Piloting an Online National Collaborative Network for Integrating Computational Thinking into Library and Archival Education and Practice

The University of Maryland seeks a two-year $299,996 **Laura Bush 21st Century Librarian Program (LB21)** *National Digital Infrastructures and Initiative (NDII)* Project grant to pilot an online national collaborative network of educators and practitioners to enable the sharing and dissemination of computational case studies and lesson plans through a Jupyter Notebook interactive computational learning platform. We focus on *Master's*-level education in order to target professional development of future practitioners across the nation. We intend to enhance education and training programs related to "Collections as Data," an approach predicated on the computational use of cultural heritage collections. We used our IMLS-funded Symposium grant to test the impact of teaching computational thinking skills to iSchool students in the fall of 2019 using the Japanese American WWII Incarceration Camps collection. Konrad Aderer, the documentary filmmaker who observed the student work, commented that "the students' [computational work] uncovered significant incidents previously unknown, synthesized data to bring out meaningful themes that light the way to further study, and contributed to the understanding of Japanese American history."

This virtual network is really a network of networks with seventeen collaborators dedicated to mapping Computational Thinking to Archival and Library practices. This Network includes: (1) a *Core Network (CN)* of seven experts in digital archives, lesson plan evaluation, project management, computational thinking, library software integration, and ethics and representation in digital collections, (2) an *Educator Network (EN)* of four educators from MLIS programs (at all ranks), and (3) a *Practitioner Network (PN)* of seven librarians / archivists representing four diverse and under-represented American collections of African, Asian, and Puerto Rican - American lineage: (a) the Maryland State Archives *Legacy of Slavery Project*, (b) the Spelman College *Department of Drama and Dance Photographs*, (c) Densho's *WWII Japanese American Camps Collections*, and (d) the 2019 *Puerto Rican Summer Protests ("RickyRenuncia")*. We are calling this cluster of *Practitioner Network* collections ***"Re-presenting America,"*** to emphasize its significance and impact of training future MLIS students and exposing them to the full diversity of the American experience. In addition, we will seek feedback from an *Advisory Network (AN)* consisting of: (1) five US experts [*three Practitioners* at Cultural Institutions: Smithsonian National Museum of American History, Harvard Library, the US Holocaust Memorial Museum, and *two iSchool Educators* from UCLA and Drexel], and (2) International experts from all six continents, many of whom attended our April 2019 IMLS Symposium workshop and who are eager to contribute.

We will work to promote digital interpretation and stewardship of *Re-presenting America*, which will form the basis of lesson plans tested and taught in the classroom across four MLIS programs with faculty offering to pilot the Jupyter Notebook lesson plans in their Spring 2022 classes. In addition, a core set of Notebooks will be tested in their Spring and Fall 2021 classes. New digital treatments and online access may raise unforeseen cultural sensitivity and privacy issues. Critical evaluation of these potential ethical concerns is a core learning objective. We seek to change the prevailing state of reticence to enroll in digital courses by our MLIS students and lack of faculty being able to teach courses needed to prepare future digital collection practitioners, by creating a supportive community of teachers and practitioners, dedicated to modernizing archival and library education. Our ultimate goal is to contribute to the development of faculty and library "digital leaders".

1. Statement of Broad Need

In April 2019, through a *Laura Bush 21st Century Librarian Program (LB21)* Symposium Grant (RE-73-18-0105-18) entitled *Developing a Computational Framework for Library and Archival Education*, we convened a panel of 50 international experts (educators, practitioners, and students), and met at the University of Maryland and at the Smithsonian National Museum of Natural History for two days, to reflect on the changing needs of Library and Archival Science education.

Participants emphasized the urgent need to better prepare students and professionals to work in an increasingly digital and even computational landscape. "The use of emergent technologies have profoundly altered the nature

of archives, by disrupting how information is created, recorded, captured, encoded, curated, shared, made available and used" (E. Goudarouli, 2019). These changes are affecting libraries and archives of all sizes. This is apparent in the decision of the National Archives and Records Administration (NARA) to stop accepting paper-based records at the end of 2022 (Fedscoop, 2019). Similarly, at the Library of Congress (LoC) digital strategies are being piloted based on how "people are using computers to do computational analysis". LoC wants to enable computational research by making sure their collections are available in machine-readable ways, increasingly providing access to materials through APIs and other bulk data interfaces (GovLoop, 2019). In the IMLS *"Collections as Data"* project *(LG-73-16-0096-16),* examples of computational treatments of collections include scenarios where "a Digital Humanities researcher engages in term frequency visualization, topic modeling, and network analysis across thousands and sometimes even millions of items." Beyond text data "the scope of data extends to images, moving images, sound, web archives, and beyond." These types of digital scholarship activities are increasingly showing up in job descriptions for positions in libraries and archives.

IMLS recently funded two Computational Thinking (*CT*) grants to facilitate the development of critical literacies applicable to STEM learning *(LG-14-19-0079-19 U. Maryland w. PI Subramaniam and RE-12-19-0094-19 CUNY w. PI Sanchez)*. *CT* is described as a form of problem solving that uses, modeling, decomposition, pattern recognition, abstraction, algorithm design, and scale (Wing, 2006). **The argument for integrating *CT* into library and archival science parallels the case for its inclusion in mathematics and science classrooms.** It is motivated by this emerging digital and computational landscape of library and archival work. For today's learners to succeed in future library and archival tasks, it is essential that *CT* be included as part of their training (W. Underwood 2018).

The current situation is one where digital and computational literacy for future librarians and archivists as taught through MLIS programs (Master's of Library and Information Science) is seen as broadly lacking and uneven at best. At the University of Maryland iSchool, members of the *Core Network* are teaching a professional certificate program in Digital Curation for Information Professionals (DCIP) and many of the participants are recent graduates of MLIS programs who upon entering the job market are realizing the urgent need for deeper technical training. The challenges are compounded by the lack of digital training of faculty themselves and even resistance from colleagues and administrators. Given this context, it makes sense to position our project, in terms of its phase of maturity, as a **community pilot project** that brings together iSchool faculty and practitioners in the field, wishing to build capacity through collaboration.

