7-11-2019

Developing, Delivering and Redesigning Metadata and Data Documentation Workshop for Graduate Students

Sai Deng
University of Central Florida, sai.deng@ucf.edu

Original Citation

Deng, S. (2019). Developing, delivering and redesigning metadata and data documentation workshop for graduate students. UCF Library Faculty Showcase “Pecha Kucha Go! Wave”. July 11, 2019.

Developing, Delivering and Redesigning Metadata and Data Documentation Workshop for Graduate Students

Sai Deng, Metadata Librarian

University of Central Florida Libraries

Intro to the Workshop

- **Part of the Dataset Metadata and Metadata Services embedded into the Research Lifecycle at UCF.**
- **One hour workshop, in lecture style, covered a large amount of information:**
  - Status of data documentation and management as revealed by a campus survey;
  - Research data documentation basics;
  - General and domain metadata standards, thesauri, data citation;
  - Data documentation practices in different disciplines, data repositories, tools…
- **It was delivered to the UCF students as part of the UCF Libraries Graduate Workshops (coordinated by Corinne) initially in 2014 and 2015.**

Context of the Initial Workshop

- **The UCF Research Data Management Survey (Beile, https://stars.library.ucf.edu/lib-docs/144/)**
- **Results from Data Collection, Recording & Analysis Sections**
  - **What type(s) of data do you generate? Please indicate an approximate percentage.**

Answer	Average Value	Responses
Numerical data, e.g. ocean temperatures (%)	62.03	73
Text, e.g. historical records and literature (%)	28.35	48
Still images (%)	24.19	37
Audio files (%)	27.37	19
Video files (%)	22.95	21
Medical data, e.g. patient health information (%)	49.47	17
Biochemical data, e.g. raw and processed "omic" data (%)	18.17	12
Tabulated data (%)	34.79	39
Other (%)	53.58	10

Context of the Initial Workshop

- **What format(s) are your data in?**

Type	Response	Annotation
Audio	9	Audio (2), .mpeg, .mp3 (2), .mp4, .wav (2), .wma
Databases	11	Filemaker, Online survey database, .dat (3), .sql (2), .mat (4)
Geographic information data organizers	4	.gis, .lyr, .prj, .shp
Graphics	18	.gif (3), .jpg (7), .png, .tif (7)
Presentation	2	.ppt (2)
Remote sensing	1	LiDAR
Scientific data	1	.fits
Simulation engines	2	.bpp BEopt, .enb
Source code	5	.cpp, .stk, hyperRESEARCH files, HDF5, VTK
Spreadsheets	59	.xls and .xlsx (45), .csv (12), .jnb (2)
Statistical analysis software	32	.dta (3), .jmp, minitab, SAS (8), SPSS (17), STATISTA, statistical files
Text	55	.pdf (8), .doc and .docx (31), .asc (2), .txt (14)
Video	2	.mov, .wmv
Virtual machines(?)	4	.sav (4)

**Note that some files may not be categorized correctly. This was a best guesstimate.*

Source: The UCF Research Data Management Survey (https://stars.library.ucf.edu/lib-docs/144/)

Context of the Initial Workshop

- **How is your data labeled or annotated?**

Answer	Response	%
Automatically, through a data collection tool	37	44%
Manually, by myself or a member of my research team	65	77%
Referentially, with an associated codebook	22	26%

- **How are you recording lab data?**

Answer	Response	%
Lab notebooks in paper	29	59%
Excel (or other) files on computers in the lab	48	98%
Electronic lab notebook (ELN) tool. Please specify which one.	3	6%

- **More popular tools**: SAS/SAS Enterprise version, MATLAB, SPSS, R-project programs, NVivo, SigmaPlot
- **Most popular choice for "Lab Notebook":** Excel (98%), Lab notebooks in paper (59%), Electronic lab notebook (6%)

Source: The UCF Research Data Management Survey (https://stars.library.ucf.edu/lib-docs/144/)

Context of the Initial Workshop

- **Do you document or record any metadata for your data or dataset?**

Answer		Response	%
Yes		21	34%
No		41	66%
Total		62	100%

- **If you record metadata for your dataset, do you use any local, agency-specific, or national standards or guidelines?**

Answer		Response	%
Yes (please specify)		5	24%
No		15	71%
I'm not sure		1	5%
Total		21	100%

Standards or guidelines filled in: HIPAA/FERPA, FITS standard, program specific, librarians are helping us with this, and all of the above.

