**Fall 2017 Sontag Center *Course Activity Grant* Application Form**

Please answer the following questions so that we may evaluate your proposal for a Sontag Center ***course activity grant***. We recommend limiting your proposal to two pages, excluding the top section. If you have any questions about how to apply, please contact Linda Shimoda, Fred Leichter, or a member of our steering committee (listed at <http://creativity.claremont.edu/team-3/>). Proposals are due **Friday, April 28, 2017**.

**Name(s) with email addresses:**Madelynn Dickerson, [madelynn\_dickerson@cuc.claremont.edu](mailto:madelynn_dickerson@cuc.claremont.edu)  
Ashley Sanders, [ashley\_sanders@cuc.claremont.edu](mailto:ashley_sanders@cuc.claremont.edu)   
Lydia Bello, [lydia\_bello@cuc.claremont.edu](mailto:lydia_bello@cuc.claremont.edu)  
Margaret Hogarth, [margaret\_hogarth@cuc.claremont.edu](mailto:margaret_hogarth@cuc.claremont.edu)  
Mariam Salloum, [mariam.saloum@claremontmckenna.edu](mailto:mariam.saloum@claremontmckenna.edu)

**College(s):** Madelynn Dickerson, Lydia Bello, and Margaret Hogarth are librarians at the Claremont Colleges Library; Ashley Sanders is Director of the Digital Research Studio (Claremont University Consortium) and faculty at Claremont Graduate University, Mariam Salloum is faculty at Claremont McKenna College.

**Department(s):** Claremont Colleges Library, Digital Research Studio, Computer Science Dept at CMC

**Course Number(s):** CMC CSCI 181: Big Data Platforms/Applications

**Course Description(s):** This semester, the topic will be "Big Data: Platforms and Applications:" This course will introduce fundamental platforms and tools currently available to collect, integrate, store and analyze Big Data. The course will also focus on data mining and machine learning algorithms for analyzing Big Data. The course will include a practical application component in which you will learn how to setup and operate the Amazon Web Services (AWS) platform to perform analysis on a large dataset. The course material will be drawn from textbooks as well as recent research literature. Topics include: Hadoop/MapReduce Environment, Stream Processing of Twitter feeds (using Spark), large scale supervised machine learning (clustering, classification and recommender systems), NoSQL databases, and analysis & visualization tools (from CMC Schedule of Classes).

1. **Please describe how you will use support (including money, space, and/or staff time) from the Sontag Center to transform elements of your course:** Students in CMC CSCI will learn approaches to managing and processing Big Data. The Claremont Colleges Library has recently embarked on its own pilot work to reconsider library collections as Big Data, and to use digital humanities tools and methodologies to analyze these collections data. Our goal is for librarians, DH faculty and staff, and computer science students to work together to build infrastructure necessary to effectively analyze and communicate the library’s large datasets for current and future research on collections. We are seeking support from the Sontag Center to facilitate discussion and active learning opportunities for the CSCI students to experience working with library data, to introduce students to theoretical concepts behind “collections as data” through guest speakers, and money for infrastructure in building a secure and stable database for the data.
2. **Please describe how the course element(s) will support the following core principles of the Sontag Center:**
   1. **Feature experiential learning**

Students will work hands on with the librarians and each other to build a functioning database that librarians and researchers will be able to use for text analysis and other experimental methodologies on library collections data.

* 1. **Bring together students and/or faculty from multiple disciplines or campuses**This project will bring together librarians working in two different library divisions “research, teaching and learning” and “information resources and systems,” DH faculty and staff currently part of the Mellon-funded Digital Research Studio, CMC computer science faculty, and students from across the 5Cs who are enrolled in CMC CSCI 181.
  2. **Involve substantial collaborative activity among students**

At the heart of CSCI 181 is a practical application component designed to give students hands-on experience building the infrastructure needed to process and analyze large datasets. Students will need to work together to tackle large, complex data structures and work directly with librarians and DH faculty and staff to learn about and refine the library collections data they will be working with.

* 1. **Present students with ambitious challenges**Library collections as data is a relatively new concept and only recently beginning to emerge on a wider scale in data research practice. Students working on this project will be at the forefront of new thinking on “collections as data” and will be working with library data in a way that no one at the Claremont Colleges (or in many other places) has ever done before. It will provide students with an opportunity to engage with Big Data through an academic, non-consumerist framework.
  2. **Support a learning environment of curiosity, openness, flexibility, and experimentation**

For the past year, librarians have been working on a pilot study to apply DH methodologies to analyze and understand library collections data. Through this, we have recently submitted a journal article on the ways in which a culture of experimentation and collaboration has contributed to the success of our work. It has not been easy! Working with Big Data presents enormous conceptual, technical and practical challenges. We are committed to continuing to build on our existing culture of experimentation and mutual support as we work with students in the next phase of this research. We will necessarily remain flexible. The very principles behind “collections as data” are still in beta, as can be seen in the ongoing work of the “Always Already Computation – Collections as Data” working group based in Santa Barbara, who are currently developing a statement on this topic that we will be drawing on in this class. FMI <https://collectionsasdata.github.io/statement/>.

* 1. **Allow for experiences that are not possible in the current course design**This project would be a new component of CSCI 181 and bring in library collections and information science principles not previously part of the course design. The project also facilitates a non-traditional student-librarian relationship who will be engaged as researchers with a common goal.

1. **What do you (individually or collectively, if applying with a team) want to learn from trying this approach in your course(s)?**

The librarians are eager to learn from the students about building data infrastructure and work collaboratively with computer scientists, DH faculty and staff and others to promote and explore concepts and skills in data analytics, DH methods and tools around text analysis and data visualization. This approach will also provide students real-world project that will facilitate future research by librarians and other scholars going forward.

1. **If there is any additional information you would like to provide, please do so here:**As part of the pilot study completed last year by librarians, a project team built a small test database to analyze historical library collection patterns from the past 20 years (1995 – 2005), specifically on the topic of terrorism.Our group experimented with DH tools such as Voyant Tools, TimelineJS, and Tableau to answer questions about how our historic collecting patterns aligned to changes in the national political landscape. The project we will be working on with students will take this pilot study and innovate with new strategies to formalize and stabilize our datasets for future research. For more information about the earlier study, see <http://madelynndickerson.wixsite.com/dh4collections>
2. **Please include a preliminary budget for the proposed funds:**

Guest facilitator for a workshop on “collections as data” and/or interface design - $800  
 Server equipment or cloud storage - $500

1. **Please describe non-monetary support you are requesting, including but not limited to space and staff time:**

Staff and space and space in the Hive for one design-thinking session to introduce students to broad collections data concepts and brainstorm functionality of proposed database for future research.

1. **Please let us know of any other funding sources you have or are applying for to support this course or course element:**

We have not applied to any other funding at this time.