2. Project Design

2.1. Project Objectives

The University of Maryland seeks a 2-year Laura Bush 21st Century Librarian (LB21) National Digital Infrastructures and Initiatives (NDII) Project Grant to pilot an online national collaborative network of educators and practitioners to enable the sharing and dissemination of computational case studies and lesson plans through a Jupyter Notebook interactive computational learning platform. "Jupyter is a free, open-source, interactive web tool known as a computational notebook, which researchers can use to combine software code, computational output, explanatory text and multimedia resources in a single document." (Int Journal of Science, 2018). The goal is to change the current MLIS-level educational culture and build a community network that supports educators and practitioners in training the next generation of library and archives leaders.

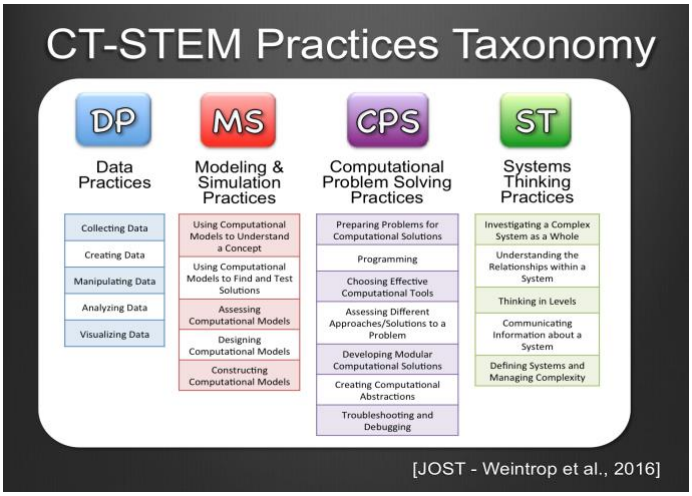
2.2. Background and Motivation

The IMLS Symposium grant we recently conducted, which provides the motivation for our current proposal (see final report), established the following results:

(1) **A demonstration that the twenty-two Computational Thinking (*CT*) practices** that have been identified as important in STEM education also occurred and were essential for performing library and archival practices when addressing digital records. These practices are organized into 4 categories: *Data Practices (DP), Modeling & Simulation Practices (MS), Computational Problem Solving Practices (CPS),* and *Systems Thinking Practices (ST*).

(2) **Jupyter Notebooks** were a useful tool for recording and organizing the tasks, data, tools and results of performing library and archival tasks on digital records.

(3) **An on-line repository called CASES** (Computational Archival Science Educational System) was constructed for storing and providing access to Jupyter notebooks that recorded the results of performing library and archival tasks on digital records.



(4) **A taxonomy of archival topics**, practices and learning outcomes was derived from Library and Archival Studies and Digital Curation courses taught in the MLIS program of the University of Maryland. This taxonomy is to be used in indexing the cases and computationally enhanced lesson plans in the CASES repository to support search and access to lesson plans relevant to Library and Archival courses. A decision to augment the taxonomy with ethics attributes was reached.

(5) **A two-day symposium** was conducted at the U of Maryland in April 2019 in conjunction with the 2019 iConference in which the results of this research was presented to some 50 national and international experts in "Collections as Data".

The idea for an online national collaborative network pilot proposal originated from feedback provided by educators and practitioners from a broad set of venues including the our IMLS Symposium, the Best Practices Exchange 2019 conference, the AERI2019 Conference in the UK, and the Archival Educators Section at the 2019 Society of American Archivists (SAA) annual conference. The need was further expressed at the 2019 IEEE Big Data International Workshop on Computational Archival Science (CAS), and the January 20, 2020 Alan Turing Institute on CAS.

2.3. Project Organization

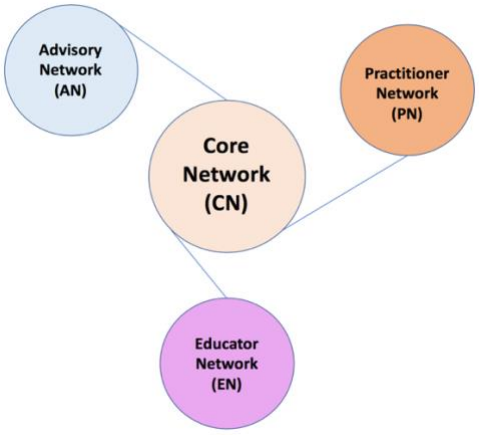
We introduce the four networks our national community network is composed of: a Core Network (CN), an Educator Network (EN), a Practitioner Network (PN), and an Advisory Network (AN). Our strategy is to bring together educators and practitioners around the use of Jupyter Notebooks *(ex. at **Supportingdoc1.pdf**):*

- **Educators** are rapidly adopting Jupyter Notebooks for teaching, use in the classroom, developing teaching materials, and creating computational stories. These are in particular inspired by the 18th and 19th century paper-based lab notebooks that were used in science research and that captured observations, experiments, ideas, notes, formulas, and data (these were used in most of the 20th century as well).
- **Practitioners** are also starting to produce Jupyter Notebooks. See in particular the work of Linda Sellars and Emily Higgs on "Implementing Named Entity Recognition in Description of Born-Digital Materials," which uses Text Processing / Named Entity Recognition (NER) techniques in automating the description of born-digital materials and is shared publicly in the form of an online Notebook. Other examples include the work of Laura Wrubel (Software Development Librarian at GWU) on "Analyzing color clusters in images in Library of Congress collections," through the production of a Jupyter Notebook tutorial.

We propose to pilot a national collaborative network that integrates Educator and Practitioner-driven endeavors and seeks to promote the creation of real-world and meaningful lesson plans to educate current MLIS students.

