# Desiloing Your Data Strategy - Worksheet

## Diving into the Details of What You’re Doing Now (~30 minutes)

* **Data Acquisition and Cleanup** 
  + Questions/Concerns: Raw data is imperfect; dubbed “messy”
  + Examples of Relevant Tools: MS Excel, paper, emails, etc.
  + Desired Output: Consistency, dependability, higher data quality, and data integrity
* **Data Storage and Management** 
  + Questions/Concerns: Housing data is a non-trivial pursuit; in cloud, in a data center, on local machines
  + Examples of Relevant Tools: MS Excel, MySQL, Workbench Oracle, NoSQL, Cassandra, MongoDB, etc.
  + Desired Output: Data that is accessible and available
* **Data Analytics** 
  + Questions/Concerns: Not every data outcome is a data insight
  + Examples of Relevant Tools: Python, R, SAS, SPSS, MS Excel.
  + Desired Output: Increased quality of data and ability to ask questions like, “what does the data mean?”
* **Data Visualization** 
  + Questions/Concerns: Visualizing data outcomes can misled decision-makers
  + Examples of Relevant Tools: Python, R, Tableau, SPSS, MS Excel, QlikView, QlikSense, PowerBI.
  + Desired Output: Summarize, digestible version of analytics and insights (when done well)
* **Data Storytelling**
  + Questions/Concerns: Communicating data insights should be informative; avoid mis/disinformation
  + Examples of Relevant Tools: GitHub, The Binder Project, DeepNote/Jupyter, MS Word, Blogs, Media articles
  + Desired Output: Communication and presentation of insights geared toward a particular audience through text, video, audio, images, and narratives

***Small Group Activity:*** What is your data tool stack for each phase of the data pipeline?

If you don’t know, leave that phase of the data pipeline blank.

***Small Group Activity Notes:***

* Data acquisition and cleanup —
* Data storage and management —
* Data analytics —
* Data visualization —
* Data storytelling —

**Operating from Ethically-Strong and Ethically-Weak Positions (~30 minutes)**

**Ethically-weak data process criteria (generic)**

* Efficiency over effectiveness
* Computational optimization and scalability
* Machine-centered over humanity-centered
* Focus on *what* the tech does over *why* or *how* the tech achieves outcomes

Examples: Automated employee candidate selection platforms, automated hate speech detection, automated writing tools, etc.

***Small Group Activity:*** Identify and explain ONE ethically weak data operation being used by your team/unit for ONE phase of the data pipeline?

***Small Group Activity Notes:***

Circle/highlight ONE phase of the data pipeline below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data acquisition and cleanup | Data storage and management | Data analytics | Data visualization | Data storytelling |

Guiding questions (adopted and adapted from *Data Conscience* book) for each phase are provided to help pinpoint the potential ethical weakness. Complete this activity ***only*** for the data pipeline phase you circled/highlighted above.

* **Data acquisition and cleanup** (Does the data we need already exist? Do we own the rights to do what we want with the data? Have we decided on the data structure or platform for our analysis?)

Your response here:

* **Data storage and management** (Does the data fit in the digital infrastructure? Is the data complete in the digital infrastructure? Is the data correct in the digital infrastructure?)

Your response here:

* **Data analytics** (How does your team select which computational procedures, models and algorithms are used? How does your team verify outcomes associated with data processes, algorithms and systems? How does your team evaluate these outcomes as algorithmic-based harms to different communities?)

Your response here:

* **Data visualization** (When are the data and software teams assured that product-users are able to make sense of any data outcomes and insights after they’re done analyzing it? When in the data pipeline is the team confident in the processes used and in their understandings of the outcomes?)

Your response here:

* **Data storytelling** (Who is your team’s audience? Is it clear who the targeted audience is? What does your tend intend to communicate? Is the narrative accurate and representative of the data insights?)

Your response here:

**Expanding Your Data Equity Deployment Options (~25 minutes)**

* How does your team/unit understand the workplace data ecosystem, e.g., data privacy and data literacy?
* How does your team/unit document the workplace data ecosystem?
* What are the preemptive measures taken to prevent algorithmic harms?
* What is the process to audit and rectify the team/unit’s data operations?

Examples: Algorithmic destruction implemented by the FTC in [WeightWatchers case](https://www.protocol.com/policy/ftc-algorithm-destroy-data-privacy), Expand [race/ethnicity standards proposal](https://www.federalregister.gov/documents/2023/01/27/2023-01635/initial-proposals-for-updating-ombs-race-and-ethnicity-statistical-standards) by US OMB

***Discussion:***

What methods and frameworks have your team/unit tried to address ethically-weak data operations? Write/type the methods and/or frameworks in the Discussion Notes section. Share one approach for each phase of the data pipeline.

***Discussion Notes:***

* Data acquisition and cleanup —
* Data storage and management —
* Data analytics —
* Data visualization —
* Data storytelling —

**Wrap-Up and Conclusion (~5 minutes)**

Key takeaways Notes:

Other Notes: