Provenance of a Tweet

DAN KERCHNER, JUSTIN LITTMAN, CHRISTIE PETERSON, VAKIL SMALLEN, RACHEL TRENT, and LAURA WRUBEL

Abstract: As part of our NHPRC-funded work to develop Social Feed Manager (h ttp://go.gwu.edu/sfm), an easy-to-use tool for gathering social media data via platform APIs, a team of developers and archivists at George Washington University have been wrestling with questions about what kind and how much metadata the system should capture to enable future use and reuse of the data (see https://scholarspace.library.gwu.edu/files/h128nd689). From social media researchers we have drawn the idea of research documentation to support reproducibility and comparability of shared research data, which is closely aligned with the archival concept of provenance, which the W3C's PROV working group defined as "information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness." In the specific instance of a tweet, we have identified dozens of categories of potential metadata related to production, selection and capture in a dataset that we could support and provide access to through SFM. We are actively seeking input from researchers, archivists, data librarians and others about which of these elements are worth creating or capturing with SFM collections of social media data.

About the Authors:

Dan Kerchner is a software developer and librarian with George Washington University Libraries.

Justin Littman is a software developer / librarian with George Washington University Libraries.

Christie Peterson is the university archivist at George Washington University. She graduated from the University of Michigan's School of Information, and has previously worked in archives at Johns Hopkins University, Princeton University and Bates College.

Vakil Smallen is the National Education Association Project Archivist at George Washington University. He graduated from the University of Maryland iSchool. Before coming to GWU, he worked in the Archives of the International Monetary Fund.

Rachel Trent is Digital Services Manager at George Washington University Libraries' Special Collections Research Center. She has worked in the area of social media archiving previously at the State Library of North Carolina and at the State Archives of North Carolina.

Laura Wrubel is a software development librarian at George Washington University and current PI for the grant "Building and Sustaining Social Media Collections: Tools and Policy Recommendations for Collection Development and Management," from the NHPRC. She has been helping faculty and students at GW collect and work with social media data since 2012.