

Responsibly Designing Open Science Data Teams

Amanda Casari
Google Cloud
Kirkland, WA
amcasari@google.com

Abstract—Much of the implementation of reproducibility, transparency, resiliency, and scalability in data-centric projects tends to focus on technology. However, examining the fundamental challenges inherent in both “open science” and “big data”, it is the decisions impacting the people and team structure that are equally, if not more important, than decisions surrounding the technology stack. Combining the lessons we have learned in industry data science, the basic scientific process, and organizational theory, we can begin to categorize the kinds of problems examined by researchers and the tradeoffs that must be evaluated to make a team most effective. We can understand how to modernize and scale your team, while you are transforming your computational stack to meet the challenges of open science with big data.

Keywords—*open science, data science, big data, organizational theory*

ACKNOWLEDGMENT

Thanks to Carol Willing for recommending this opportunity to me

REFERENCES

[1] TBD