

# MAPPING THE HUMANITIES DATA ECOSYSTEM

- + Nathan D. Woods
- Postdoctoral Research Fellow
- Research Talks and Coffee Breaks, University of Lethbridge, July 09, 2024



Humanities  
Innovation  
Lab



University of  
Lethbridge

With Funding from SSHRC-CRSH



Conseil de recherches en  
sciences humaines du Canada

Social Sciences and Humanities  
Research Council of Canada

Canada



---

# **I. BACKGROUND AND THE NATURE OF THE PROBLEM**

- + • **II. THE HUMANITIES DATA ECOSYSTEM**

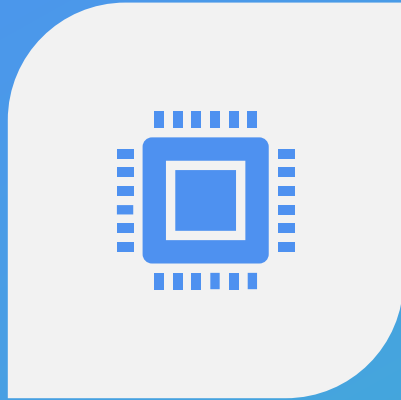
- **III. PRELIMINARY FINDINGS AND TAKEAWAYS**



# SCALE OF THE DATA REVOLUTION

- **2003:** “Between the dawn of civilization and 2003, we only created five exabytes of information; now we’re creating that amount every two days” (quoted in Einav and Levin 2013).
- **2023:** The world produced and consumed approximately 120 zettabytes of data (IDC and Statista 2024).

# DIGITAL TRANSFORMATION



WIDESPREAD  
INCORPORATION OF  
INFORMATION AND  
COMMUNICATION  
TECHNOLOGIES (ICTS).



CHANGES TO THE  
POWER AND MAGNITUDE  
OF COMPUTING  
CAPABILITIES.



SOCIAL AND CULTURAL  
CHANGES ALLOWING  
SUPPORTING NEW TYPES  
OF BEHAVIOR.

# DRIVEN BY THE MULTIPLICATION OF DATA ASSEMBLAGES



BIG DATA



MASS  
DIGITISATION



SCALING  
SMALL DATA  
SETS



DATA  
BROKERAGE



OPEN DATA  
STANDARDS  
AND  
PRACTICES



NEW DATA  
INSTITUTIONS

# DATA INTENSIVE SCHOLARSHIP



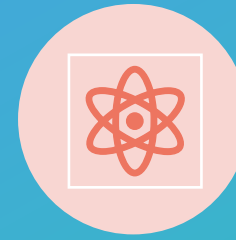
DATA SCIENCE



DIGITAL  
HUMANITIES



HUMANITIES  
ANALYTICS



COMPUTATIONAL  
SOCIAL SCIENCE



SOCIAL  
COMPUTING

# DATA AND SCHOLARLY INFORMATION OUTPUT

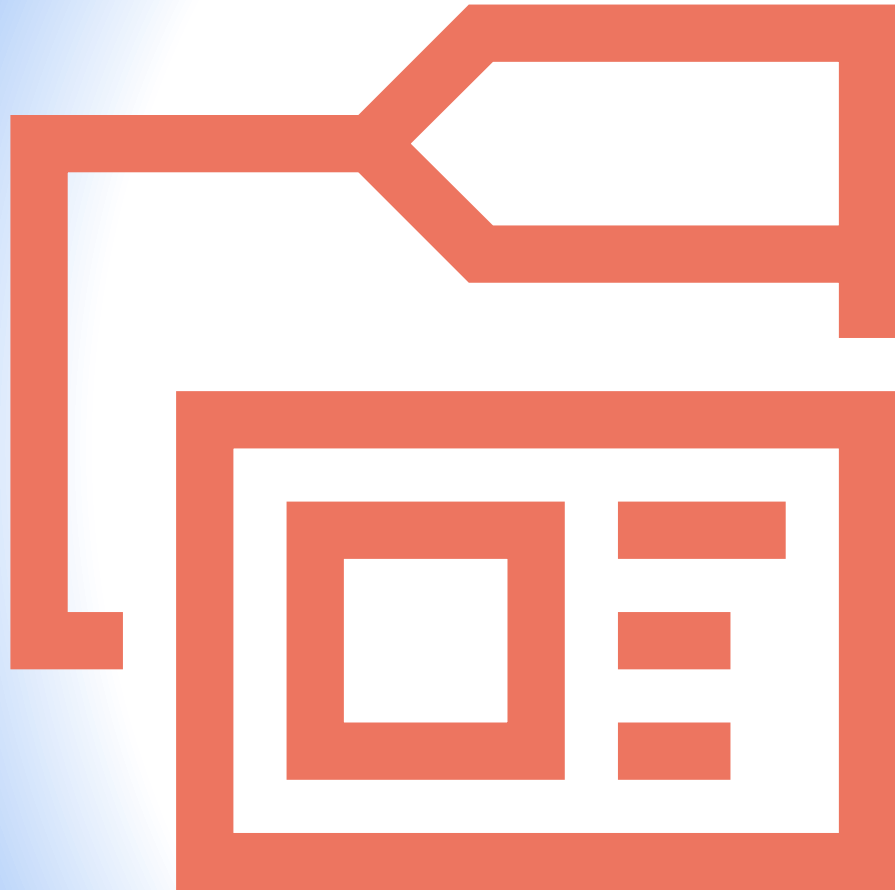
4 MILLION RESEARCH PAPERS PRODUCED LAST YEAR in 2023,  
ESTIMATED

CHALLENGES:

MAKING PAPERS AVAILABLE AND DISCOVERABLE

BUT WHAT IF THESE PAPERS ARE EACH DATA POINTS IN A  
LINKED DATA SET?