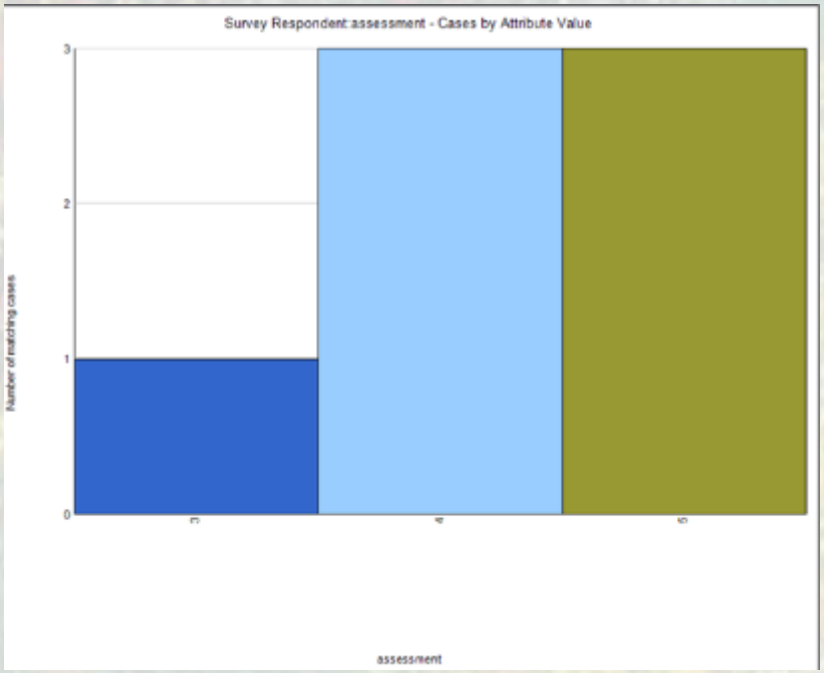
Source: The UCF Research Data Management Survey (https://stars.library.ucf.edu/lib-docs/144/)

Data Documentation in 1 minute
(or 20 seconds)

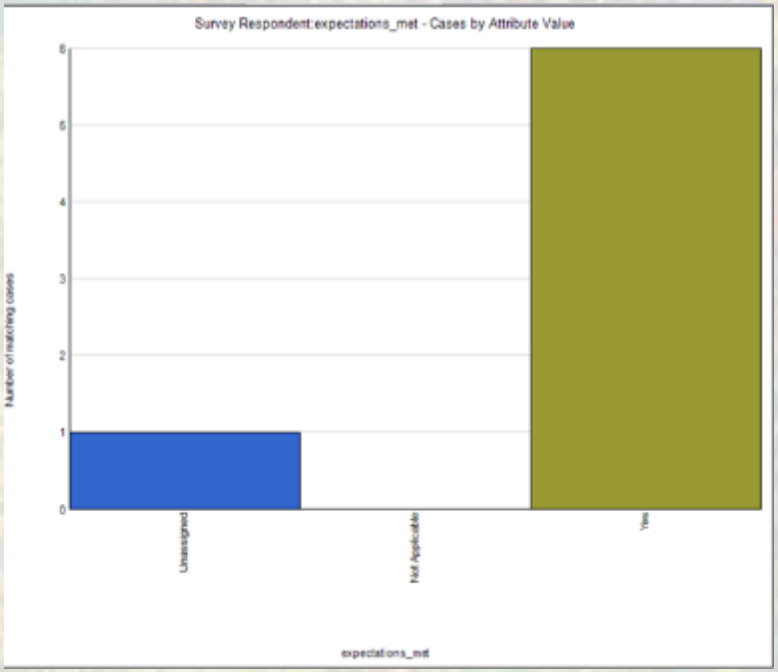
- **Data documentation:** records data's creation, meaning, content, structure, manipulation; necessary for data sharing, reuse and long-term preservation.
- **The various kinds of documentation** may include:
  - **Embedded documentation** (included within the data, e.g., code, field and label descriptions, syntax, descriptive headers or summaries, transcripts)
  - **Supporting documentation** (in separate files, e.g., readme, project information, methodology report, working papers, lab books, questionnaires or interview guides, reports & publications)
  - **Catalog Metadata** (for data archiving, identification and locating)
- In other words, Research data can be documented at various levels: **Project level, File/database level** and **Variable/item level**.
- **Metadata** can be taken as **a type of data documentation**; can be embedded in data files, produced and recorded in the research lifecycle.
- It is recommended to document **all data collected and generated through your research lifecycle**: materials, research data format…
- **Consider:** what information is needed, funder requirements, field standards (discipline metadata & data documentation standards), appropriate tools, identifier, thesauri, data citation, (data) repositories…