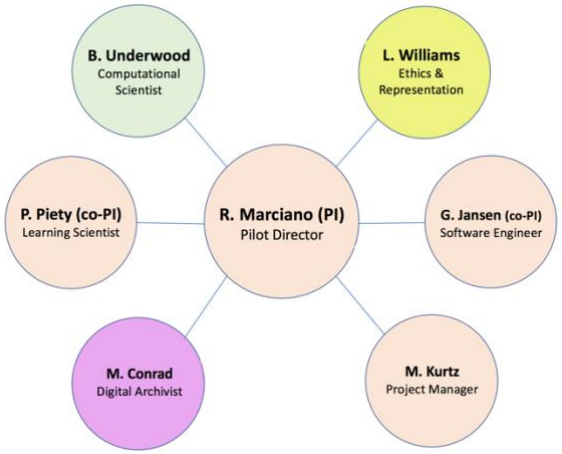
In preparation for this national network, on February 28, 2020 we decided to pro-actively launch a virtual organization that will support this national pilot. We are calling it the Advanced Information Collaboratory (**AIC**). The **AIC** was devised as part of our IMLS-funded Symposium grant and was put in motion at a recent workshop at The Alan Turing Institute at the British Library on January 20, 2020. The **AIC** builds on four years of CAS community development, projects, and funded grants from the Institute of Museum and Library Services, the National Science Foundation, and the National Park Service in particular. The **AIC** brings together the partners referenced in this proposal from leading academic and cultural institutions spanning six continents. The primary focus of the **AIC** is technical, collaborative, educational, and library and archives-centric.

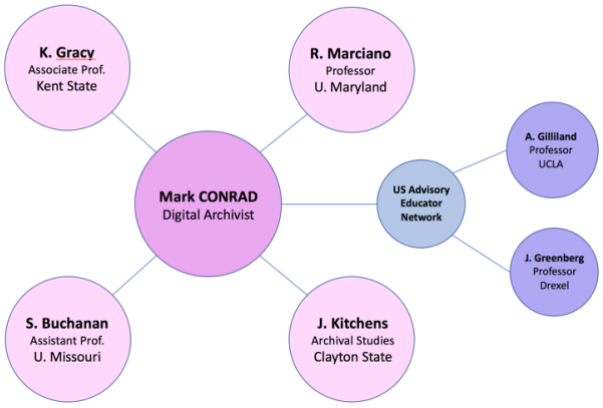
With this Collaboratory in place, our project is designed as a network of interacting networks including: (1) a *Core Network* of experts from the University of Maryland iSchool supporting (2) an *Educator Network*, (3) a *Practitioner Network*, and receiving feedback from (4) a national and international *Advisory Network*.



*Our project is directed by: (1) a **Core Network (CN)** from the University of Maryland iSchool, interacting with three networks: (2) **Educator Network (EN)**, (3) **Practitioner Network (PN)**, and (4) **Advisory Network (AN)**.*

*(1) The **Core Network (CN) at U. Maryland** includes 7 experts in project management, lesson plan evaluation, digital archives, computational thinking, software integration, and ethics and representation.*

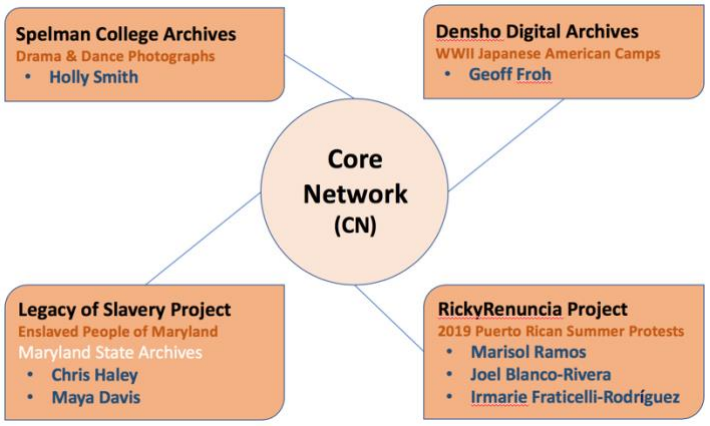


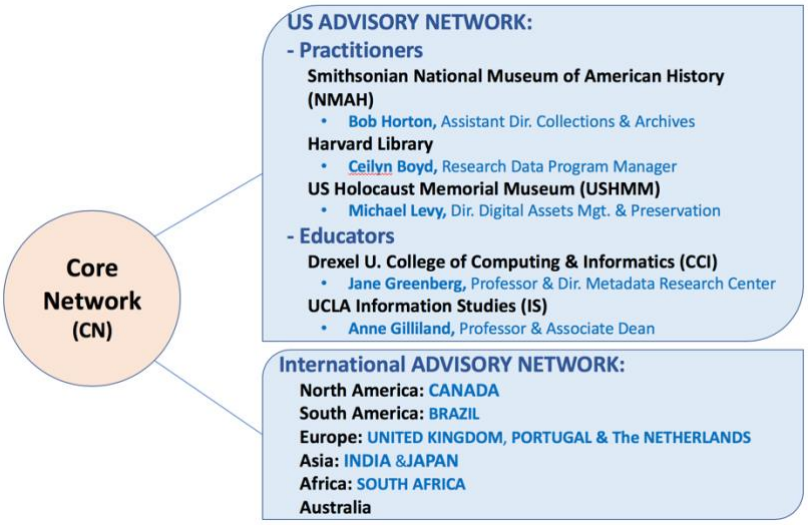


*(2) The **Educator Network (EN)** includes 4 archival and library educators from iSchools and professional programs: Kent State U., U. of Missouri, U. Maryland, and Clayton State U. (with faculty at all ranks). Two senior educators from the Advisory Network (AN), from Drexel U. and UCLA, will provide oversight.*

*(3) The **Practitioner Network (PN)** features underrepresented American collections of African American, Japanese American, and Puerto Rican content, from: (1) the Maryland State Archives Legacy of Slavery Project, (2) the Spelman College Department of Drama and Dance Photographs, (3) Densho's WWII Japanese American Camps, and (4) the 2019 Puerto Rican Summer Protests.*

We are calling this cluster of collections the "Re-presenting America", to emphasize its foremost significance and impact in training future MLIS students and exposing them to the full diversity of the American experience.