# **SCHOLARLY RECORD:**



**RECORD OF  
RESULTS OF  
ORGANIZED  
INQUIRY**



**SCHOLARSHIP**

**SCIENTIFIC RECORD**

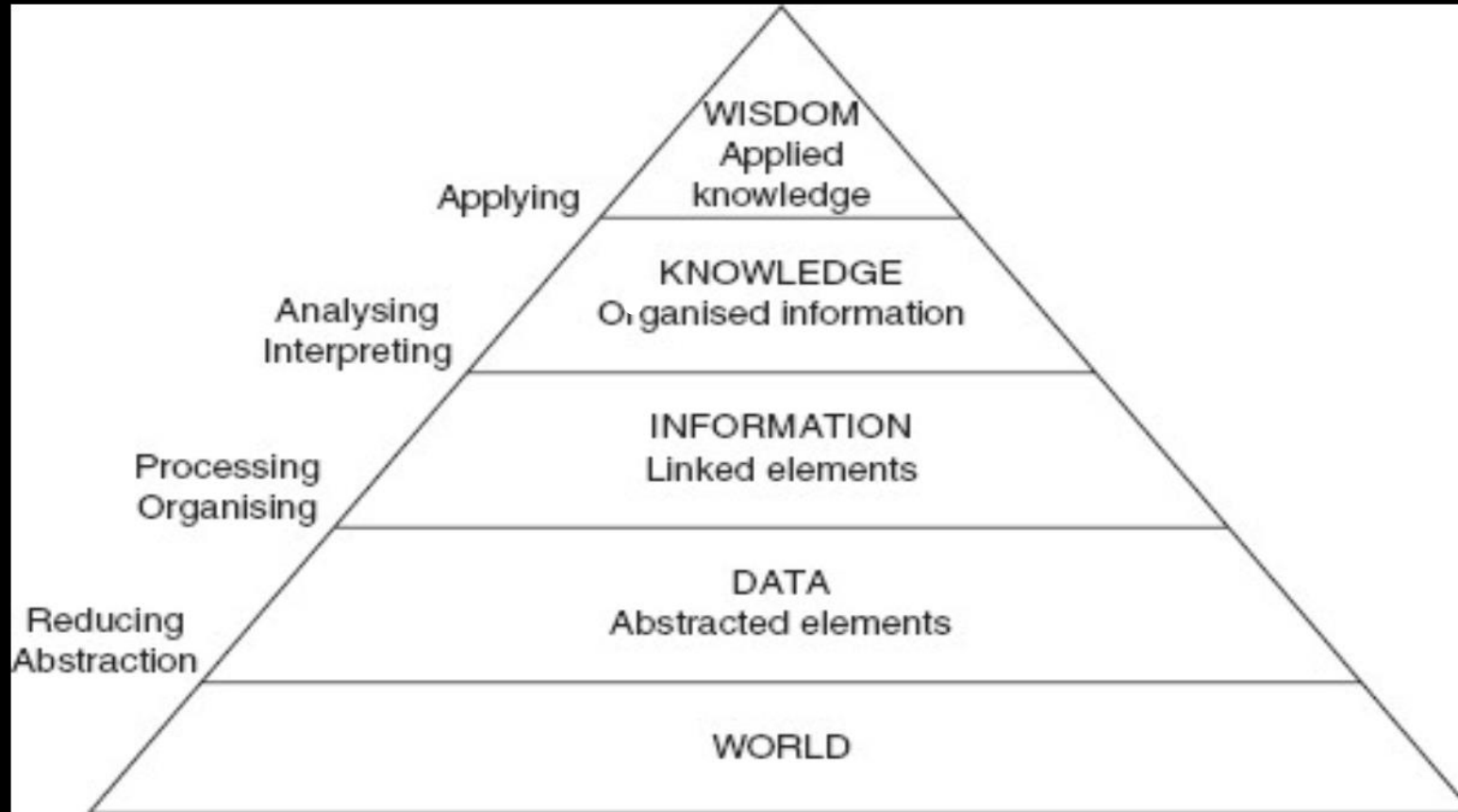
**HUMAN CULTURAL RECORD**

**HISTORICAL RECORD**

**ETHNOGRAPHIC RECORD**

**EVIDENCE BASE**

# CLASSIC KNOWLEDGE HIERARCHY



**Figure 1.1** Knowledge pyramid (adapted from Adler 1986 and McCandless 2010)

## A collection of colorful stars and hearts scattered across the page. The shapes include red, orange, blue, green, yellow, purple, brown, and grey stars, as well as red, green, blue, yellow, pink, brown, and grey hearts. Some shapes are solid, while others are outlined. They are arranged in a random pattern across the white background.

	★	▲	♥
001	★	▲	♥
002	★	▲	♥
003	★	▲	♥
004	★	▲	♥
005	★	▲	♥
006	★	▲	♥
007	★	▲	♥

FORM ABC-12

★ ★ ♥

— — — — —

★

— — — — —

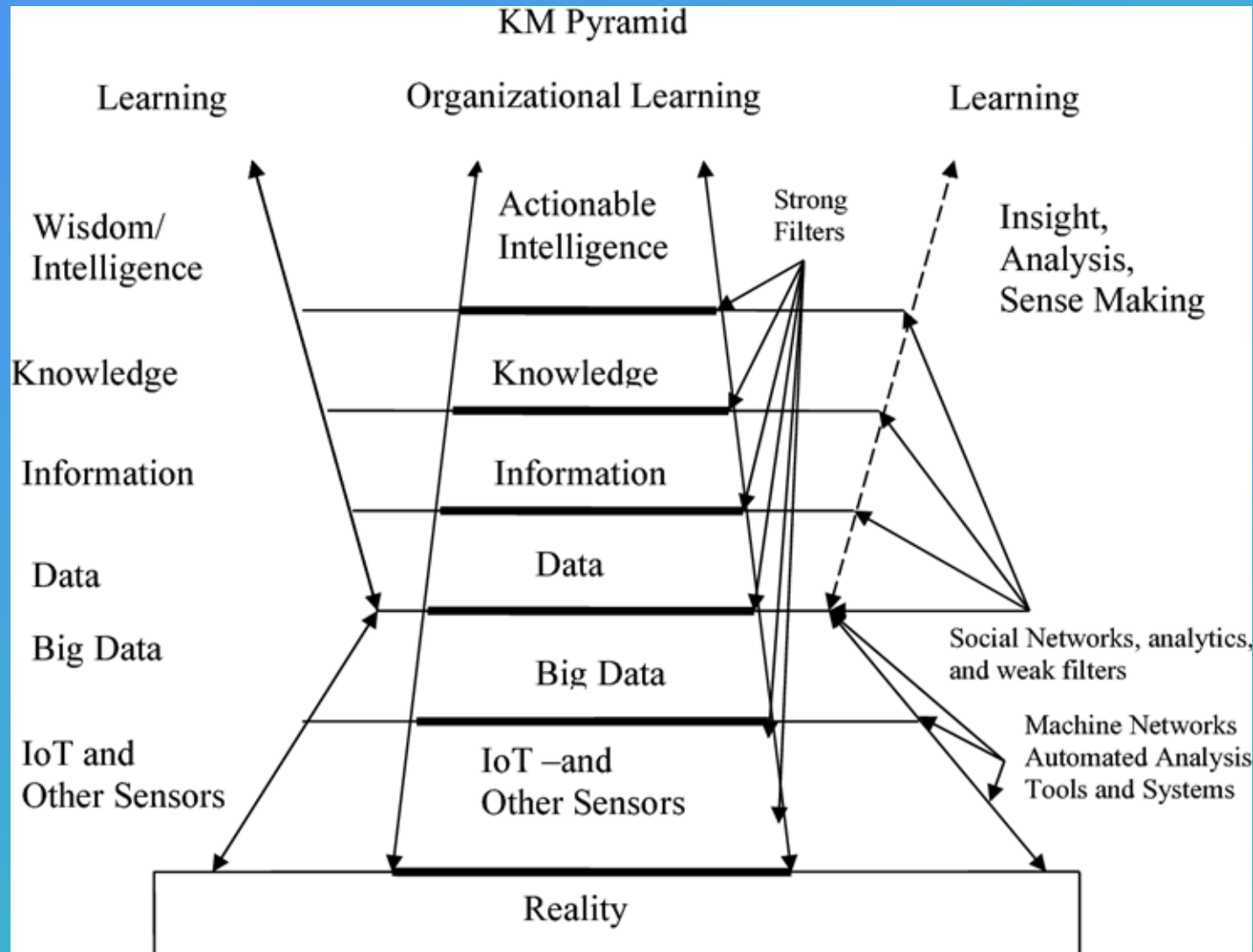
★ ★

— — — — —

★ ~~~~~

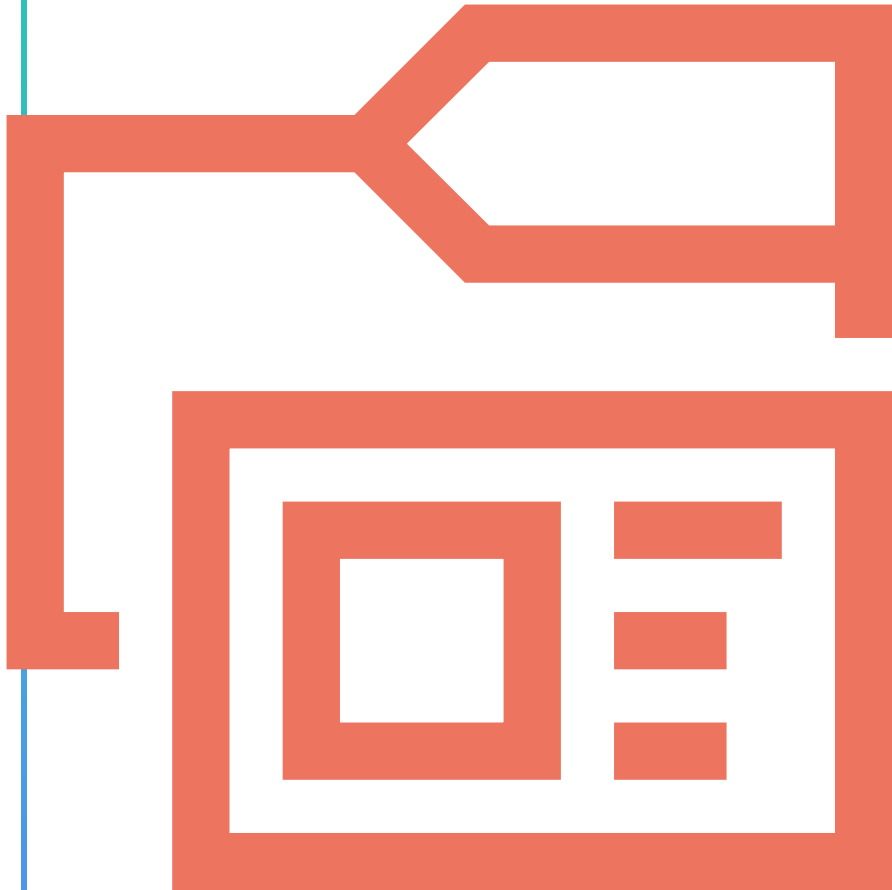
Texas State Library and Archives Commission. “The Relationships Between Data, Information, and Records.” The Texas Record. May 18, 2022.