Student Experiences and Feedback

- **Short surveys were conducted in two workshops; received participants' responses.**



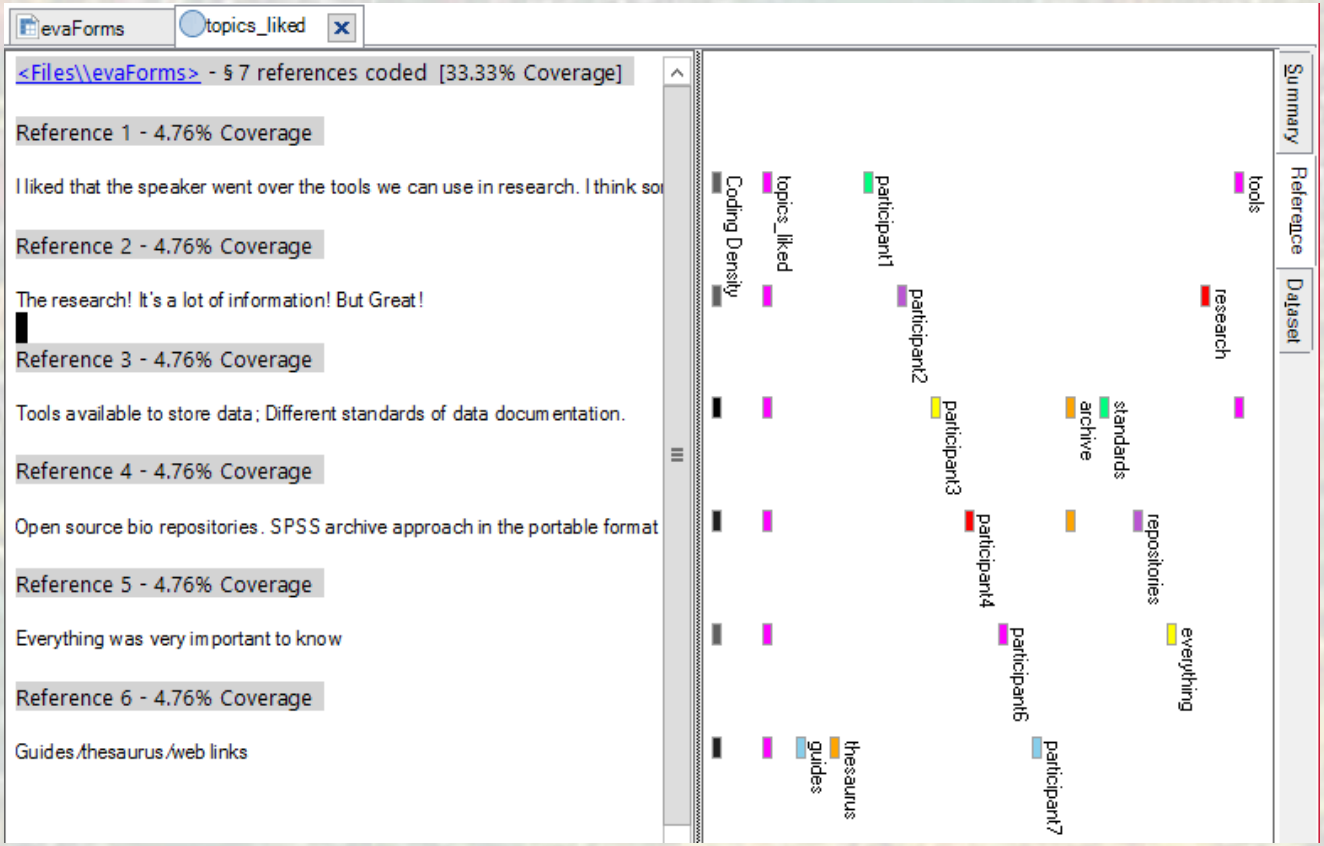
What is your overall assessment of the event? (1=insufficient, 5=excellent)