*(4) The **Advisory Network (AN)** consists of three Practitioners at US institutions (Smithsonian National Museum of American History, Harvard Library, and the US Holocaust Memorial Museum), two Educators at US institutions (UCLA and Drexel), and an International body of partners from all six continents, many of whom attended our April 2019 IMLS Symposium workshop and have strong interest in mapping Computational Thinking to Archival and Library practices.*

2.4. Project Activities

Development of Computationally Enhanced Lesson Plans (T1.1)

Lesson plans, or educational units of several lesson plans *(see **Supportingdoc2.pdf** for an CT-STEM example)*, will be developed within key Archives & Libraries, Knowledge Areas and Computational Thinking Practices, in collaboration / conversation with the Instructors. Computational archival and library research cases in Jupyter notebooks will serve as the basis for computationally enhanced lesson plans, also in Jupyter notebooks. These Lesson plans will include a teacher's guide, links to computational tools, examples of library and archival tasks that exhibited CT, and exercises that require the use of CT skills. The idea behind lesson plans and units is that they are convenient educational building blocks that can be adopted in a classroom setting at various levels (within a class, for use in a hands-on lab, across multiple classes, or even to construct an entire syllabus). A lesson plan can be constructed as a 20-minute, 45-minute, or even 90-minute class., and a unit could be used for 3 to 4 class segments for example. We will explore how to size these segments for MLIS teaching, in an iterative way, and by collecting feedback from the classroom.

An appropriate notebook-based course for elementary Python programming will be identified. If necessary, we will adapt this course to our MLIS Student audience. The following are a possible initial set of Jupyter Notebook-based lesson plans:

Lesson Plan or Unit	MLIS Course Knowledge Areas
1. Auto-categorization of Email	Records Management, Computational Thinking
2.Preserving Records in a Java ARchive (JAR)	Digital Preservation, Records Management
3.Auto-extraction of Metadata	Archival Description, Indexing and Retrieval
4.Analysis of Historical City Directories	Digital Curation, Archival Science, Digital Humanities
5.Python Programming	Python Programming for Library and Archival Scientists

We have already developed case studies for all five of these topics but not lesson plans yet. For example, topic 4. is being tested in the Spring 2020 DCIP professional certificate at the U. Maryland and we have identified a Unit of 7 lesson plans including: (1) Digitization management of early 1900 City Directories, (2) Cleaning and transforming data, (3) Data wrangling in the Cloud, (4) Text Processing, (5) Geospatial

transformations, (6) Data visualization, and (7) Building networks. We will leverage this earlier work to ensure that an initial set of Lesson Plans will be ready for use in classes in the Spring and Fall of 2021.

Development of Assessment Tools (T1.2 & T3.3)

As we indicated earlier in this proposal, we believe that capabilities of Computational Thinking (*CT*) are essential to Library and Archival practice in the 21st century. However, some students in MLIS programs are hesitant to enroll in courses that involve programming or computational methods. Also, some faculty are doubtful of the importance of computational thinking in the MLIS curriculum and may lack the knowledge to teach computational thinking.

We plan to develop *CT* enhanced lesson plans that faculty will use to enhance required core MLIS courses that they teach. Faculty teaching these courses will be involved in the refinement of the units or lesson plans that include computational thinking practices.

It is planned to develop and conduct pre-course and post-course assessments of student and faculty knowledge of computational thinking practices as they relate to library and archival topics and practices. Pre-course and post-course attitudes of students and faculty toward computational thinking practices will also be assessed.

Jupyter Notebook Training (T1.3)

Train EN and PN in use of Jupyter Notebooks, the MyBinder service (which allows the virtual execution of and interaction with Jupyter Notebooks in the cloud without having to install any software), and recommend the elementary Python online learning resource.

Lesson Plans Used in Classes (T2.2 & T4.2)

The 4 members of the **Educator Network (EN)** have offered to pilot the Jupyter Notebook lesson plans in their Spring 2021, Fall 2021, and Spring 2022 classes in the following manner:

Educator Network	Spring 2021 & Spring 2022	Fall 2021	MLIS Course Knowledge Areas
Clayton State U. Archival Studies Program	• ARST 5110: Archives and the Web	• ARST 5300: Digital Preservation	• Preservation
Kent State U. School of Information	• LIS 60633: Digital Curation • LIS 61095: Intro. to Digital Humanities	• LIS 60631: Introduction to Digital Preservation • LIS 60652: Foundations of Recordkeeping	• Digital Curation • Digital Preservation • Records Management • Digital Humanities
U. Missouri School of Information Science & Learning Technologies	• ISLT 9492: Data & Records Management	• ISLT 9490: Archival Practice	• Records Management • Archival Science
U. Maryland College of Information Studies	• INST Topics: Computational Thinking for MLIS Students • INST 747: Research in Advanced Digital Curation	• INST448: Digital Curation Research in Cultural Big Data Collections • INST 747: Research in Advanced Digital Curation	• Digital Curation • Computational Thinking

The initial lesson plans we intend to have ready by Spring 2021 are described above in the "Development of Computationally Enhanced Lesson Plans (T1.1)" section. These will be both units and lesson plans and are intended to be 45 to 50 minutes in length.

Lesson plans will be refined based on feedback, so educators are partners in the creation of lesson plans. Educators will revise the notebooks and lesson plans to improve their learning impact and student engagement, based on classroom experiences and teaching expertise.

The ***Educator Advisory Panel***, composed of senior educators from Drexel U. and UCLA (resp. Dr. Jane Greenberg and Dr. Anne Gilliland) will provide further feedback on opportunities to scale these lesson plans to other universities and programs. Relevant classes in those programs include:

- At Drexel U.: (1) INFO560 Introduction to Archives I, (2) INFO561: Introduction to Archives II, (3) INFO590 Foundations of Data and Information, (4) INFO591: Data and Digital Stewardship, (5) INFO750 Archival Access Systems, (6) INFO755 Electronic Records Management, (7) INFO662 Metadata for Resource Description, and (8) INFO676 Applied Ontology.
- At UCLA: (1) IS241 Digital Preservation, (2) IS282 Management of Digital Records, and (3) IS438A Archival Appraisal.

It is both deliberate and significant that the members of the ***Educator Network*** are of all tenure ranks (assistant, associate, full, and professional) and advised by senior faculty. Too often, the tenure process gets in the way of taking pedagogical risks. Our Jupyter Notebooks are meant to support taking innovation in pedagogy.