# KNOWLEDGE HIERARCHY: BIG DATA



Adapted from Kitchin, *The Data Revolution* (2014)

# ECOSYSTEM MODEL OF COLLECTIONS



**SCHOLARLY CULTURES**

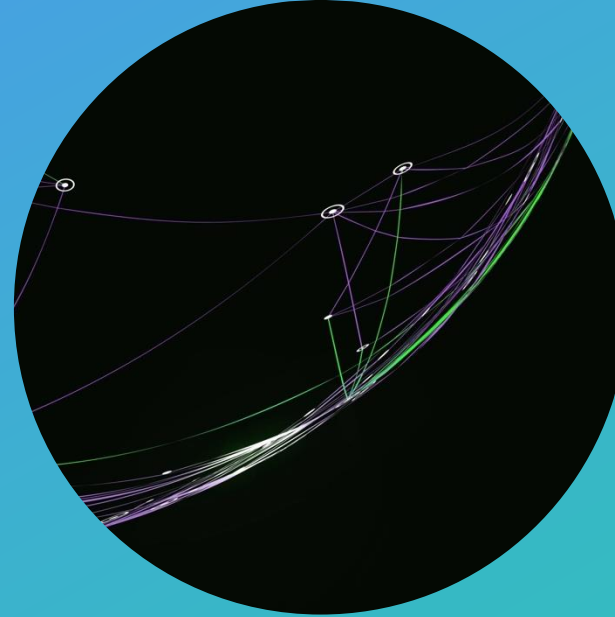
**DOMAINS OF EVIDENCE**

**SCHOLARLY PUBLICS**

**KNOWLEDGE INFRASTRUCTURE**

+  
○ ●

# HUMANITIES DATA INQUIRY



# HUMANITIES DATA INQUIRY

BIG DATA  $\neq$   
HUMANITIES  
PRACTICE

DATA = DIVERSE

DATA  $\nabla$  SOURCE

Humanities methods and assumptions are not reflected in emerging data practices.

Lack of shared understanding about data types and problems.

A terminological disjunct between a 'source' orientation and a 'data' orientation.

# HUMANITIES DATA INQUIRY

REFLECTION

COMMUNITY  
OF PRACTICE

DOCUMENTATION

PRACTICE  
PROBLEMS &  
USE CASES

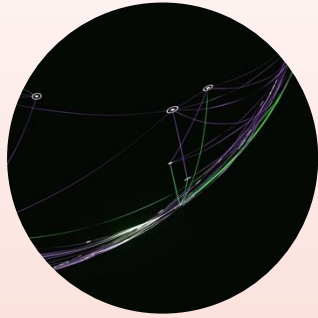
AGENDA SETTING

PROBLEM  
CLARIFICATION

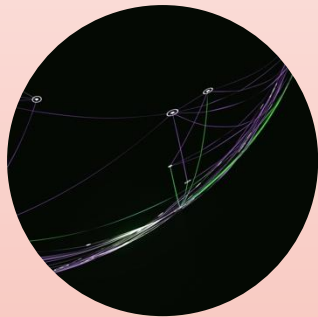
# THE PROBLEM WITH HUMANITIES DATA

An abstract network diagram is overlaid on the slide. It features a black background with a complex web of thin, curved lines in green and purple. Small white circles, some with black centers, are positioned at various points where the lines intersect or terminate, representing nodes in a network. The lines and nodes are more densely packed on the right side of the slide, creating a sense of depth and complexity.

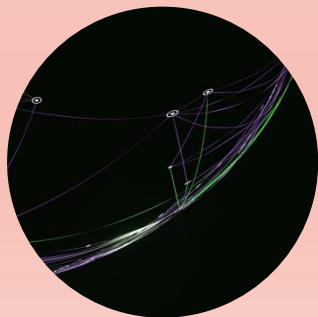
- **VARIABLE DEFINITIONS AND RESISTANCE TO DATA**
- **FEW EMPIRICAL STUDIES OF HUMANITIES IN PRACTICE**
- **LACK OF GENERALIZABLE STATEMENTS OF METHOD**
- **NO COHERENT ‘HUMANITIES POLICY’**



EMPIRICAL RESEARCH OBJECTS



INFORMATION--FACTS AND  
STATISTICS--COLLECTED FOR  
ANALYSIS



INFORMATION THAT CAN BE  
PROCESSED BY A COMPUTER

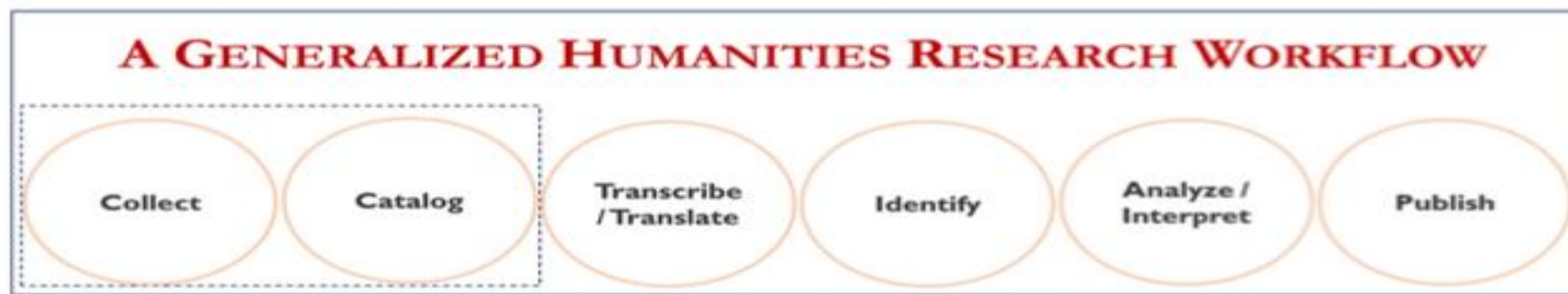
# THE DATA PROBLEM

---

**“HUMANISTS—WHO HAVE LARGELY EXISTED ON THE MARGINS OF MAJOR TRENDS TOWARDS MANDATORY DATA SHARING THAT ARE TRANSFORMING RESEARCH PRACTICES AND SCHOLARLY COMMUNICATION IN OTHER FIELDS—MUST NOW CONSIDER THEIR PLACE IN THIS POLICY LANDSCAPE.”**

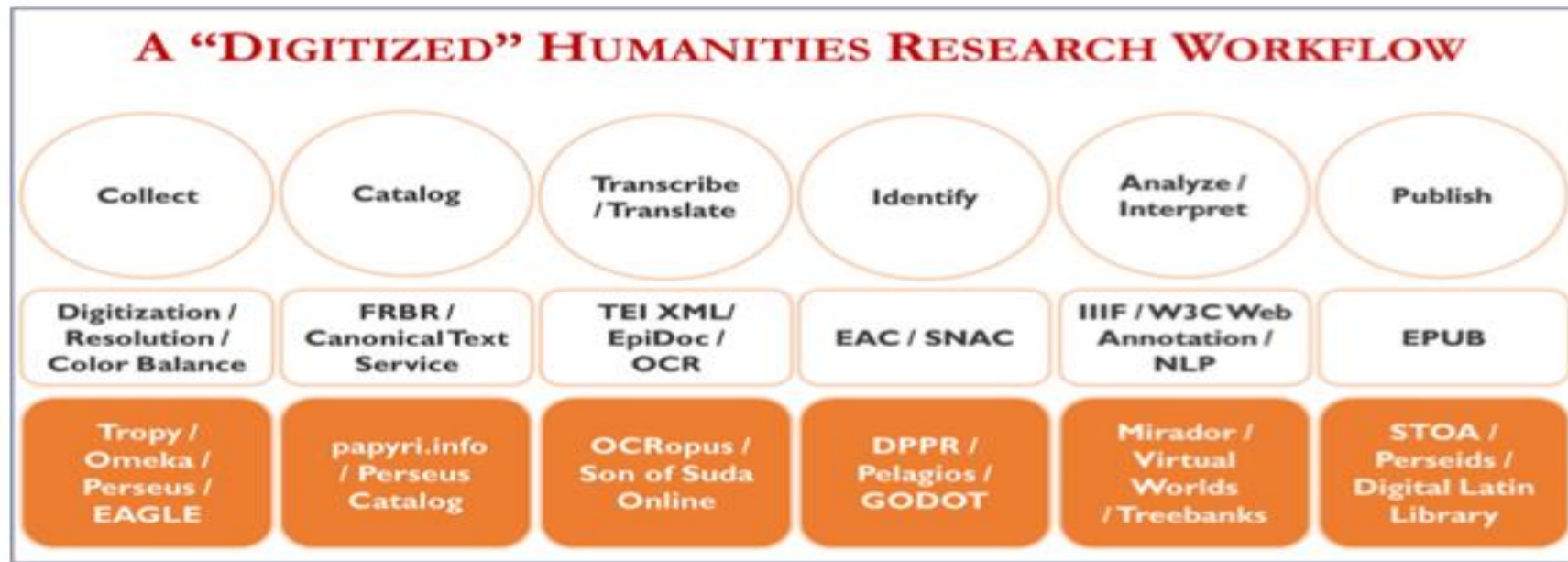
----Ruediger, Dylan and Ruby MacDougall. "Are the Humanities Ready for Data Sharing?"

