

Research Data EcoSystem

- FAIR data is a cyberinfrastructure challenge
- Researchers increasingly expect AI-ready data and infrastructure
- RDE = national cyberinfrastructure for Social Science Data

AI making FAIR Data at Scale in RDE

AI in RDE production workflows

- [TurboCurator](#) uses AI to enhance metadata at ingest
- Study HomePage as an AI-ready analytical environment
- AI ready cloud infrastructure (COBRE) - coming in 2026
 - Secure cloud environment for computation with sensitive data
- Machine-Consumable Access - coming in 2026
 - MCP infrastructure enables AI systems to interact directly with RDE data

Future Role of AI in RDE

- Private LLMs and/or SLMs
- AI for Search
- Data Chatbots
- Automating data readiness
- Integrating additional notebook types

AI and the Social Science Data LifeCycle

CURRENT USES

FUTURE POSSIBILITIES

Data Discovery & Planning

- Conduct preliminary research/ literature review
- Explain research methods
- Discover well-known datasets

- "Data Concierge" chatbots
- Stronger explanation of existing methods/literature
- Strategies for addressing research questions

Data Collection

- Draft survey questions
- Draft data collection protocols/procedures
- AI as survey respondent

- Conducting multilingual survey interviews
- Ensuring cultural/social/temporal consistency

Data Preparation & Analysis

- Write, clean, & analyze code
- Optimize code efficiency
- Categorize qualitative data
- Extract data from nonstandard formats

- Improving qualitative data categorization
- Analyzing datasets with AI models

Publication & Sharing

- Draft research papers and data summaries
- Create replication code/text

- Assisting metadata creation/organization
- Identifying papers using a dataset
- Detecting PII/PHI before sharing

Long-term Management

- Convert data to accessible formats
- Identify access barriers

- Digitizing records from scanned documents
- Identifying mergeable datasets/measures

AI Considerations

What is critical to implement for the research data pipeline?

When to be a “first mover” vs using existing tools/software?

How to use AI responsibly?

- Smaller LLMs might be better and less costly