To an ever greater degree, librarians “Data” is a nebulous concept. Through the lens of a data management plan, we determined that A preservation program that seeks to Recognizing that data are a product What counts as data and furthermore what counts as a dataset is Data are often a form of scholarship in A focus on “data curation” risks the To determine how to actually build the tool, we conducted 20 in 2012 EcoMob tracks Tweets about the While interest in managing data has Our library collection.” From the National Science The UCLA Registry is a new online tool designed to assist researchers Re • EcoMob tracks Tweets about the weather – can we learn about climate change by looking at what people say on Twitter? • 2012 – Consulted with EcoMob to develop a strategy for the preservation of their data. • Through the lens of a data management plan, we determined that the data for the project were the Tweets, stored in a MySQL database, along with the text files containing the scripts. This was what could be deposited in a repository. • But what about the dynamic visualizations? Aren’t these important products of the research? • Re-creation of the visualization from the scripts requires that a future user has the necessary software and know-how. • Our library-centric preservation plan was unable to capture the true intellectual contribution of the project.

Case Study 2: UCLA Data Registry

The UCLA Registry is a new online tool designed to assist researchers in managing and discovering surrogate records of datasets. To determine how to actually build the tool, we conducted 20 in-depth interviews with researchers from disparate fields—from neurology to archaeology, and everything in between. In some disciplines, such as the Digital Humanities, we found that research often involves a multitude of material—such as archival documents, computer code, and visualization software. What counts as data and furthermore what counts as a dataset is often very difficult to define. For the purposes of the Data Registry, we left these definitions up to the researcher to decide, but it left us convinced that “data” was not enough.

References


Acknowledgements: We would like to thank the members of the UCLA Information Studies department, where the research for this poster was conducted. Special thanks to the members of the EcoMob team and Dr. Jean-Francois Blanchette for their help and guidance with Case Study 1, and to Dr. Christine Borgman, Dr. Chris Kelly, Dr. Jonathan Furner, and Todd Grappone, for their help with Case Study 2.