

Part 1: SABBATICAL PLAN

Board of Regents and CU System policies require the following information be provided by each faculty member applying for a sabbatical assignment. Thank you for completing this public document in a clear and substantive way. Each response should be a minimum of 300 words.

Applicant Name: [REDACTED]

Title of Sabbatical Project: Investigating Document Theory in Relation to Large Language Models

(1) Describe your sabbatical's academic objectives including its contribution to your professional growth and expertise.

The advent of large language models (LLM) and artificial intelligence has brought libraries and information science to a new inflection point. In particular, the subfield of document theory, which examines the nature and function of documents in various contexts, is now contrasted with the seemingly autopoietic nature of large language models. It is this tension that I would like to explore during my sabbatical. As a research project it will examine three areas:

- How do we understand LLM query responses as a new form of document? Is it possible to understand these responses as documents through a new lens concerning the velocity of lexia, and how does that inform our understanding of the preservation of documents.
- Examining autopoiesis, or the apparent self organizing aspect of LLM, through materialist analysis. This will illuminate the value of generated documentation through LLM and interrogate critical blindspots in how we assess LLM outputs.
- Evaluate the impacts of LLM on how libraries manage records. In particular I will examine how LLM can inform new document types, and the implications for storing, retrieving, and contextualizing them.

This examination will lead to a peer-reviewed research article.

I am also a recipient of a University Libraries Collaborative Research Grant. The objective of this project is to catalog, archive, and make discoverable segments of marigram records on which are recorded measurements of historical tsunami events. NOAA National Centers for Environmental Information (NCEI) has more than 3,700 tsunami marigram records in both image and paper format, capturing worldwide observations of more than 390 tsunami events from 1854 to 1994. The majority of these tsunami marigram records were scanned to high-resolution digital tiff images during the NOAA Climate Data Modernization Program (CDMP) which ran from 2000 to 2011. For many tsunami events prior to 1994, data resides only on the marigram records, making them of great historical significance. Six of the 13 uncatalogued microfilm rolls have been scanned by NCEI to produce 3,548 TIFF images (about 5 GB). The proposed effort will catalog these TIFF images, identify any duplicates by comparing to the existing catalog of marigrams, and answer the questions below. In addition, given the size of the collection, it would be desirable to develop an automated approach to harvest metadata from the images to aid in cataloging. I will be working on this grant with my research partner, Aaron Sweeney.

(2) Describe your work plan including all anticipated professional activities (i.e., where will you spend your sabbatical, what you will do, your work timeline, etc.).

My work will be done remotely during this time within the state of Colorado. The first two months will be dedicated to focused research on the algorithmic underpinnings of Large Language

Models. Additional research will be conducted into previous theoretical examinations of documents. Most of these research documents are readily available online and if not they can be readily obtained through interlibrary loan. In particular the *Journal of Documentation*, *Journal of Library Metadata, Theory, Culture, and Society*, *Computers and Computation*, *Journal of the Association* and *Signs and Society* are readily available through the Libraries journal subscriptions or through partner institutions. It is possible I will also request records from the Mundaneum, an archive dedicated to Paul Otlet's work, but that will be determined by the trajectory of my initial research. The remainder of the sabbatical will be used to write the article, as well as initiating any new projects that emerge from this research but which may be out of scope for the sabbatical period. My collaborative research grant will have started prior to my sabbatical and will continue through this period. I may travel to Boulder, Colorado to supervise parts of this grant but most of it will be done remotely.

(3) Describe how meeting your sabbatical objectives will benefit the academic, clinical, and/or pedagogical goals of your primary unit.

Continuing to examine the function of documents in a records management setting can illuminate new ways of orienting our work. This project can lead to a critical reassessment of our priorities and workflows, opening them up to refinement which make our resources more discoverable. This is a pivotal time as the increased adoption of Large Language Models is poised to change aspects of our workflows. Rather than blindly use these tools, or secure funding to purchase them, it is important that we be able to assess the documents they ultimately produce for bias, accuracy, and alignment with our own metadata practices. Fundamentally I see this examination as an opportunity to demystify Large Language Models and map them to similar disruptive technologies and how they evolved or were ultimately abandoned. To summarize, Large Language Models represent a new tool for library workflows. How we reconcile its use with our own priorities and standards is open for examination. By using the lens of document theory we can better assess the role of Large Language Models with other similarly significant technological developments.