Knowledge and information gained from participation at this event?
(Met your expectations Yes No Somehow Unassigned NA)

Student Experiences and Feedback

- **Which topics or aspects of the workshop did you find most interesting or useful?**

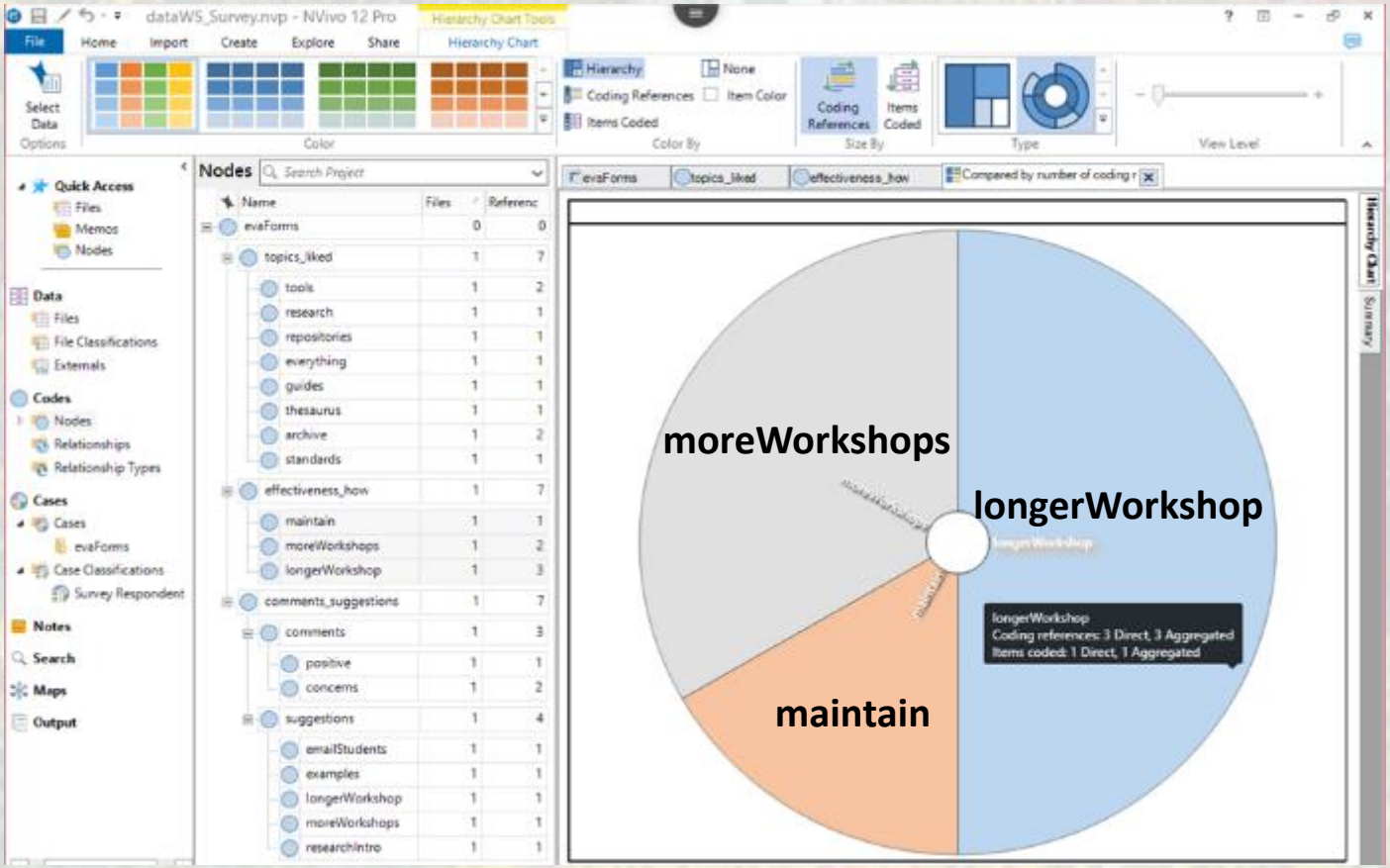


topics_liked

topics_liked: Word frequency query results

Student Experiences and Feedback

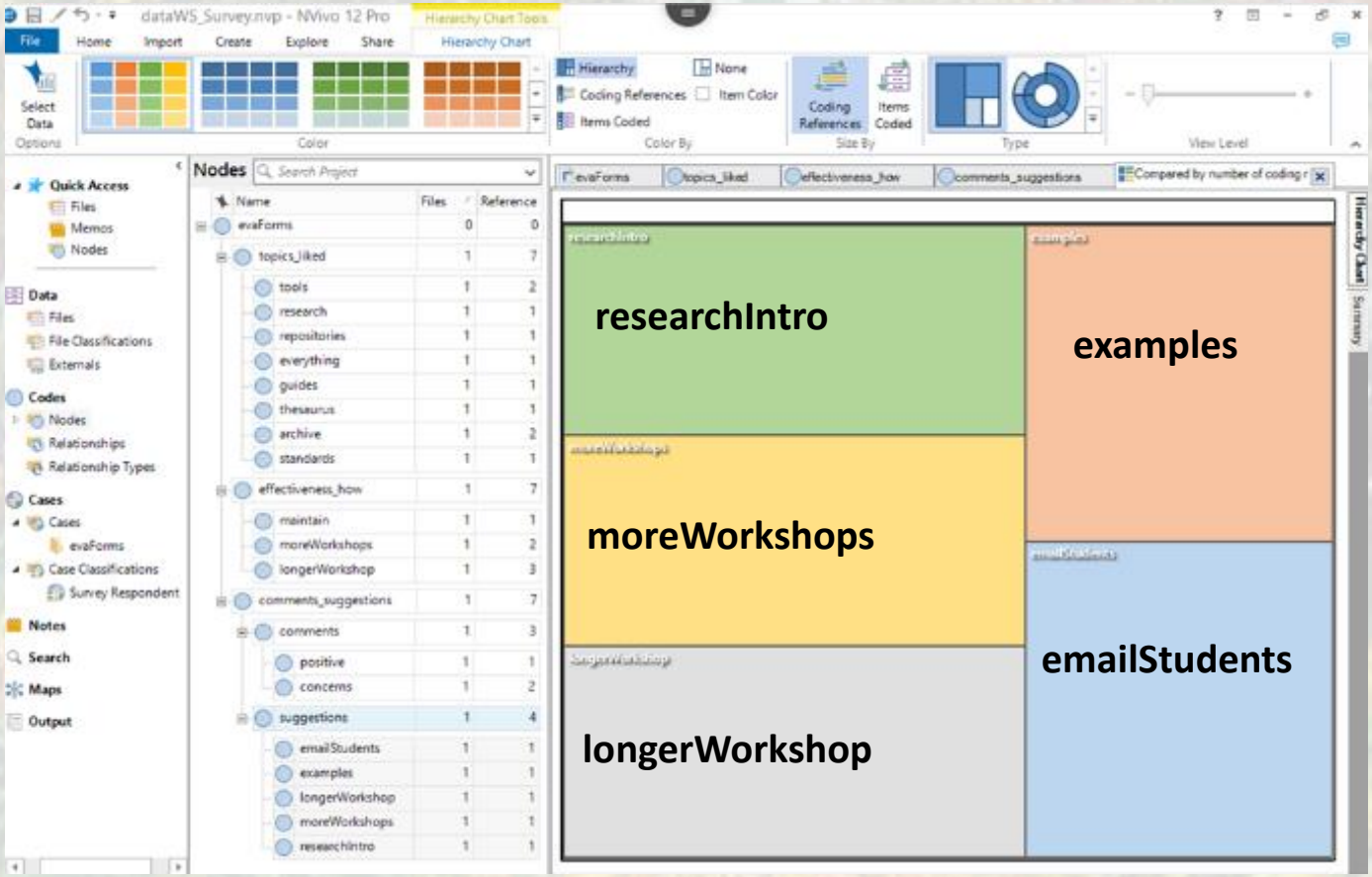
- **How do you think the workshop could have been made more effective?**



effectiveness_how (Hierarchy chart: Compared by number of coding references)

Student Experiences and Feedback

- **Comments and suggestions** (including topics or activities you think would be useful, for the future)



suggestions (Hierarchy chart: Compared by number of coding references)

ICPSR Data Fair

- **ICPSR Data Fair 2018**
- **Learned more about data related concepts and aspects**
  - Data deposit, Access, Sharing, Data transparency, Research reproducibility, Community tools, Ethics in social research, Data training, Diversity, Equity and inclusion in data...
- **Sessions by both data curators and practitioners; learned some real cases**
  - Census Bureau case, data portal, community data archive, colleges' & universities' data practices
- **Got to know more about good data habits and practices** (e.g., in data analysis & presentation).