# CONSOLIDATING DIGITAL WORKFLOW



**Fig. 1** "A generalized humanities research workflow." It contains six circles to represent the elements of the workflow. They are labeled from left to right: Collect, Catalog, Transcribe/Translate, Identify, Analyze/Interpret, and Publish. The Collect and Catalog circles are enclosed

in a dotted box, which is meant to depict the functions typically associated with libraries



# DIVERGENT RESEARCH INFRASTRUCTURES

---



EUROPE



CANADA



UNITED  
STATES



# COMPARATIVE RESEARCH INFRASTRUCTURES



# CONVERGING DATA PRESSURES

---



**DIGITIZATION**

**COLLECTIONS AS DATA**

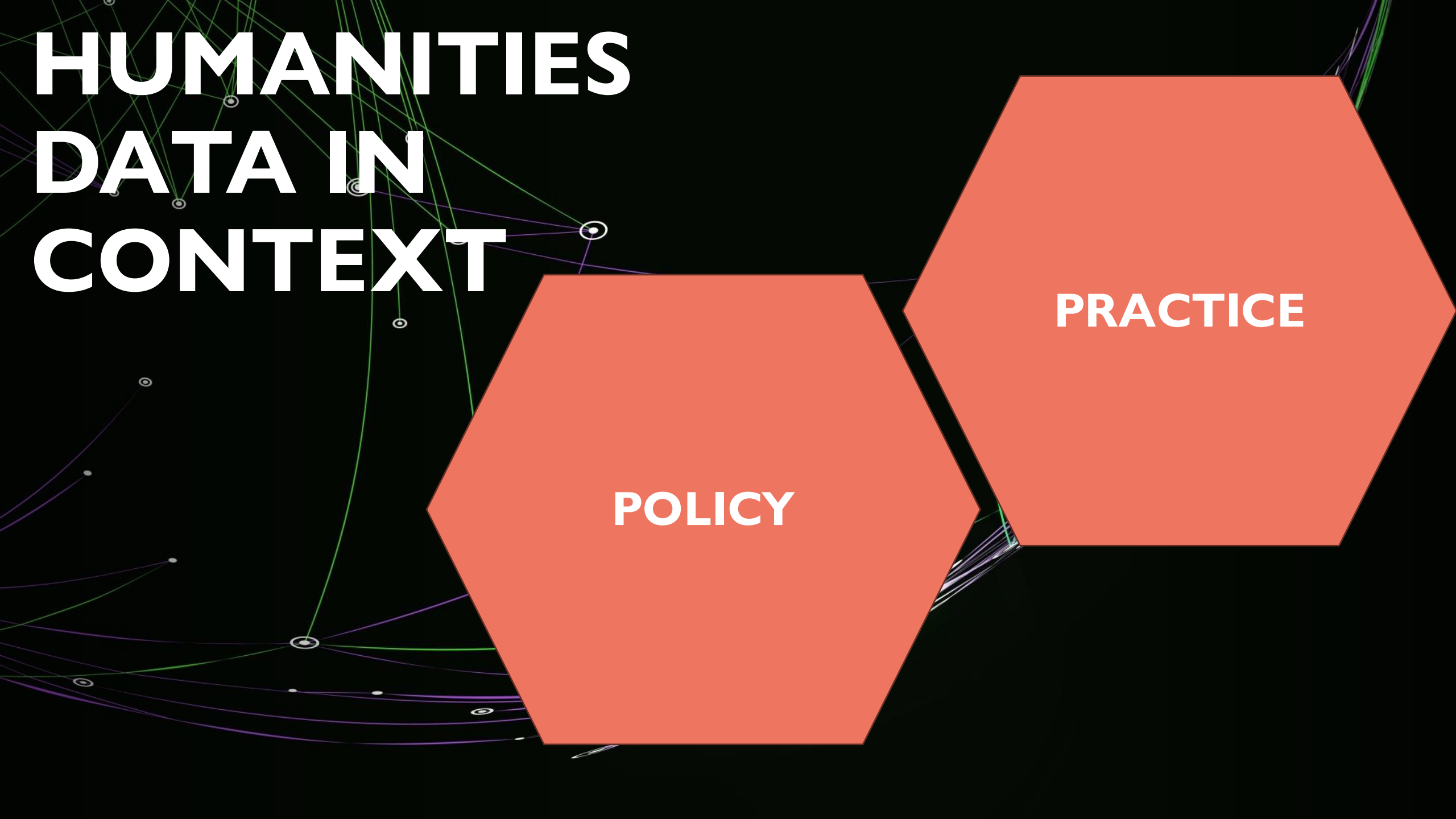


**HUMANITIES**

**DATA INTENSIVE  
SCHOLARSHIP**



**DIGITAL RESEARCH  
INFRASTRUCTURE**



# HUMANITIES DATA IN CONTEXT



# ETHNOGRAPHIC FRAME: PRACTICE<sup>+</sup> •<sub>o</sub>



**WHAT DO HUMANITIES RESEARCHERS DO?**



**WHAT DO HUMANITIES SCHOLARS DO  
WITH DATA?**

# ETHNOGRAPHIC FRAME: PRACTICE<sup>+</sup><sub>o</sub>



**WHAT SCHOLARS DO IN PRACTICE?**



**REFLECTIONS ON PRACTICE**



**INTERPRETATIONS OF PRACTICE**

# ETHNOGRAPHIC FRAME: POLICY



**HOW IS HUMANITIES PRACTICE DEPICTED?**



**HOW IS HUMANITIES DATA DEPICTED**



**INTERPETATIONS OF PRACTICE**

# MULTIMETHOD DATA COLLECTION

---