As for my research grant, the objective of this project is to catalog, archive, and make discoverable segments of marigram records on which are recorded measurements of historical tsunami events. By extracting this data into a computational format, more researchers will be able to use this valuable historical data which will lead to a better understanding of this natural phenomenon. Specifically we will be closer to answer the following questions:

1. How many tsunami events are captured on the six microfilm rolls? What is the oldest and what is the most recent?
2. For each tsunami event identified in the collection, what water level stations are included?
3. How many water level stations are included in the collection of the microfilm rolls?
4. Do any of the microfilm rolls contain observations from any of the tsunami events specifically sought after by the researchers below?

(4) Describe how your sabbatical project will enhance the university's reputation.

Examining the relation between Artificial Intelligence and records management is an understudied area with substantial implications for how libraries operate within this new technological paradigm. The ability to assess the relation between the two is sorely needed before Artificial Intelligence simply becomes incorporated into how libraries conduct their mission. By examining this area critically the reputation of the university will be enhanced as one that can be both forward looking and innovative but also thoughtfully restrained when it comes to adopting new technologies.

Documentation studies have yet to substantively address the role of Large Language Models as it impacts document production and function. As such this positions the University of Colorado Boulder to be at the edge of a developing discourse that will have implications beyond just its role in the Library. By positioning the University of Colorado Boulder at the forefront of

[REDACTED]

this topic, it can have a decisive impact on how the questions that arise from this new technology develop. As a public university it is in our mission to examine these questions on the behalf of the public and to ensure that our adoption of these tools aligns with the needs of Coloradans.

As for my research grant, it is hoped that this project will create the groundwork for additional partnerships between the University of Colorado and NOAA. Such relations exist but this work may open up additional research pathways and a sharing of resources towards new scientific outputs.

(5) Describe how your sabbatical will contribute to the educational experience of students.

This study will allow the libraries to assess certain record management workflows in light of a new technology. Ultimately this can aid students in making resources more discoverable and creating better access to resources. It also positions the libraries as a place where AI can be thoughtfully incorporated in a way that supports student learning, furthering the libraries' strategic initiative to seamlessly connect users and information. In particular as student data gets fed into various services to train Large Language Models, being able to articulate how this happens and how the Library and University are carefully allowing or preventing it will be essential to building trust with students.

Concerning my research grant, this will improve the student experience, particularly for those in geological, atmospheric, and related departments by making this data more widely available. This grant will also hire a student to develop code, giving them a greater insight into both NOAA/CIRES and Library practices.

Part 2:

REMUNERATION AND FUNDING PLAN

Regarding external funding, faculty members applying for sabbatical assignments are expected to apply for external funding (such as fellowships, grants, or clinical work) when appropriate. The total university salary to the faculty member, from sabbatical pay and any contract or grant administered through the university, shall not exceed university limits. There is no restriction on additional non-university income, subject to the faculty member satisfying the duties of the sabbatical plan and any contract/grant requirements. If a faculty member on sabbatical anticipates funding, sponsorship, employment, gifts, non-financial support, or other benefits from foreign institutions or sources, these should be detailed in the sabbatical remuneration plan, and all appropriate Export Control procedures should be followed. In addition, faculty members on sabbatical leave are not permitted to be paid for any administrative appointments or extra teaching during the sabbatical period.

- (6) Per CU System APS 1024, it is expected that faculty members applying for sabbatical will also apply for **external** funding to the extent it is available. Describe any anticipated external funding sources, amount of funding from sources external to CU, and attempts to obtain such funding.

I have not applied for funding for this project. I could not find relevant funding sources, specifically to support archival research at the Mundaneum in Belgium, to support this project. Ultimately it is unlikely any funding would be needed as I can request records from them if needed.

- (7) Describe the source and amount of any additional funding to support your sabbatical, including departmental or gift funding.