Image source: https://library.temple.edu/beyondthepage/event/learning-good-data-habits

Data Curation Network

- **Data Curation Network** (**DCN**, https://datacurationnetwork.org/)
  - A collaborative model for curating research data across academic and general data repositories.
- Participated in the DCN Specialized Data Curation workshop @ DLF;
- **Created the SPSS Primer** (Deng, Dull, Finn & Khair)
  - https://conservancy.umn.edu/handle/11299/202812
- **Reflections:**
  - DCN and DCN primers are more designed for the curators, however,
  - They can help data practitioners to pay more attention to good data practices (including data documentation practices) for different data formats.
  - Shared training materials with Digital Initiatives.

DATA CURATION NETWORK

Image source: https://datacurationnetwork.org/

Redesign

- Make the workshop more **researcher centered** rather than librarian centered;

- Use examples from the survey to address issues and make **recommendations for best practices**;

- Add more emphasis on **tools**, specially on the quantitative analysis software SPSS and qualitative analysis software NVivo, and provide details on how to use them for data documentation;

Redesign

- Add more **activities** on how to document research datasets;
- Continue to cover data repositories, metadata standards and vocabularies, but include more information in **appendixes**;
- Cover less information in the workshop, supplement it with **handouts** of resources;
- Could consider making it a series in the future (if attendance is good).

Modules

Part I: The Data Basics

- **Understanding Data, Research Data and Datasets**
- **Why data documentation** (Q)
- **Data documentation & Metadata**

Part II: Data Documentation

- **Practices & Recommendations** (Q, E, D)
- **Data Documentation: Study-level** (E, Ex, D)
- **Data Documentation: Data-level** (Quantitative/ Qualitative data; SPSS/ NVivo) (E, Ex, D)

Q: w/ question. E: w/ examples. Ex: w/ exercise. D: w/ discussion.

Modules

Part III: Dataset Metadata

- **Dataset Metadata** (Q, E)
  - Dataset Metadata ABC
  - Metadata Standards
  - Vocabularies and Thesauri
  - Disciplinary Metadata & Data Repositories
- **Curation Tools for Datasets** (E, D)

Part IV: Dataset/Metadata Service

- **Dataset and Metadata Service at UCF Libraries** (D)
- **Library & Campus Resources**
- **Appendixes**

Q: w/ question. E: w/ examples. Ex: w/ exercise. D: w/ discussion.

Deng, S. Introduction to Data Documentation: https://www.academia.edu/39625696/Introduction_to_Data_Documentation

Exercises & Activities

Questions & Discussions:

- **Current Data Documentation Practices**
- **Tools used**
- **Lab Notebooks used**
- **Standards/Guidelines used**

Exercises:

- **Case Analysis for Data Documentation**
- **Project/Study Level Metadata Choices**
- **Data Level Metadata Exercises** (e.g., Variable Naming, Label, Value)

Final Thoughts

- The Data Documentation Workshop at UCF, as part of the Metadata and Dataset Metadata Services embedded into its Research Lifecycle, contributes to the overall goal of serving its graduate students and researchers better;

- The researcher's view and the librarian or curator's view on data documentation differ in some ways which will affect how the workshop can be designed and delivered;

- The workshop can also provide information on other research data related services (such as institutional repository, research data management, scholarly communication, research services);

- Librarians need to be aware of developments in data related practices and services and learn new resources, knowledge and skills to meet the needs of our users.

Contact:
Sai Deng, Metadata Librarian &
Associate Librarian
sai.deng@ucf.edu
407-823-4312 (Office)

Thank you!

Credits:
Cover & Background:
Claude Monet, Garden at Giverny Arches, 1900.