**MAPPING REVIEW OF THE LITERATURE**

**PARTICIPANT OBSERVATION**

**COMPARATIVE CASE STUDIES**

**SEMI-STRUCTURED INTERVIEWS  
(N=100)**

# PURPOSIVE SAMPLING

---



**HISTORY**

**PHILOSOPHY**

**LINGUISTICS,  
LITERATURES,  
LANGUAGES**

**ART HISTORY-  
ARCHAEOLOGY**

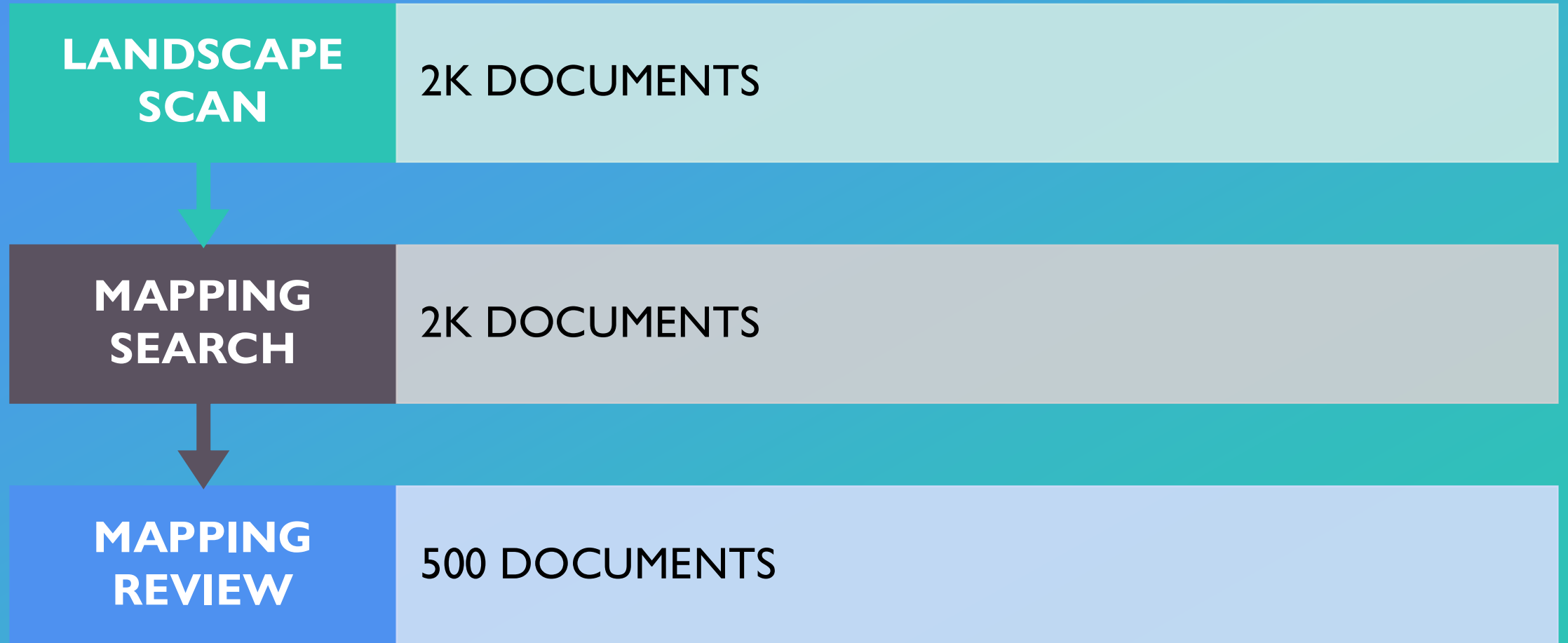
**RELIGIOUS  
STUDIES**

**MUSIC-  
MUSICOLOGY**

**MEDIA-MEDIA  
STUDIES**

**CULTURAL  
STUDIES**

# REVIEW OF LITERATURE



# MULTI-SITED DATA COLLECTION



## MAPPING SEARCH OF LITERATURE

- LITERATURE IN ENGLISH

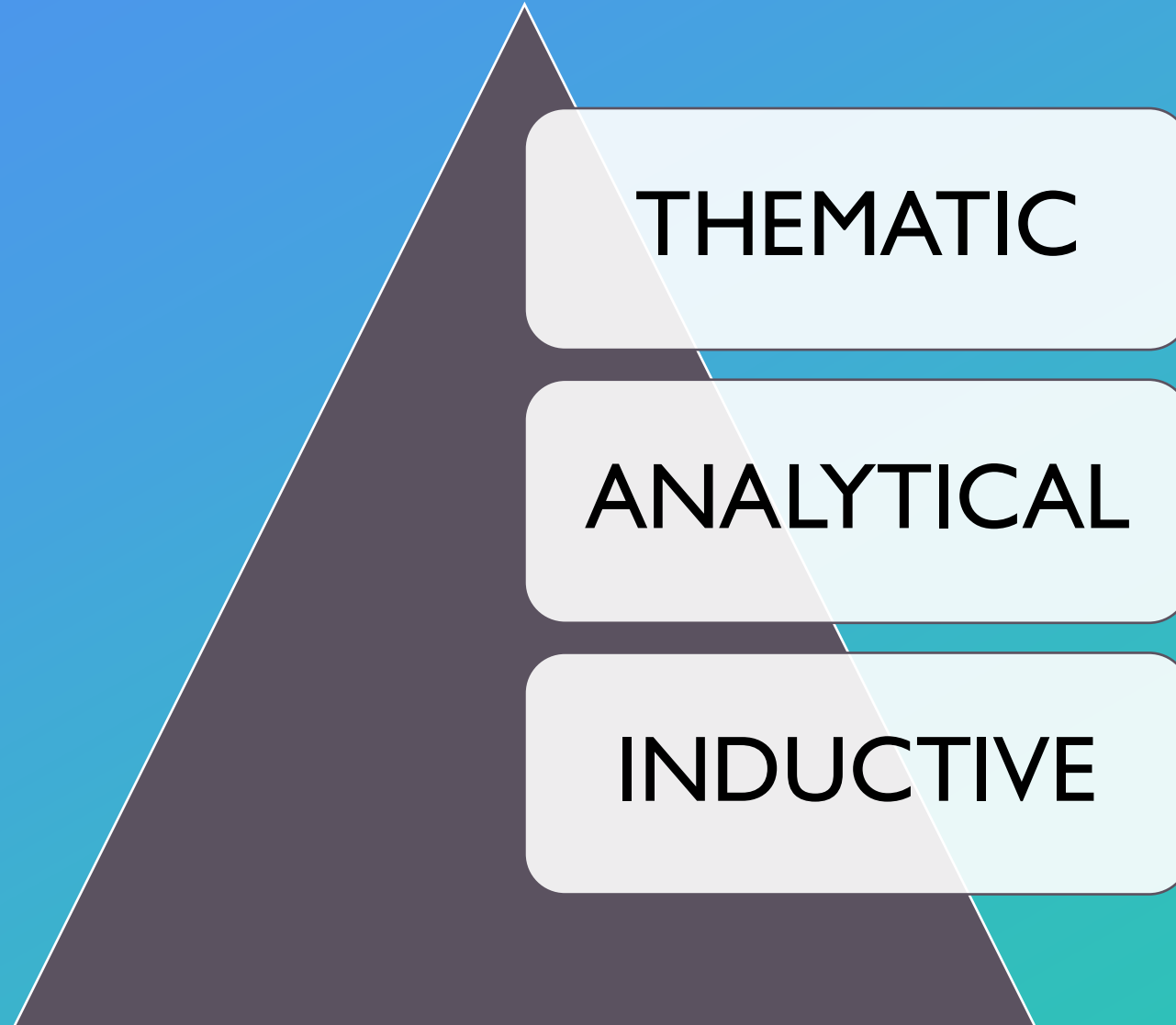
## VIRTUAL INTERVIEWS

- CANADA
- UNITED STATES

## IN-PERSON INTERVIEWS

- ALBERTA
- MONTANA

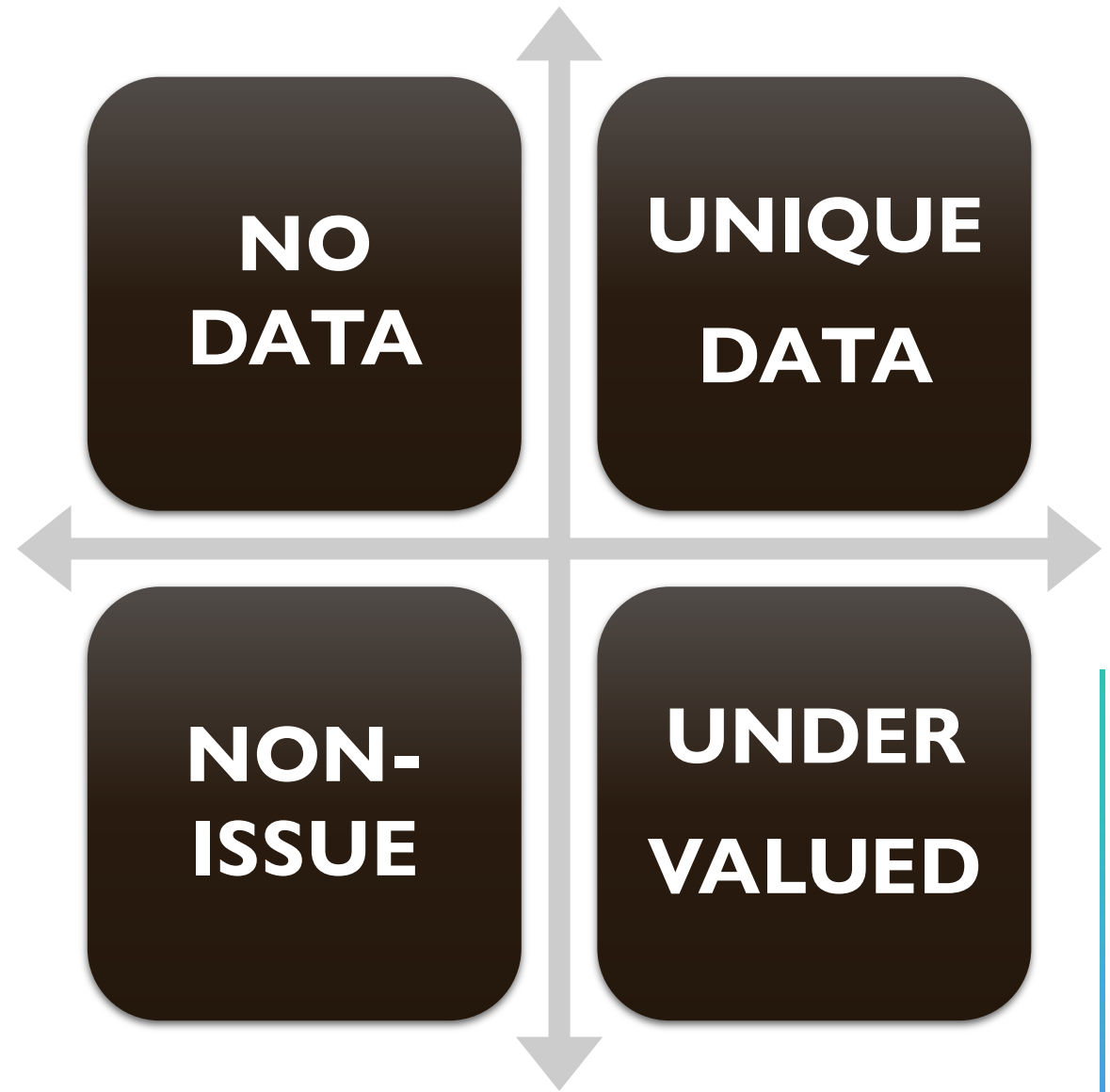
# CODING STRATEGY



# DATA AS A CULTURAL OBJECT

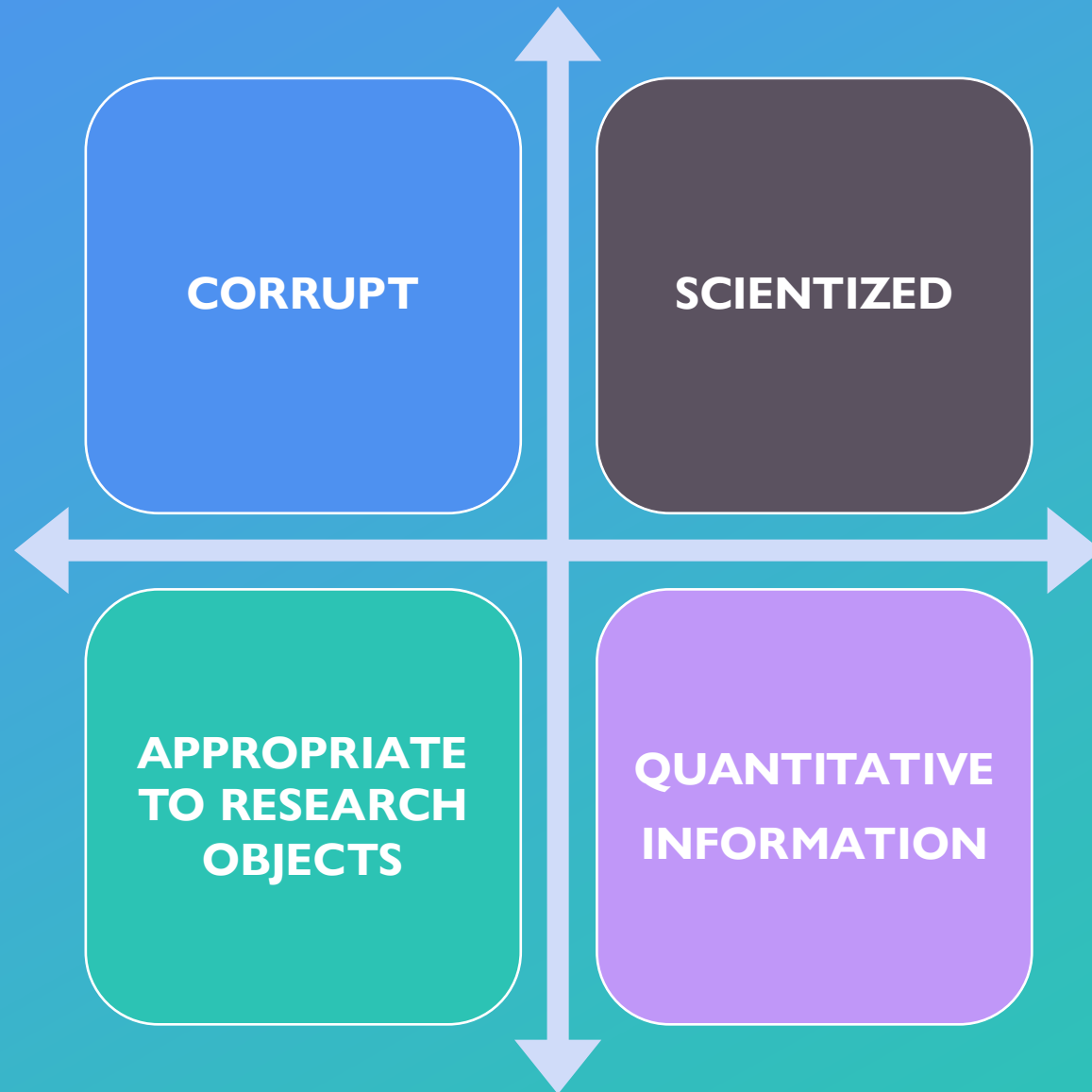
An abstract network diagram on a black background. It features numerous nodes, represented by small white circles with black outlines, connected by thin, curved lines. The lines are primarily green and purple, creating a complex web of connections. The network is more densely packed on the right side of the image, where many lines converge, and more sparse on the left. The overall effect is one of dynamic, interconnected data.