Collaborative Digital Curation of Library and Archival Collections (T3.1)

The Core Network (CN) will work with the Practitioner Network (PN) to develop digital projects that will lead to the development of a second set of Jupyter Notebooks around the theme of **"Re-presenting America"** (see *Diversity Plan* section for more details). Those projects will address records from African American, Japanese American, and Puerto Rican content, from: (1) the Maryland State Archives *Legacy of Slavery Project*, (2) the Spelman College *Department of Drama and Dance Photographs*, (3) Densho's *WWII Japanese American Camps*, and (4) the *2019 Puerto Rican Spring Project*. The PN case study collections will be reviewed and brainstorm about the challenges and opportunities with respect to computational treatments. The resulting notebooks will be added to the corpus of educational resources through the CASES (Computational Archival Science Educational System) Infrastructure.

Lesson Plans from Practitioner Case Studies (T3.2)

Computational archival research cases in Jupyter notebooks will be the basis for computationally enhanced lesson plans, also in Jupyter notebooks. These Lesson plans will include a teacher's guide, links to computational tools, examples of library and archival tasks that exhibited CT, and exercises that require the use of CT skills. Assessment tools similar to those developed earlier will also be developed for these lesson plans. If these lesson plans are developed before the spring semester of 2022, there may be an opportunity to incorporate them into MLIS courses taught by educators during that semester.

Student, Faculty and Practitioner Assessment (T2.1 / T2.3 & T4.1 / T4.3 & T5.1 & T5.3)

As we indicated earlier in this proposal, we believe that capabilities of computational thinking (*CT*) are essential to Library and Archival practice in the 21st century. However, some students in MLIS programs are hesitant to enroll in courses that involve programming or computational methods. Also, some faculty are doubtful of the importance of computational thinking in the MLIS curriculum and may lack the knowledge to teach computational thinking. Finally, there are librarians and archivists who are eager or willing to explore computational practices in the digital curation of their organization's collections.

We plan to develop CT enhanced lesson plans that faculty will use to enhance required core MLIS courses that they teach. Faculty teaching these courses will be involved in the refinement of the units or lesson plans that include computational thinking practices. We plan to work with practitioners in curating some of their collections using computational thinking practices. We also plan to introduce practitioners to computational thinking in a workshop.

It is planned to develop and conduct pre-course and post-course assessments of student, faculty and practitioner knowledge of computational thinking practices as they relate to library and archival topics and practices. Pre-course and post-course attitudes of students, faculty and practitioners toward computational thinking practices will also be assessed.

Computational Thinking Workshop for Practitioners (T5.2)

Practicing librarians and archivists also need to be able to integrate computational thinking into their management of their digital collections. While there is hesitation among some of these professions to consider learning anything computational, there are others who are willing or even eager to explore the explore the possibility via continuing education.

A workshop will be developed for the 7 members of that *Practitioner Network (PN),* that introduces them to computational thinking via some of the lesson plans developed during this project. The student assessment tools will be adapted to assess acquisition of knowledge and possible change of attitude toward computational library and archival methods. We will pursue conducting these workshops and the annual professional meetings of librarians and Archivists.

Maintain Computational Thinking Library and Archival Studies Website (T6)

New CASE lesson plans will be added to the CASES repository and website. A "notebook finder" feature will be added to the CASES website, based on Archives & Libraries course topics.

2.5. List of Key Project Staff and Consultants

Core Network (CN):

1. **Dr. Richard Marciano** is a professor in the College of Information Studies at the University of Maryland. His research interests center on digital curation, sustainable archives, cyberinfrastructure, and big data. He is also the 2017 recipient of Emmett Leahy Award for innovation in records and information management.
2. **Dr. William Underwood** is an Adjunct Research Scientist with the College of Information Studies at the University of Maryland. His current research interests are in records management and archival science, and the application of natural language processing to the support of automated archival description.
3. **Dr. Michael J. Kurtz** was the project manager on our IMLS Symposium grant and will continue in this role on this grant. Prior to this he worked at the U.S. National Archives and Records Administration for 37 years as a professional archivist, manager, and senior executive, retiring as Assistant Archivist in 2011.
4. **Dr. Philip Piety** is a Senior Lecturer and Learning Scientist at the iSchool at the University of Maryland. He is an expert in learning technologies and analytics and founder of the Maryland Education Digital Infrastructures and Analytics Lab (MEDIAL) that explores the tools that mediate pedagogy.
5. **Greg Jansen** is the AIC Collaboratory's Principal Software Engineer and builds data repositories with new capabilities for computation and analysis. He has a focus on creating traceable data curation workflows that connect new data or holdings to evidence, through open workflows and chains of provenance.
6. **Mark Conrad** is an archival practitioner, educator, and researcher. He retired from the National Archives and Records Administration in October 2019 after 28 years. He spent his entire time at NARA working with electronic records as an appraisal and accessioning archivist, program officer for electronic records grants, collaborator on research projects, and educator on electronic records issues
7. **Dr. Lyneise Williams** is Associate Professor of Art History at the University of North Carolina at Chapel Hill (PhD Yale 2004). She is the author of *Latin Blackness in Parisian Visual Culture, 1852-1932, (*February 2019, Bloomsbury Academic Publishers), which examines how Parisians' visual language of Latin Americans in popular imagery inextricably links blackness to Latin American identity beginning in the mid-nineteenth century and into the early twentieth century.

Educator Network (EN):

1. **Dr. Karen Gracy, [Kent State U. School of Information]** has developed and delivered coursework in digital preservation and curation, moving image archiving, preservation and conservation of cultural heritage materials, and archival description and representation.
2. **Dr. Joshua Kitchens [Director of the Clayton State Archival Studies program]** teaches a variety of courses at Clayton State that cover archival topics including Digital Preservation, Law and Ethics, Appraisal, and various special topics courses.

3. **Dr. Sarah Buchanan [Assistant Professor, Library and Information Science (LIS), School of Information Science & Learning Technologies (SISLT), U. Missouri]** studies the creation and curation of data in the humanities, focusing on the storytelling potential of archaeological collections. Being an archival educator, she is active in the Archival / Preservation Education ALISE SIG, in SAA CORDA and the Missouri Association for Museums & Archives, and in AERI and ASIS&T.
4. **Dr. Richard Marciano [UMD College of Information Studies]** is referenced earlier.

Practitioner Network (PN): See *Diversity Plan* section for more details.

