

Assignment 1: Data Science 2026 ITWS/CSCI/ERTH 4350/6350 (10% of overall score)

Due: September 25th, 2026 (by 10:00 pm ET)

Submission method: LMS

Document naming: DataScience_2026_Assignment1_TEAM_NUMBER_OR_NAME.ext (e.g. txt, pdf, doc, zip).

Late submission policy: first time – no penalty, otherwise 20% of score deducted each late day.

Use the numbering below when completing your responses to this assignment.

Assignment: propose **two different data collection exercises (labeled A and B)** and perform a survey of data formats, metadata standards, software tools, cloud services, etc. for the management of data to be collected in Assignment 2. This is the planning process for data collection, and you must plan properly and in details before you collect.

Remember!

Modes of data collection:

- Observation
- Measurement
- Generation

Driven by

- Questions
- Research idea
- Exploration

You MUST complete the following questions for BOTH collection ideas.

1. Data collection – propose two data collection options - 4%

a. State the details of the **mode** of each collection and what the data collection need is being **driven** by. Suggested minimum response is 6-7 sentences. Include any drawings/diagrams/sketches if available (these are not mandatory, feel free to include them if you create them).

b. Describe a **data management plan** for the data and metadata acquisition, and initial curation using the **9 components** of Data Management in the lecture slides from week 2. Minimum 4-6 sentences per category.

2. Survey of data storage/ formats - 3%

a. Based on Q1 (for both proposed collections), research and describe existing suitable data formats that could be used. Provide references (URLs) for uncommon data formats (e.g. FASTA for bioinformatics, KML for geospatial). If no suitable format matches your data needs exactly, consider how you may restructure the data to fit existing formats and describe these choices. Minimum 5-6 sentences.

3. Survey of metadata conventions, standards - 3% (4000-level), 2% (6000-level)

a. Based on Q1 (for both proposed collections), research and describe existing **metadata conventions/ standards** that could be used. If none are suitable, describe the metadata needs and how you may record these metadata without a standard. Minimum 4-5 sentences.

4. 6000-Level student question - 1%

a. Describe the provenance information you plan to collect (for both proposed collections) and how that may support the overall data collection and goals potential analyses/investigations. Minimum 6-7 sentences.

That's it for now..