

Ethical Data Design

Overview

- Investigate ethical tensions in gathering, assessing and sharing natural disaster data
- Analyze Harvey damage datasets
- Kubernetes case study – analyze commits across security issues
- Semi-structured interviews with data creators, repository managers, system administrators

Some Highlights

- Organizational values influence incentives to produce & share data
- Platforms embed constrained practices
- “Open” does not always mean complete and consistent

Meet the Team

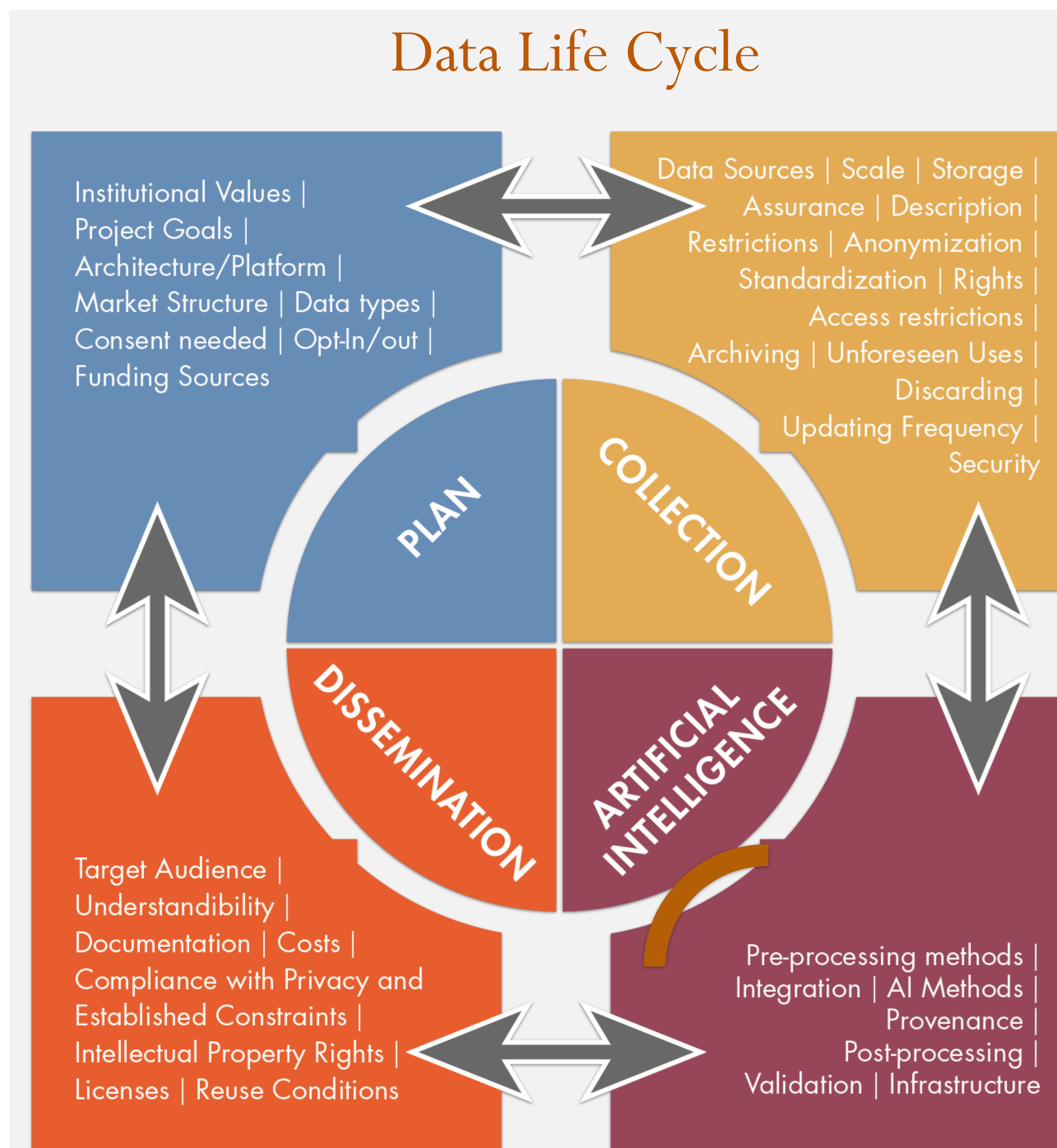
Maria Esteva
Data Analyst, TACC

Sharon Stroker
Professor, Journalism

Chris Rossbach
Professor, Computer Science

Soyoung Park
Doctoral Student, RA,
Radio-TV-Film

John Thywissen
Doctoral student,
Computer Science



Key Questions

- (1) What are key ethical data features from the perspectives of data producers, repositories, platforms and users?
- (2) How can ethical features, including the stakeholders' and data subjects' best interests, be effectively managed across the data lifecycle? What are the difficulties?
- (3) What techniques can help identify and track ethical features in data?

ISSUES

Data quality varies;
compartmentalized tasks

Ethical concerns align with community affected and/or providing the data:
no universals

Open access technologies and values may not be well understood by data creators

Data creators are not thinking about future data uses, including AI/ML.

Risks of building weak or biased AI models

Risks of distributing data to decision makers - quick data gathering, uncertain quality, constraints on analysis.

Low incentives for data curation/publication in the academy.

AI possibilities not well considered