I will be working on a University Libraries Collaborative Research Grant during this time. The award is for [REDACTED]

- (8) Describe any anticipated support from foreign or international entities, including research collaborators, host universities, or other institutions, (e.g., equipment use, office/lab space, lodging or travel).

None.

- (9) Under the University's [APS 1024](#), faculty must identify business expenses to be reimbursed in connection with a Sabbatical Plan. Please describe anticipated business expenses and the funding source for those expenses.

None.

- (10) Will this plan require international travel? If so, please describe the travel including anticipated destination(s).

No.

Part 3:
DEPARTMENT CHAIR/UNIT HEAD
PLAN FOR COVERAGE AND REMUNERATION

Based on the corresponding CU System [APS 1024](#), the dean of the school/college shall ensure that the costs associated with the sabbatical are covered, including teaching replacement expenses. The dean may suspend a sabbatical if funding is not available in the school/college. Remuneration (from university resources such as state funding, university administered grants or contracts, or any other university managed sources) for the sabbatical assignment shall be as follows: for full-time faculty on nine-month appointments, either full salary for one semester or half salary for two semesters; for full-time faculty on 12-month appointments, six months full salary or 12 months half salary. Please note that remuneration from university-managed funds or university-administered grants or contracts should not exceed 100% of the faculty member's base salary in the case of full-pay sabbaticals, or 50% in the case of half-pay sabbaticals. For two-semester sabbaticals, remuneration sourced from General Funds (Fund 10) should not be used to supplement a half-salary. Stipends for administrative duties, such as chair or center director stipends, are not included in "base salary" and shall not be taken into account in calculating the individual's salary while on sabbatical.

- (11) Describe the plan for coverage of the faculty member's teaching responsibilities and replacement of teaching expenses. If it is known at the time this application is being made, include specific course names or numbers to be replaced each semester.

responsibilities will be addressed as described below.

Directs and coordinates the operations of the Metadata Production and Curation (MPC) Section.
[REDACTED], will work with the Digitization, Description, and Discovery Services Team Director and the Assistant Dean of Innovation and Research Strategies to coordinate the activities of the section and ensure it meets the needs of the Libraries.

Provides leadership for metadata practitioners in conceiving, designing, planning, and implementing appropriate metadata schemes and innovative metadata projects to facilitate user discovery of library resources.

During this period [REDACTED] will serve as [REDACTED] to the Colorado State Library and Digital Public Library of America. [REDACTED] will facilitate transfer of metadata to them on their quarterly cycle. [REDACTED] will also coordinate with the lead of the Metadata Optimization and Discovery section on any discovery issues that may arise. Project Proposals for the CU Digital Library will be reviewed by MPC personnel in addition to colleagues from the Digital Asset Management and Production Services section.

[REDACTED] a co-recipient of a University Libraries Collaborative Research Grant. [REDACTED] will work on this grant during the sabbatical with [REDACTED]

Analyzes metadata use cases and requirements to develop strategies for metadata capture, data quality, quality control, normalization, transformation, storage, and discoverability.

Members of the Metadata Curation and Production (MPC) Section will work collaboratively to determine any adjustments needed to workflows to ensure these requirements are met. When needed they will consult with Resource Description Services personnel to

[REDACTED]

ensure strategies are aligned. Geolibrary's workflows are currently under revision but any ingests will be covered by members of MPC.

Participates in planning and executing data migrations between information systems.

Should such a migration take place during the sabbatical period [REDACTED] [REDACTED] work collaboratively with members of the Libraries' Digital Asset Management and Production Section and the Core Technology and Applications Team to plan and execute the migration.

Identifies, monitors, and engages in areas of interest in metadata development (e.g., BIBFRAME, linked data, machine learning), particularly focusing on institutional participation.

Members of MPC will monitor these trends. They will be encouraged to apply for funding to attend conferences such as the Digital Library Federation Forum to learn more about metadata developments during this period.

- (12) If the applicant is in an additional administrative position (e.g., chair or director), indicate how the administrative responsibilities will be covered.

[REDACTED] will manage leadership of the Metadata Curation and Production Section on an interim basis.