

# Data Inspection Method (WiP)

The Data Inspection Method (name is a WiP) is an ethnomethod designed to investigate core human-data interaction activities and data literacy in the environment. Although part of the method involves examining the features of the data, all inquiries are done from the human perspective while objective facts are gathered from technology that interfaces with the data. This allows for a balance of subjective and objective data collection for analysis.

## Key Concept Definitions

<b>Human-Data Interaction</b>	The set of interactions between humans and data to accomplish key tasks.
<b>Data Literacy</b>	Set of core competencies for interacting with data according to the needs of the business environment.

## Why the Data Inspection Method?

- Identify what data is being interacted with.
  - Who is interacting with the identified data?
  - When are they interacting with it?
  - Why are they interacting with it?
- What are the features/qualities of the data being interacted with.
- Identify steps in the interactions.
- Map how:
  - Users organize around the data.
  - Technology mediates HDI.
  - Processes organize around the data.
  - Data interactions chain together.
- Understand communication about the data.

## Credibility, Transferability, Confirmability, and Dependability

Creating a new research method means ensuring 4 criteria are met: Credibility, transferability, confirmability, and dependability. Taken together, these 4 criteria create a level of *trustworthiness* in the study.

- **Credibility**- confidence in the truth of the research study's findings.
  - Employ *triangulation*, multiple methods of getting at a set of information.
    - This method will use interviews, individual ratings, observations, and artifact collection. A mix of subjective and objective information will paint a rich and accurate picture of the experience in the data environment.
    - If possible, more than one user should participate in data collection in the same environment.
- **Transferability**- applicability of findings to other environments.
  - Document environment and observations via *thick description*, a way of providing in-depth descriptions of observed phenomena.
    - Researcher will take notes and collect artifacts of observed phenomena.
- **Confirmability**- neutrality/lack of bias in the findings.
  - Creation of an *audit trail*, detailed records of every step of data collection and data analysis.
    - All data collection and analysis tools will be documented with instructions for use.
    - All data analysis will be held to pre-determined standards, and any deviations noted with the justification.
- **Dependability**- repeatability of the study with the finding of consistent results.
  - Use of *inquiry audits*, providing other researchers with methodology documentation and research artifacts to ensure that the method is sound.
    - Researcher must turn over research documentation when requested for other researchers to review.

## When to use the Data Inspection Method

Right now, this method is intended for use early on in the AI-SDLC. Over the course of an ML/AI application's lifecycle, as new users are onboarded, it might be possible to adapt the methodology to investigate their HDI activities.

## Data Inspection Method Tool Kit Overview

Being an in-development ethnomethod, some basic tools are defined below. Like the method, these tools are new and over time will be refined via use.

## Interview Guide

As with many interview guides, the recommended general format for the session is to open with a set of introductory questions to assess the user's background in the data environment, follow up with a bulk of questions centered around human-data interactions (technologies used, processes, use, interaction chains), and end with questions to wrap up any questions from the interviewee. Please note that what is outlined below is not an exhaustive list, and will be open to change as the method evolves.

- **Introduction**
  - Background Assessment
  - Role + Responsibilities in the Company
  - Team in the Company
- **Body of the Interview**
  - What data do you interact with?
  - How do these interactions link together?
  - What is the value of these interactions to you?
  - Who do you communicate with about the data?
  - Participant's subjective assessment of their data fluency.
- **Wrap Up**
  - What could be done to improve how you conduct your interactions with data?
  - What could be done to improve communication about the data in your work environment?
  - (Potential, not always applicable) Would you be interested in participating in future research?
  - Any questions for the interviewer.

## Data Evaluation Guide

This inspection method relies on collecting objective information about the data to be used by an ML/AI application. This information can be collected via technology, and should focus on the following aspects:

- Content- what specifically is the data about?
- Format- how is the data stored?
- Structure- what is the organization of the data?
- Cleanliness- how orderly is the data within the structure?
- Relevance- how well does the stored data serve the problem to be solved by ML/AI?

## Human-Data Interaction Flow Diagrams

Just as products have flows for user interactions, data has flows for user interactions. These interactions will be impacted by the processes put in place for interacting with data. Researchers should document these flows as there might be a flow that is prescribed by a defined process, but the actual implementation of that process (the steps taken) might deviate due to user or technology issues.

### **What is being depicted?**

- Directionality of information movement: Backward/forward, one-way, bi-directional, conditional.
- Start points, end points, jumping off points.
- User interactions

### **Data Environment Map**

Processes, technology, and users all organize around data. That organization—the data environment—should be documented. This information should have formal business documentation, but the user summary of the environment should also be recorded.