

# Curation of ongoing and long-running projects

**Greg Janée & Renata Curty**



UCSB Library - Research Data Services  
University of California, Santa Barbara



# Common research data lifecycle

- Curation workflows are often built on a “default” research data lifecycle

Researcher...

- starts project  DMP consultation!
  - creates/gathers data
  - submits the data  Curation!
- Data is published and researcher is done with the data

# But not all workflows look like this...

- Data collected over the course of a researcher's career
- Research labs and ongoing projects
- Long-running field monitoring sites



# Common characteristics

- There is no point at which data is “done”
- Data is incrementally built up/shared/published (evolves, constant flux)
- Data = collections of heterogeneous but related objects
- Complex relationship between data and publications
- Often highly collaborative with different people acting at various steps
- Uses various tools often lacking in data repositories
- Multiple types of licenses and restrictions may apply
- Not necessarily grant funded
- Long-lived, on the order of years/decades



# Problem

Even today, they remain almost completely uncurated...

# Lingering Questions

When and how should we offer curatorial support to such projects? With what infrastructure? How scalable?

# Where are we at?

Evaluating solutions with the following features/requisites:

- Self-deposit
- Researchers to manage/partially curate their own data
- Whole lifecycle
- Computationally integrated
- Space-making and collaborative
- Pre-validation (manual and automated) and curation workflow
- Incremental publication and versioning
- Flexible licensing and access control

# Moving Forward

- Interviews with PIs and research teams
  - Data management practices
  - Curatorial expectations
  - Tools and workflows
  
- DCN Interest Group

