

Building Infrastructure for Scientific Data: Contrasting Curation Approaches across the Lifecycle

Dharma Akmon, SEAD
Kevin Dyke, Terra Populus

IDCC15 BoF Session, London, UK
February 9, 2015



Framing Questions for the Session

How can we more effectively engage scientists in data curation?

How can we facilitate dataset interoperability?

How do we most effectively integrate early and later lifecycle data curation activities?

Two Example Projects: SEAD and TerraPop



facilitates curation of “active” data and transition to long-term repositories
<http://sead-data.net/>



TERRA POPULUS

facilitates data integration, processes completed data to conform to standards and augments metadata
<http://www.terrapop.org/>

Active and Social Curation with SEAD

Active curation: recording data and metadata as close to the source as practical

Social curation: leverage cross-group interactions to motivate best practices

SEAD Project Spaces

Space that encourages scientists to upload data while working on them.

Metadata made immediately useful

Metadata addition is flexible and incremental

Geared toward self-interest rather than altruism

TerraPopulus:

Microdata, aggregate data, raster data

Making data interoperable while not deviating from the source

Dynamically derived datasets via extract system

Metadata standardization/augmentation/creation

Some source datasets lacking clear lineage

Questions for the Group

How can we more effectively engage scientists in data curation?

How can we facilitate dataset interoperability?

How do we most effectively integrate early and later lifecycle data curation activities?

When it comes to documentation and metadata, how can we ask for enough without asking too much?

Other topics?