

# Diversity in Data: Lessons from Interdisciplinary Practices

Jonathan Petters (presenter), Virginia Tech

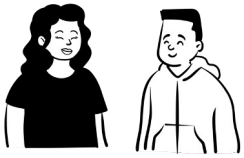
Inna Kouper, Indiana University

Thea Lindquist, University of Colorado Boulder

Ece Gumusel, Indiana University

# Background

- Interdisciplinary and highly collaborative research (IHCR) combines knowledge and expertise across domains, communities, and organizations
- Project is funded by IMLS ... a multi-method study that included interviews, observations, and document analysis
- Six interdisciplinary research projects across a wide range of sciences, computational sciences, social sciences, and humanities



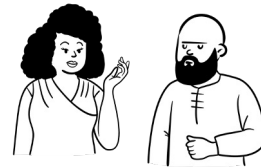
Midwest Mapping  
and Ecology



Sustainable  
Food Systems



Histories of Slave  
Trade



Specialty Crops  
Production



Urban Green  
Infrastructures



Trends in Rural Communities  
(Rural America)

# This Study

- Previous studies have primarily focused on team dynamics, rather than the technical and ethical challenges of working with complex, heterogeneous datasets.
- Our study fills this gap by providing a detailed exploration of interdisciplinary/highly collaborative data management practices and the challenges that arise.
- Q - How do interdisciplinary project members negotiate their work around diverse datasets?
- Q - What role do ethics play in their data practices?
- Q - How do interdisciplinary teams approach privacy and confidentiality as they collect, store, and organize their heterogeneous datasets?

# Diversity in Research Teams

- Defined in various ways (race, gender, nationality, career stage, experience, skill level)
- General awareness (DEI is important) and topical awareness (how to study DEI in the context of historical injustice or food security)
- Often deferred to leadership (responsibility of the PI)
  - Research quality vs diversity
- Perceptions differ across gender and career stages

# Diversity in Research Data

- Studying diversity across populations and samples (e.g., economic status, racial diversity in humans, genetic diversity in crops)
- Sources (temporal and spatial scales, frequent changes, access constraints, human participant pool)
- Technologies must support multiple data / metadata types
- Vocabularies – different ways objects and subjects of studies can be described
  - e.g. regions involved in the slave trade

# Diversity in Challenges within IHCR Projects

- Negotiating disciplinary background and experiences
- Finding consensus and/or reaching compromise
- Implementing a top-down vs collaborative model
- Using multiple tools and data sources
- Lack of training and support for handling IHCR data and metadata

# Lessons Learned

- Better teams are created with intention and awareness of interpersonal dynamics and diversity in data
  - “You should talk about things from the outset!”
- Importance of documentation and transparency in data management
- Need more ethics education related to how data work affects communities (crops, food, historical injustice)
- Emphasize actions to make data reusable during its active management

# Conclusion

- Effective interdisciplinary team dataset practices, ethics, privacy, and confidentiality in data, and diversity and heterogeneity in data are crucial for successful research projects.
- Researchers must prioritize open communication and collaboration within their teams, ensure proper data management and organization, address ethical and privacy concerns, and consider the diversity and heterogeneity of the data being collected and analyzed.
- **These statements are relevant for effective research in general, but even moreso for IHCR**



Project outputs available on  
[Github](#)

Thank you!

[jpetters@vt.edu](mailto:jpetters@vt.edu)  
@jon\_petters

IASSIST 2023  
Diversity in Data: Lessons from Interdisciplinary Practices