

デジタルトランスフォーメーション(DX)時代のデータ キュレーションと情報管理

Imker, Heidi
University of Illinois at Urbana-Champaign

Mischo, William H.
University of Illinois at Urbana-Champaign Library

Wiley, Chris
University of Illinois at Urbana-Champaign

Schlembach, Mary C.
University of Illinois at Urbana-Champaign Library

他

<https://hdl.handle.net/2324/4738597>

出版情報 : 2021-12-09. University of Illinois Urbana Champaign
バージョン :
権利関係 :

Teaching Data Curation at Illinois: *Two Courses*

Allen H. Renear
School of Information Sciences
University of Illinois Urbana-Champaign

December 8, 2021

Overview

A question for discussion:

What programs and courses do we need in data curation?

To stimulate reflection:

School of Information Sciences

IS547: [data curation for *curation services managers*]

Department of Computer Science (in the College of Engineering)

CS598: [data curation for *data scientists*]

Data curation: the Illinois definition

From Wikipedia:

“According to the University of Illinois [School of Information Sciences],

"Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education; curation activities enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time.”

“Data Curation”, Wikipedia. , 2017.

Original source: “An Educational Program on Data Curation”. *ALA Science & Technology Section Conference*. Cragin, Melissa; Heidorn, P. Bryan; Palmer, Carole L.; Smith, Linda C. (2007).

Or, alternatively:



WIKIPEDIA
The Free Encyclopedia

*Ensuring that data can be used **effectively**, **efficiently**, and **reliably**.*

Data Curatorial Activities

Collection:	Support the collection and acquisition of data
Organization:	Employ an appropriate data model and use appropriate standards
Storage:	Support reliable and effective storage
Preservation:	Ensure that data will be understandable and useable in the future
Discoverability:	Support the ability to search for and locate relevant data
Access:	Support the ability to retrieve and distribute data
Workflow:	Support the ability to systematize data workflows
Identification:	Support the ability to identify, authenticate, and validate data
Integration:	Support integration of data from different sources using different data models
Reformatting:	Support reformatting for use by different tools or to match new format standards
Reproducibility:	Support ability to reproduce results, ensuring scientific validity
Sharing:	Support sharing data between researchers, teams, and institutions.
Communication:	Support representation, publishing, and visualizations that provide insight
Provenance:	Support identifying what inputs and calculations are responsible for data values
Modification:	Support management of corrections and updates
Compliance:	Ensure compliance to legal, regulatory, local policy requirements; ethics, bias, equity, inclusion
Security:	Ensure that data is secure from tampering or inappropriate access and distribution

Data Curatorial Methods

Analysis

Documentation

Classification

Standardization

Validation

supported with:

policies, systems, and computational infrastructure.

School of Information Sciences

“... human-centered interdisciplinary approach to the understanding, design, and management of information systems and services...”

A top-ranked iSchool

known internationally for data curation research and teaching.

A particular focus on programs providing professional preparation:

MS in Library and Information Science

MS in Information Management

Data Curation for Curation Services Managers

IS547: Foundations of Data Curation (School Information Sciences)

Designed particularly for students who anticipate:

Designing and managing data curation services in various organizational settings

Managing organizations that provide data curation services

Working with data curators

Some distinctive topics:

Data practices in different disciplines

Organizing and managing data curation services

Data governance and legal and ethical issues

Agencies, organizations, and associations related to data curation

Mode of delivery:

Face to face and online synchronous; 20-35 students

Department of Computer Science

- Relevant Program:** Master of Computer Science – Data Science
- Collaborating units:** Computer Science, Statistics, Information Sciences
- Focus:** Advanced applications of data science methods
- Program Modality:** Asynchronous (Coursera)
- iSchool contributions:** *Foundations of Data Curation,
Theory and Practice of Data Cleaning*

Data Curation for Data Scientists

CS598: Foundations of Data Curation (Department of Computer Science)

Offered every year. 300 students. Asynchronous Coursera course.

Students have a computer science background; most already employed as software engineers.

Optimized for preparing...

- Data scientists

- Software engineering

- Managers of data-intensive engineering

Examples of distinctive Emphasis (in addition to fundamentals)...

- data models

- data/schema integration (heterogeneity, interoperability...)

- workflow and data provenance

- data cleaning

- reformatting, serialization, validation

- transformation and extraction

- identity, integrity, and scientific equivalence

For discussion

How should programs and courses be designed?

What combination of courses?

for what sorts of professional roles?