1. **Maryland State Archives** Legacy of Slavery Project
   - **Chris Haley:** Director and Supervising Archivist
   - **Maya Davis:** Research Archivist and Legislative Liaison
2. **Spelman College Archives**, Black Archives Alliance
   - **Holly Smith**, College Archivist
3. **Densho Digital Archives** WWII Japanese American Incarceration Experience
   - **Geoff Froh**, Deputy Director
4. **RickyRenuncia "2019 Puerto Rican Summer Protests":**
   - **Marisol Ramos**, UConn Library, Digital Scholarship, Humanities/Social Sciences Librarian
   - **Dr. Joel Blanco-Rivera**, Adjunct Professor, National Institute of Anthropology and History. National School of Conservation, Restoration, and Museography (ENCRyM)
   - **Irmarie Fraticelli-Rodríguez**, Special Collections Librarian—Colección Puertorriqueña, University of Puerto Rico and MLIS grad student at the School of Information, and 2019-2021 ARL/SAA Mosaic Program Fellow

Advisory Network (AN): *many of whom attended our April 2019 IMLS Symposium workshop*

US Advisory Network (5 people):

Practitioners:

(1) SMITHSONIAN National Museum of American History (NMAH) [**Bob Horton,** Asst. Dir. Coll. & Archives]
(2) HARVARD Library [**Ceilyn Boyd,** Research Data Program Manager]
(3) US HOLOCAUST MEMORIAL MUSEUM (USHMM) [**Michael Levy,** Dir. Digital Assets Mgt. & Preservation].

Educators:

(1) Drexel U. College of Computing & Informatics (CCI) [**Dr. Jane Greenberg,** Prof & Dir. MRC]
(2) UCLA Information Studies (IS) [**Dr. Anne Gilliland,** Professor & Associate Dean].

International Advisory Network (18 people across 6 continents):

(1) NORTH AMERICA: Canada -- Blockchain@UBC [**Dr. Victoria Lemieux**, Associate Professor]
(2) SOUTH AMERICA: Brazil -- University Brasilia [**Dr. Claudio Gottschalg-Duquee**, Professor]
(3) EUROPE: United Kingdom -- King's College London [**Dr. Mark Hedges**, Reader in Cultural Informatics in the Digital Humanities Department], University College London [**Dr. Jenny Bunn**, Lecturer in Archives and Records Management], Oxford University [**Dr. David De Roure**, Professor of e-Research], The National Archives [**Pip Willcox**, Head of Research, **Dr. Eirini Goudarouli**, Head of Digital Research Programmes, **Mark Bell**, Big Data Researcher, **Paul Young**, Digital Preservation Specialist / Researcher, and **Sonia Ranade**, Head of Digital Archiving], The Alan Turing Institute [**David Beavan**, Senior Research Software Engineer], **Continent: The Netherlands** -- University Amsterdam [**Dr. Tobias Blanke**, University Professor of Humanities and AI], **Portugal** -- INESC-ID [**Dr. Diogo Proença**, eArchiving Project Researcher]
(4) AFRICA: South Africa -- University South Africa [**Dr. Shadrack Katuu**, Information Science]
(5) ASIA: India -- Central University of Gujurat [**Dr. Bhakti Gala**, Assistant Professor at the School of Library and Information Science], and Indian Institute of Management [**Dr. H. Anil Kumar**, Librarian of the Vikram Sarabhai Library (VSL)], **Japan** -- Kyushu University [**Dr. Yoichi Tomiura**, Deputy Director General of U. Libraries, Professor Department of Informatics, Professor Department of Library Science]
(6) AUSTRALIA -- University of Canberra [**Dr. Tim Sherratt**, Associate Professor of Digital Heritage]

3. Diversity Plan

Dr. Lyneise Williams, through her newly founded VERA Collaborative (Visual Electronic Representations in the Archive), will embed herself in this online collaborative network as part of the Core Network. VERA addresses the distortions and erasures in visual representations, linked to reproduction technology, that particularly impacts communities of color and other under-represented communities.

Based on reviewer feedback of our Preliminary Proposal, we specifically requested additional funds as part of the full proposal to develop a strong, coherent, and thematic **Practitioner Network** (PN), with a unique cluster of collections representing the full American experience. We are calling this the ***"Re-presenting America" Collection***, to emphasize its foremost significance and impact in training future MLIS students and exposing them to the full diversity of the American experience. The partners themselves bring under-represented American collections of African, Asian, and Puerto Rican -American lineage: (1) the Maryland State Archives *Legacy of Slavery Project*, (2) the Spelman College *Department of Drama and Dance Photographs*, (3) Densho's *WWII Japanese American Camps Collections*, and (4) the 2019 *Puerto Rican Summer Protests ("RickyRenuncia")*.

Representation will be a component of every aspect of this project as it touches on non-represented people in the archives: who gets represented and what are the aesthetics of representation? It will also bring forward under-recognized sources produced by and featuring diverse American populations. These sources are often sidelined in favor of a narrative of a white America. Integrating technological training into a story about America, encouraging users to internalize it as American, a part of a whole, is an important goal in developing future faculty and library leaders.

The archivists and librarians in this network involved are deeply connected with their collections and have received accolades and national recognition for the quality and diversity of their work:

- **Holly Smith** (Spelman College) was a speaker at the SAA 2019 Conference on *"Building Digital Capacity in HBCU Libraries through Collaboration".*
- **Marisol Ramos** (UConn) was the recipient of the *2019 SAA Diversity Award* for the Puerto Rico Citizenship Archives Projects (PRCAP)
- **Irmarie Fraticelli-Rodríguez** (U Michigan) is a *2019-2021 ARL/SAA Mosaic Program Fellow*.
- **Dr. Joel A. Blanco-Rivera** is the 2018 co-author of *"Puerto Rico's Archival Traditions in a Colonial Context."*
- **Chris Haley and Maya Davis** (Maryland State Archives Legacy of Slavery Project) organized a 2-day datathon on computational archival processing of slavery records October 28-29, 2019.
- **Geoff Froh** (Densho.org) was the keynote speaker at the 2018 IEEE International Big Data "Computational Archival Science" Workshop and presented on "Reclaiming our Story: Using Digital Archives to Preserve the History of WWII Japanese-American Incarceration"

4. Broad Impact

This project will pilot an online national collaborative network. It is expected that the approach will be unifying in nature: not only impacting the way MLIS Master's courses are taught but also naturalizing an image of a diverse America. We expect this pilot project to be beneficial to other teaching contexts as well (pre-professional, doctoral, early career development, and even continuing education). Our approach is based on a unifying framework (based on introducing Computational Thinking - *CT* - skills to Archivists and Librarians through the use of open-source Jupyter Notebooks) that allows case studies and lesson plans to be shared, re-purposed, and searched based on a common reference set of practices in three areas: libraries & archives, computational thinking, and ethics/values considerations. International partners interested in the integration of *CT* into library and archival education and practice will benefit by validating the approach. We seek to change the prevailing state of reticence to enroll in digital courses by our students and lack of faculty being able to teach courses needed to prepare future digital collection practitioners, by creating a supportive community of teachers and practitioners, dedicated to modernizing archival and library education.

