

INFX 551: Fundamentals of Data Curation

Nicholas Weber

Contents

Introduction to INFX 551	5
1 Data	7
1.1 Types & Roles	7
1.2 Metadata & Provenance	8
1.3 Cleaning & Standardization	8
1.4 Assignment	8
2 Systems	9
2.1 Repositories & Storage	11
2.2 Preservation & Levels of Service	11
2.3 Cost Models & Sustainability	11
2.4 Assignment	11
3 Policy, Privacy, & Ethics	13
3.1 Policy: Data Management Plans & Attribution/Citation	15
3.2 Privacy: Institutional Review Boards & Data Minimization	15
3.3 Ethics	15
3.4 Assignment	15
4 Assignments Overview	17
4.1 Assignment 1 Details	17
4.2 Assignment 2 Details	17
4.3 Assignment 3 Details	17

Introduction to INFX 551

This is where I give you an overview of the course and tell you how much \$ you will make being a data curator.

Talk about how the modules will work: Monday lecture about overview of the topic Thursday deeper dive into the topic.

Chapter 1

Data

This three-week module will focus on data.

1.1 Types & Roles

1.1.1 Overview

1.1.2 Lectures

The video below is just a test of embedding youtube videos. The link to use is the one you find in the embed code. Not the url to the YouTube video. Note the word “embed” within the url:

1.1.3 Required Readings

1.1.4 Recommended Readings

1.1.4.1 Science

1.1.4.2 Humanities

1.1.4.3 Public Sector

1.2 Metadata & Provenance

1.2.1 Overview

1.2.2 Lectures

1.2.3 Required Readings

1.2.4 Recommended Readings

1.2.4.1 Science

1.2.4.2 Humanities

1.2.4.3 Public Sector

1.3 Cleaning & Standardization

1.3.1 Overview

1.3.2 Lectures

1.3.3 Required Readings

1.3.4 Recommended Readings

1.3.4.1 Science

1.3.4.2 Humanities

1.3.4.3 Public Sector

1.4 Assignment

The assignment for Module 1 is to create a data dictionary.

Chapter 2

Systems

This three-week module focuses on systems.

2.1 Repositories & Storage

2.1.1 Overview

2.1.2 Lectures

2.1.3 Required Readings

2.1.4 Recommended Readings

2.1.4.1 Science

2.1.4.2 Humanities

2.1.4.3 Public Sector

2.2 Preservation & Levels of Service

2.2.1 Overview

2.2.2 Lectures

2.2.3 Required Readings

2.2.4 Recommended Readings

2.2.4.1 Science

2.2.4.2 Humanities

2.2.4.3 Public Sector

2.3 Cost Models & Sustainability

2.3.1 Overview

2.3.2 Lectures

2.3.3 Required Readings

2.3.4 Recommended Readings

2.3.4.1 Science

2.3.4.2 Humanities

2.3.4.3 Public Sector

2.4 Assignment

The assignment for Module 2 is...

Chapter 3

Policy, Privacy, & Ethics

Finally, we talk policy, privacy, and ethics...

3.1 Policy: Data Management Plans & Attribution/Citation

3.1.1 Overview

3.1.2 Lectures

3.1.3 Required Readings

3.1.4 Recommended Readings

3.1.4.1 Science

3.1.4.2 Humanities

3.1.4.3 Public Sector

3.2 Privacy: Institutional Review Boards & Data Minimization

3.2.1 Overview

3.2.2 Lectures

3.2.3 Required Readings

3.2.4 Recommended Readings

3.2.4.1 Science

3.2.4.2 Humanities

3.2.4.3 Public Sector

3.3 Ethics

3.3.1 Overview

3.3.2 Lectures

3.3.3 Required Readings

3.3.4 Recommended Readings

3.3.4.1 Science

3.3.4.2 Humanities

3.3.4.3 Public Sector

3.4 Assignment

Chapter 4

Assignments Overview

Here is a detailed description of all three assignments.

4.1 Assignment 1 Details

Data Dictionary

4.2 Assignment 2 Details

4.3 Assignment 3 Details