# HUMANITIES DATA FROM THE LITERATURE



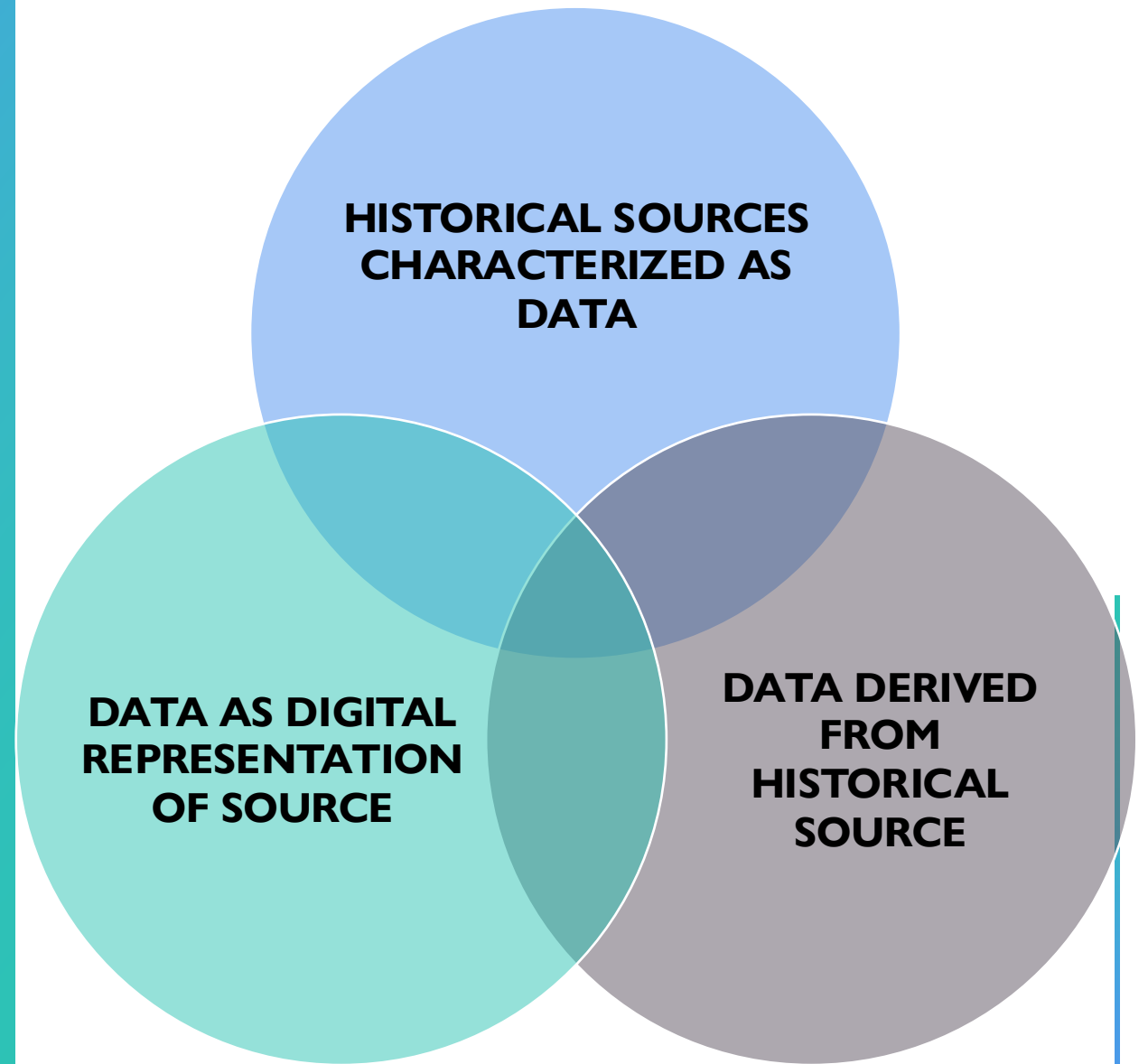
# HUMANITIES DATA AS A CONTENTIOUS OBJECTS

---



# ETHNOGRAPHIC SCENE:

## Humanities Data in Conversation



The background features a complex network of thin, glowing lines in green and purple. These lines connect various small, circular nodes, some of which are highlighted with a white ring. The lines and nodes are scattered across the black field, with a higher density of connections on the right side, creating a sense of dynamic movement and interconnectedness.


