

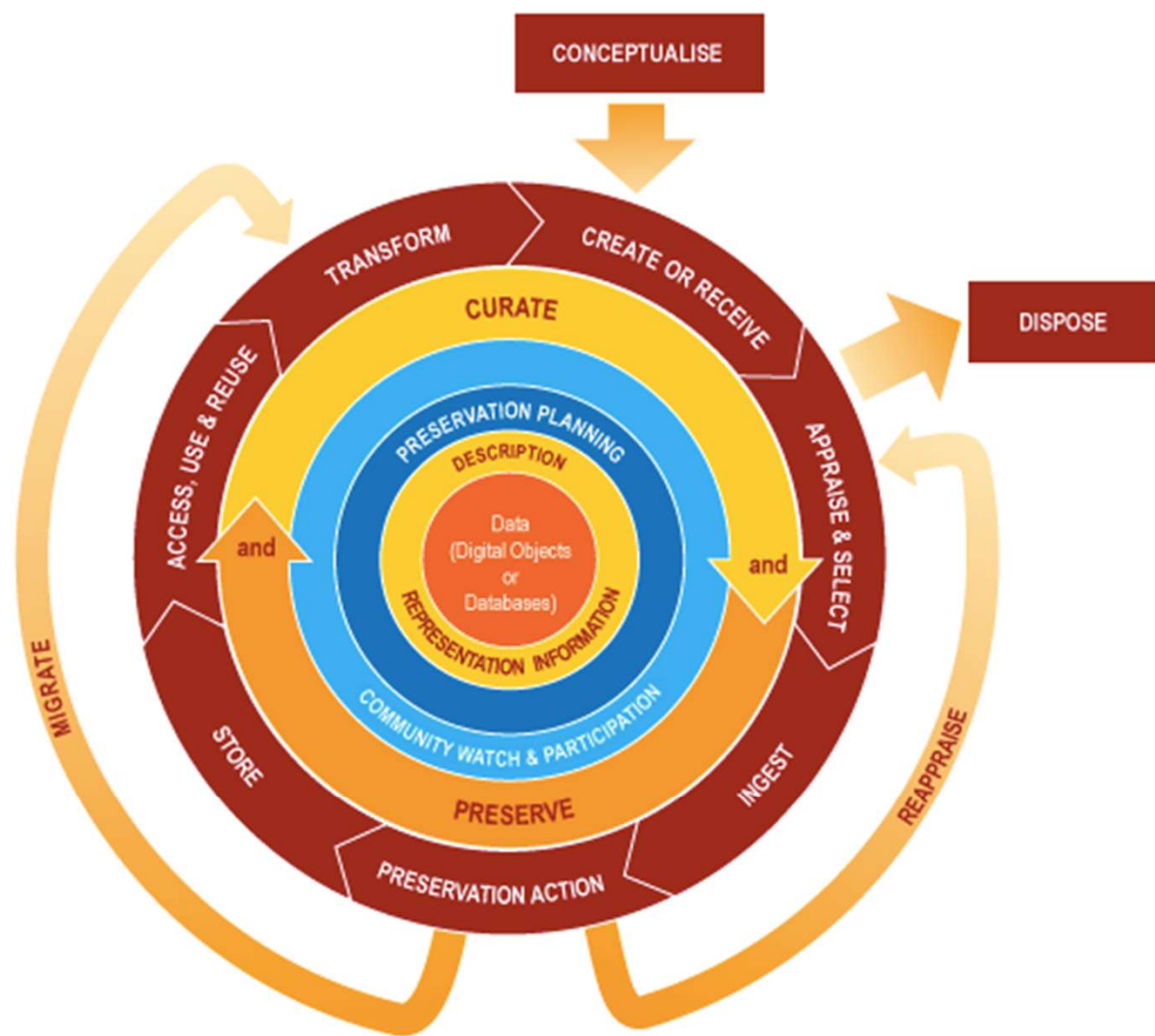
DCC Lifecycle Curation Model 2.0

Sayeed Choudhury, Johns Hopkins University

Carole Palmer & Rainbow Huang, University of Washington

2019 IDCC

February 5, 2019



Rationale for Research

- DCC Lifecycle Model has play an **important role** in digital curation from both theoretical and practical perspective
- **Context and environment** – particularly related to data – have **changed notably** since the inception of the model

Alfred P. Sloan Foundation Grant

- Consideration of **data** in terms of scope, complexity, and machine actionable nature (especially as it relates to connections between data archives and **high performance computing**)
- Assessing impact of **machine learning** for **data processing** and analytics
- Casting the Lifecycle model in terms of the growing movement of **integrated research workflows**
- Accounting for **algorithmic bias**, particularly as it relates to issues of **diversity and inclusion**

Goal and Methodology

- **Recommendations** for next steps – **not a new version** of the model (if that's even necessary)
- **Review** of lifecycle models – both **content and research**
- Interviews with **DCC staff** and analysis of documentation including original (Sarah Higgins) paper and subsequent **DCC reports**
- **I am not representing the DCC** – findings and recommendations from my, Carole Palmer's and Rainbow Huang's (both University of Washington) **synthesis**

Original DCC Lifecycle Model

- Motivated by desire to organize **DCC website** materials (and associated requests)
- Based on **archival science principles** – possible connections to the OAIS reference model
- Adapted to **multiple purposes** – some of which have been unintended or unplanned
- **Pedagogical** applications were perhaps most **(pleasantly) surprising**
- **Guide** for both institutions and researchers for **conversation**

Recommendations

- Adopt a more **process oriented approach** accounting for workflows, sheer curation, and agile software development practices
- **Modular approach** – focus on **most relevant components** for context or domain rather than entire lifecycle
- **Decision tree approach** – extend lifecycle with framework for researchers and curators to make **specific interventions**
- While there are some **initial resources** (e.g., values levers, data sheets) for considering **algorithmic bias**, there remains **much to explore**
- Incorporate **data science principles**...

Acknowledgements

- Kevin Ashley and colleagues at DCC
- Carole Palmer, Rainbow Huang and colleagues at University of Washington
- Sheridan Libraries at Johns Hopkins University
- Alfred P. Sloan Foundation