Scheduleofcompletion.pdf

Sep. 1, 2020 – Aug. 31, 2022	NETWORKS			2020				2021												2022							
	CN	EN	PN	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A
Task 1: Teaching Preparation																											
T1.1: Develop Computationally Enhanced Lesson Plans																											
T1.2: Develop Assessment Tools																											
T1.3: Jupyter Notebook Training																											
Task 2: Fall/Spring 2021 Teaching																											
T2.1: Student Pre-assessment																											
T2.2: Lesson Plans used in Classes																											
T2.3: Student Post-assessment																											
Task 3: Working with Practitioners																											
T3.1: Research with Practitioner Collections																											
T3.2: Develop Computationally Enhanced Lesson Plans from Practitioner Cases																											
T3.3: Develop Assessment Tools																											
Task 4: Spring 2022 Teaching																											
T4.1: Student Pre-assessment																											
T4.2: Lesson Plans used in Classes																											
T4.3: Student Post-assessment																											
Task 5: Practitioner Workshop																											
T5.1: Practitioner Pre-assessment																											
T5.2: Workshop with Lesson Plans																											
T5.3: Practitioner Post-assessment																											
Task 6: CASES website																											
Task 7: Write Final Report																											

DIGITAL PRODUCT FORM

INTRODUCTION

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to digital products that are created using federal funds. This includes (1) digitized and born-digital content, resources, or assets; (2) software; and (3) research data (see below for more specific examples). Excluded are preliminary analyses, drafts of papers, plans for future research, peer-review assessments, and communications with colleagues.

The digital products you create with IMLS funding require effective stewardship to protect and enhance their value, and they should be freely and readily available for use and reuse by libraries, archives, museums, and the public. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

INSTRUCTIONS

If you propose to create digital products in the course of your IMLS-funded project, you must first provide answers to the questions in **SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS.** Then consider which of the following types of digital products you will create in your project, and complete each section of the form that is applicable.

SECTION II: DIGITAL CONTENT, RESOURCES, OR ASSETS

Complete this section if your project will create digital content, resources, or assets. These include both digitized and born-digital products created by individuals, project teams, or through community gatherings during your project. Examples include, but are not limited to, still images, audio files, moving images, microfilm, object inventories, object catalogs, artworks, books, posters, curricula, field books, maps, notebooks, scientific labels, metadata schema, charts, tables, drawings, workflows, and teacher toolkits. Your project may involve making these materials available through public or access-controlled websites, kiosks, or live or recorded programs.

SECTION III: SOFTWARE

Complete this section if your project will create software, including any source code, algorithms, applications, and digital tools plus the accompanying documentation created by you during your project.

SECTION IV: RESEARCH DATA

Complete this section if your project will create research data, including recorded factual information and supporting documentation, commonly accepted as relevant to validating research findings and to supporting scholarly publications.

SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS

A.1 We expect applicants seeking federal funds for developing or creating digital products to release these files under open-source licenses to maximize access and promote reuse. What will be the intellectual property status of the digital products (i.e., digital content, resources, or assets; software; research data) you intend to create? What ownership rights will your organization assert over the files you intend to create, and what conditions will you impose on their access and use? Who will hold the copyright(s)? Explain and justify your licensing selections. Identify and explain the license under which you will release the files (e.g., a non-restrictive license such as BSD, GNU, MIT, Creative Commons licenses; RightsStatements.org statements). Explain and justify any prohibitive terms or conditions of use or access, and detail how you will notify potential users about relevant terms and conditions.

This project will result in many digital products that will have copyright assigned to their creators or home organizations. These digital products will be free for distribution and use under a Creative Commons license, requiring attribution only. Software products will be owned by the University of Maryland and licensed for free distribution and use under the Apache 3.0 open source license. Both of these licenses make digital products widely available while also encouraging communication among stakeholders and future contribution of enhancements. Free and active participation in enhancement activities is of central concern to the project, as we collaboratively develop learning resources and test them in classrooms.

Copyright claims and licenses will be clearly listed on the front page of Jupyter notebook-based learning materials. Software products will be distributed with license and copyright information included in user interfaces, as well as in public source code repositories, following industry standard practices.

Other digital products, such as surveys, reports, and classroom assessments, will be cleaned of any personally identifiable student information and published on a relevant project or portal web page.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

University of Maryland will assert copyright on digital products created primarily by its faculty and staff. Proper attribution given to the creator and the institution will be the only condition for use of the digital products. Users of lesson plans, notebooks, and other products are

encouraged to modify and then redistribute their own versions. Therefore, the project will promote a standard guideline for giving proper attribution in derivative products.

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

Digital products involving students will be cleansed of personally identifiable information. If any documents are found later to raise privacy concerns, we will immediately remove them from the project.

Archival materials that receive attention and processing in Jupyter Notebooks will already be available to the public via the host institution. However, as a disruptive technology, the digital treatments and online access may raise new cultural sensitivity and privacy issues. These ethical concerns will be raised in all interactions with the Practitioner Network and continually examined as digital treatments are developed. Critical evaluation of ethical concerns is a core learning objective in this project's lesson plans. Therefore, we do hope to raise a number of such concerns in the lesson plans, explore the issues, and then demonstrate how they may be mitigated or avoided.

SECTION II: DIGITAL CONTENT, RESOURCES, OR ASSETS

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and the format(s) you will use.