# PRELIMINARY FINDINGS



**EPISTEMIC  
CULTURES**



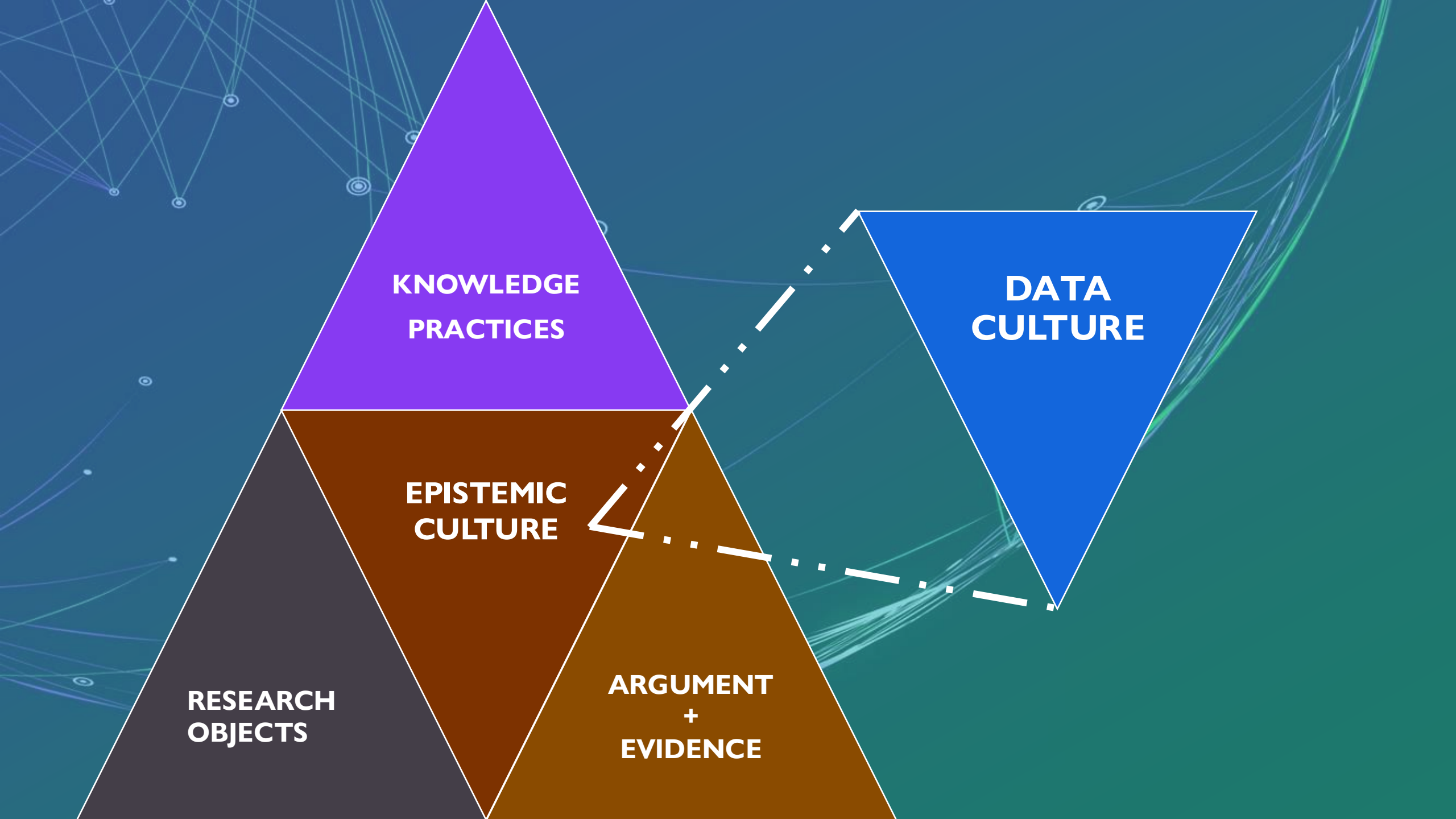
**CREATE  
KNOWLEDGE**



**WARRANT  
KNOWLEDGE**



**Research Objects  
Knowledge Practices**  
**Argument  
Evidence**



# DIVERGENT DATA CULTURES

---

+

•

○

Representational Data

Cultural Data

Distant Evidence

Humanities Analytics

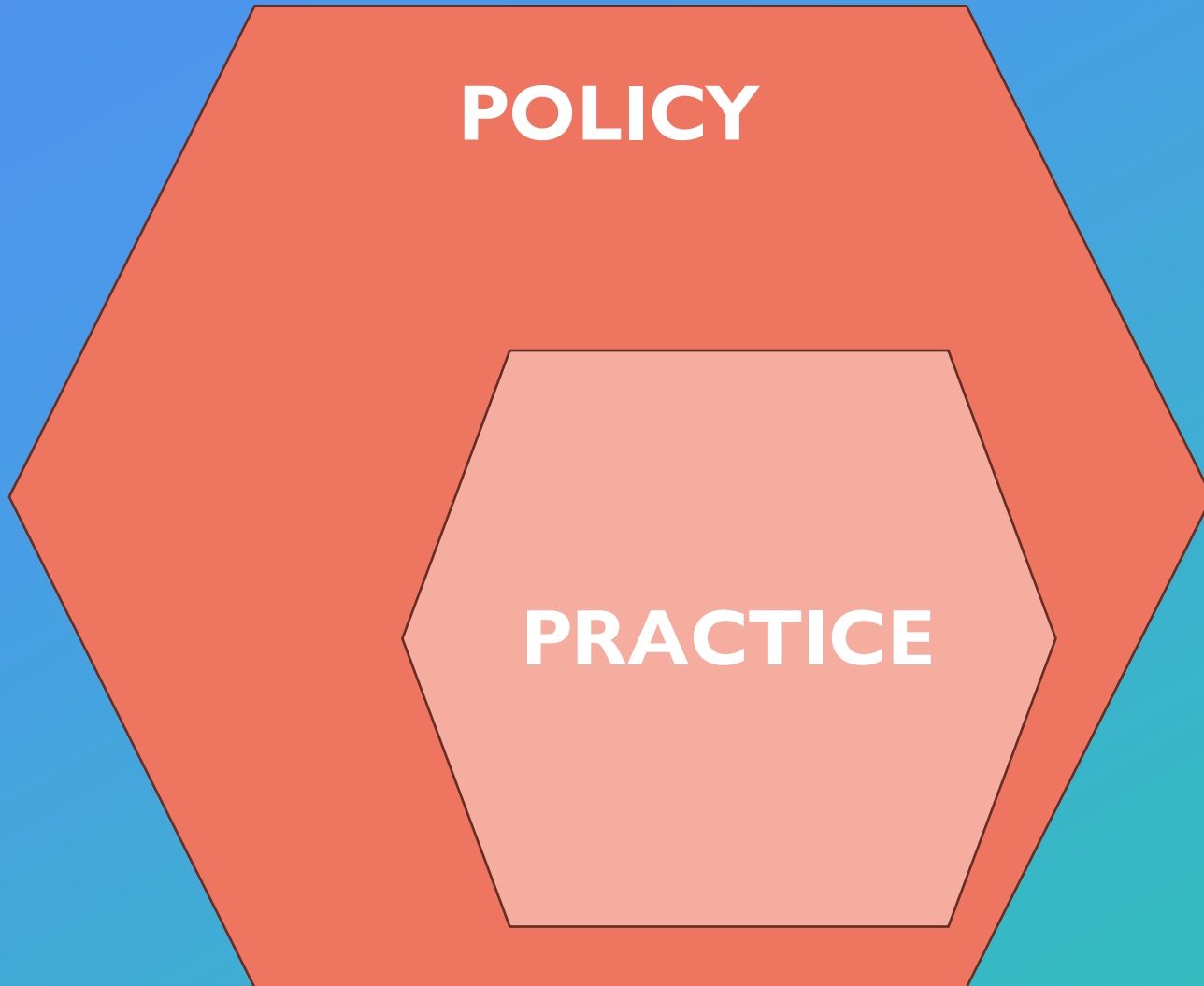
# COMPOUND PROBLEM

---

+

•

○



POLICY PROBLEMS ARE  
MISALIGNED WITH  
PROBLEMS IN PRACTICE



# NEXT STEPS

---

# EXTENDING THE ETHNOGRAPHIC FRAME



The diagram features a dark blue background with a network of white and light blue lines and nodes, suggesting a complex system or data network. On the left, a large orange hexagon contains the text 'INFRASTRUCTURE EXPERTISE'. To its right, three stacked orange rounded rectangles contain the text 'INFORMATION PROFESSIONALS', 'TECHNOLOGISTS', and 'FUNDERS AND POLICY PROFESSIONALS' respectively. The overall layout suggests a conceptual model for extending ethnographic research into technical and policy domains.

**INFRASTRUCTURE  
EXPERTISE**

