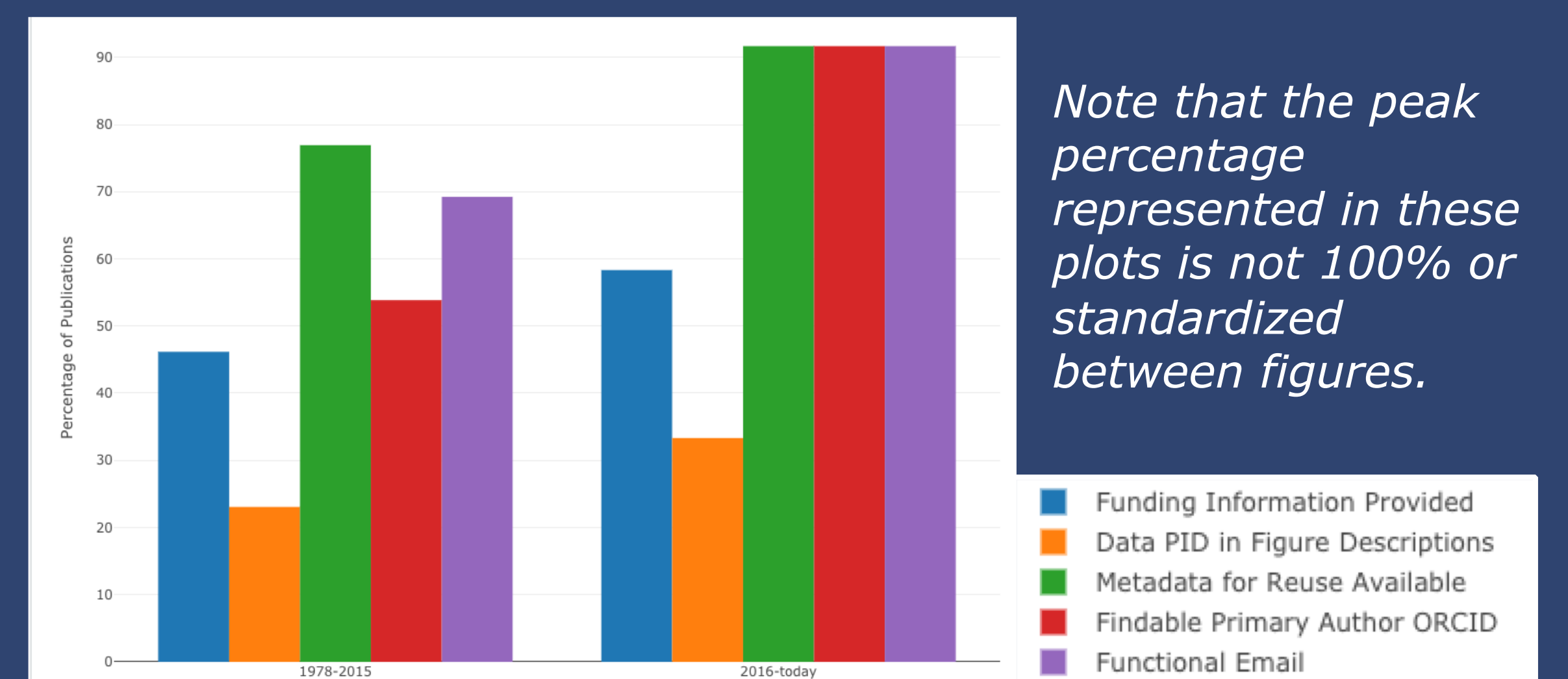
Note that the peak percentage represented in these plots is not 100% or standardized between figures.

Legend: Funding Information Provided, Data PID in Figure Descriptions, Metadata for Reuse Available, Findable Primary Author ORCID, Functional Email

Improvements In Progress

- ❏ Author Contact & ORCIDs
- ❏ Transparent Funding Info
- ❏ Findable Data & Software
- ❏ Extensive & Reusable Metadata

Improvements Still Needed

- ❏ Data Persistent Identifiers (PIDs) Provided
- ❏ Data Published in Repositories
- ❏ Software PIDs Provided
- ❏ Software Published in Repositories
- ❏ Relevant Data PIDs in Figure Descriptions
- ❏ Relevant Computational Provenance in Figure Descriptions

Methods Roadmap

We chose and analyzed 25 publications associated with USGS, based on publication year (pre-2016/post-2016) and discipline (Ecology/Astrogeology/Geology).

We created a rubric to assist in the determination of factors which contribute to the reproducibility of each paper's results. 18 of these factors were quantifiable and comparable.

We compiled data for all 25 publications into a machine readable format. We used R-Studio Plotly package to create visual comparisons of reproducibility factors before and after 2016.

References

For a complete works cited, or for more information about this project, please visit
DOI: 10.6084/m9.figshare.8947310

Geo-Launchpad

FRONT RANGE COMMUNITY COLLEGE

RED ROCKS COMMUNITY COLLEGE

USGS science for a changing world