This project will produce learning materials in the form of at least five sets of Jupyter Notebooks and accompanying documentation. The Jupyter Notebooks have their own format, which is a type of JavaScript which embeds Markdown and Python text blocks. The project will also produce a website and perhaps twenty or more web pages with project information. This website will also catalog and disseminate learning materials to faculty and practitioners. The project may also produce tabular classroom assessment data in comma-separated files (CSV).

A.2 List the equipment, software, and supplies that you will use to create the digital content, resources, or assets, or the name of the service provider that will perform the work.

The project will use several University of Maryland hosted and third-party services to create digital projects. UMD will host the project website and an instance of the JupyterHub software for developing Jupyter Notebooks. Participants of all kinds, including students in classrooms, will be encouraged to use the Binder service (mybinder.org) in order to view, execute, and further develop Jupyter Notebooks. The GitHub service (github.com) will be used as a hub of source code development and sharing. Various open source software packages will be used within Jupyter Notebooks in order to perform digital treatments. These may include, but are not limited to, DROID, Sigfreid, QGIS, Pandas, Tesseract, SpaCy, NLTK, and OpenCV.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG, OBJ, DOC, PDF) you plan to use. If digitizing content, describe the quality standards (e.g., resolution, sampling rate, pixel dimensions) you will use for the files you will create.

HTML, iPython Notebooks (ipynb), PDF, CSV. Practitioner institution data may also include TIFF, MPEG, and PDF.

Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan. How will you monitor and evaluate your workflow and products?

The Core Network team will convene regular meetings and follow the project timeline. Feedback and community input will be solicited and collected throughout the project and shared back with the community.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period. Your plan should address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

During the project our assets will largely remain hosted on the project website and in GitHub repositories. After the project period, we plan to leave these outputs on the website where they will continue to receive attention and modifications. An archival package containing all relevant materials will be produced and stored in the UMD DRUM digital repository.

Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata or linked data. Specify which standards or data models you will use for the metadata structure (e.g., RDF, BIBFRAME, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

Learning materials will be tagged with Dublin Core metadata fields to describe their content and provenance. Whenever possible this will include systematically embedded metadata fields in the corresponding Python Notebook files. This strategy will be adjusted based on feedback and technical design. In particular we need to provide term-driven browse and search capabilities on the website.
The project will produce or reproduce an ontology (SKOS or similar) of terms for tagging notebooks, to include computational practices, archival practices, and ethical considerations.

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

The metadata will remain embedded in project files and source code repositories, where it will travel with the data.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

The project website will include presentations of the relevant ontologies, and term-driven browse and search interfaces for finding learning materials. We will provide persistent and versioned online links to any original ontology files that are created.

Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content, delivery enabled by IIIF specifications).

Digital products will be openly available online through our project website, which will include a custom search engine and the specialized free software, Notebook Viewer, which allows us to showcase Python notebook content. Products, such as code, notebooks, and lesson plans will also be distributed via git software and the GitHub service. This channel will facilitate others making their own copies of products so they can adapt them to their own classrooms.

D.2. Provide the name(s) and URL(s) (Universal Resource Locator), DOI (Digital Object Identifier), or other persistent identifier for any examples of previous digital content, resources, or assets your organization has created.

CASES Prototype Website: http://cases.umd.edu
DRAS-TIC Testbed: http://drastic-testbed.umd.edu

SECTION III: SOFTWARE

General Information

A.1 Describe the software you intend to create, including a summary of the major functions it

will perform and the intended primary audience(s) it will serve.

We are not primarily creating a software application or software products. While the Jupyter notebooks do employ code, this code is for demonstration only as a teaching tool. It is not code that is formally packaged and distributed as an application or service.

We will create some ancillary code for our project website and that site will serve the information needs of instructors and practitioners.

A.2 List other existing software that wholly or partially performs the same or similar functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

We intend to reuse existing software to achieve most of our website goals, including the open source Jupyter Notebook Viewer software from the Jupyter organization.

Technical Information

B.1 List the programming languages, platforms, frameworks, software, or other applications you will use to create your software and explain why you chose them.

Instructional Jupyter Notebooks will employ Python, as well as several library modules, including DROID, Sigfreid, QGIS, Pandas, Tesseract, SpaCy, NLTK, and OpenCV.

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

Our website will extend and interoperate with the Jupyter Notebook Viewer software, by adding page and new functions to the existing web application framework.

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

N/A

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

All user-facing software will be third-party and not maintained by this project. Documentation for the example code in the Jupyter Notebooks will be given alongside that code within the Jupyter Notebook itself.

B.5 Provide the name(s), URL(s), and/or code repository locations for examples of any previous

software your organization has created.

https://github.com/UMD-DCIC/gatling-testbed
https://github.com/UMD-DCIC/vcl-dashboard
https://github.com/UMD-DCIC/IRP2

Access and Use

C.1 Describe how you will make the software and source code available to the public and/or its intended users.

We intend to make Jupyter Notebook content available to users through GitHub.

C.2 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository: Various Git repositories, one for each set of notebooks.

URL: To be determined

SECTION IV: RESEARCH DATA – N/A

As part of the federal government's commitment to increase access to federally funded research data, Section IV represents the Data Management Plan (DMP) for research proposals and should reflect data management, dissemination, and preservation best practices in the applicant's area of research appropriate to the data that the project will generate.

A.1 Identify the type(s) of data you plan to collect or generate, and the purpose or intended use(s) to which you expect them to be put. Describe the method(s) you will use, the proposed scope and scale, and the approximate dates or intervals at which you will collect or generate data.

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

A.3 Will you collect any sensitive information? This may include personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information. If so, detail the specific steps you will take to protect the information while you prepare it for public release (e.g., anonymizing individual identifiers, data aggregation). If the data will not be released publicly, explain why the data cannot be shared due to the protection of privacy, confidentiality, security, intellectual property, and other rights or requirements.

A.4 What technical (hardware and/or software) requirements or dependencies would be necessary for understanding retrieving, displaying, processing, or otherwise reusing the data?

A.5 What documentation (e.g., consent agreements, data documentation, codebooks, metadata, and analytical and procedural information) will you capture or create along with the data? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the data it describes to enable future reuse?

A.6 What is your plan for managing, disseminating, and preserving data after the completion of the award-funded project?

A.7 Identify where you will deposit the data:

Name of repository:

URL:

A.8 When and how frequently will you review this data management plan? How will the implementation be monitored?