**INFORMATION  
PROFESSIONALS**

**TECHNOLOGISTS**

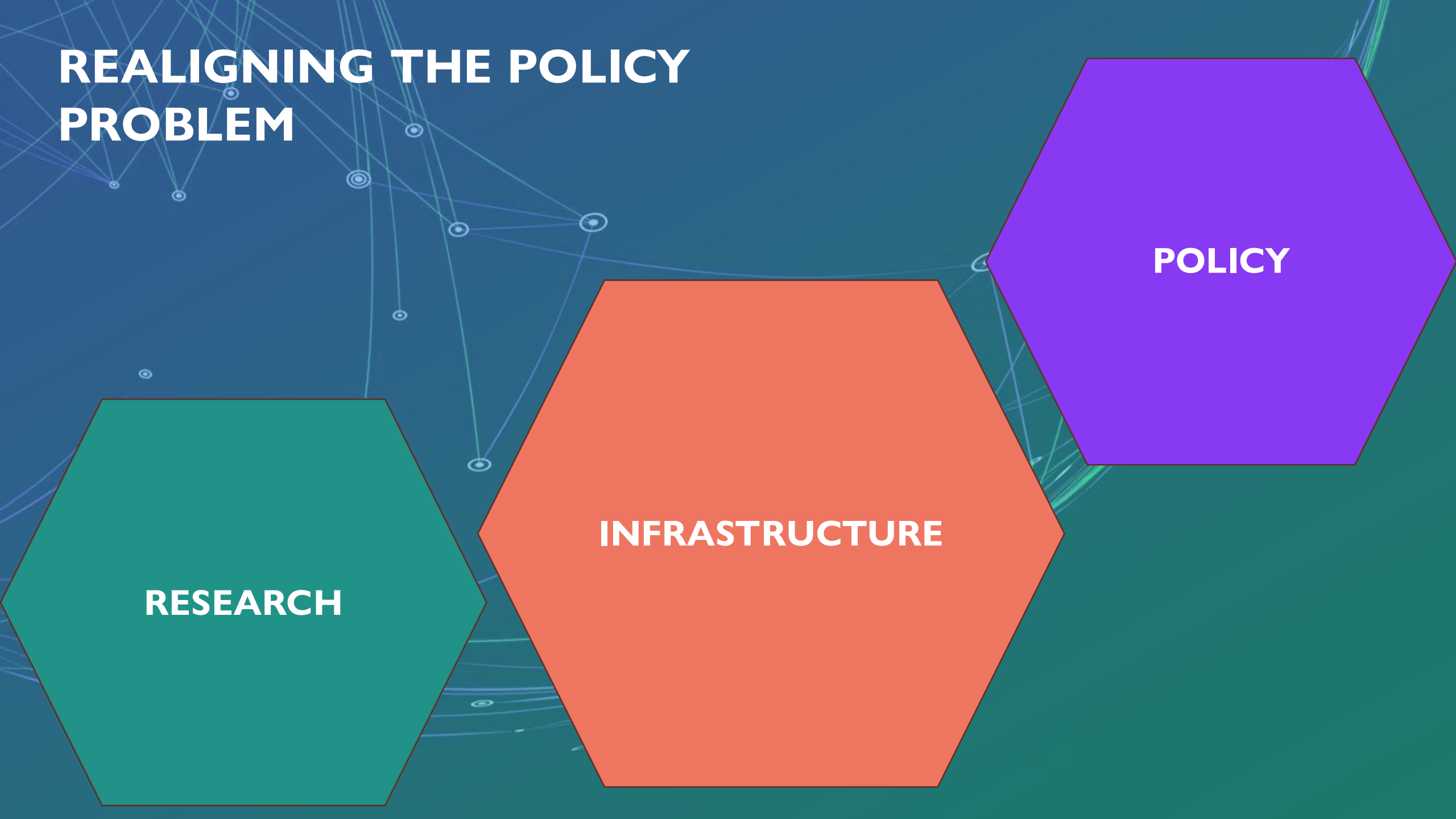
**FUNDERS AND POLICY  
PROFESSIONALS**

# REALIGNING THE POLICY PROBLEM

**RESEARCH**

**INFRASTRUCTURE**

**POLICY**



# MAPPING THE USER ECOSYSTEMS OF LARGE-SCALE HERITAGE COLLECTIONS



HERITAGE PUBLICS



HERITAGE  
COLLECTIONS  
AS DATA





# TAKE AWAYS

- 1. FUNDAMENTAL CHANGE TO THE ORGANIZATION OF SCHOLARSHIP•**
- 2. DATA AND DATA INTENSIVE SCHOLARSHIP IS FUNDAMENTALLY ABOUT COLLABORATION.**
- 3. RESEARCH INFRASTRUCTURES SHAPE COLLECTIVE AND AGENCY—AS SCHOLARS BUT ALSO AS PARTICIPANTS IN SCHOLARLY DIALOGUE**

# REFERENCES

---

Einav, Liran, and Jonathan Levin. 2013. “The Data Revolution and Economic Analysis.” *NBER Working Paper No. 19035*. Cambridge, MA: National Bureau of Economic Research.

Kitchin, Rob. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE Publications, 2014.

Texas State Library and Archives Commission. “The Relationships Between Data, Information, and Records.” *The Texas Record*. May 18, 2022. [Texas State Library and Archives Commission article](#)