

A Graduate Certificate in Web Archiving

Old Dominion University (ODU) requests \$98,361 (no cost share) from the IMLS Laura Bush 21st Century Librarian Program for a one year planning grant to establish a graduate certificate in web archiving at ODU. The project category is “National Digital Infrastructures and Initiatives” and the project type is “Continuing Education”. This project directly addresses the IMLS agency-level goal of “Build Capacity” through supporting the objectives of “Recruitment, training, and development of library and museum workforces” and “Share and adopt best practices and innovations”. This project will secondarily address the goal of “Increase Public Access” through the objectives of “Support stewardship of museum and library collections” and “Tools, technology, and training enabling discovery and use of collections and resources by all”. The deliverables for this project will include: 1) an environmental scan documenting the need of industry, university, and government organizations for employees with experience in web archiving, 2) materials for a graduate-level course in the technology, theory, and ethics of web archiving, and 3) a lessons learned report of our experiences working in a multidisciplinary team to deliver graduate and professional education that integrates library science, archival science, computer science, and cybersecurity.

1 Statement of Broad Need

We will address the lack of a formal curriculum to support organizations that have web archiving as part of their mission. This project will have a broad impact by providing a multi-disciplinary training curriculum for both new hires and existing employees of GLAM (galleries, libraries, archives, and museums) institutions and other organizations, such as government, military, and intelligence, charged with preserving web content. The web touches all parts of our lives, including culture, education, commerce, government, and entertainment. There is a broad-based industry for advancing the core technologies necessary to construct web-scale information systems that power and enrich these areas of human interaction. The nation’s computer science departments provide the research, development, and training of graduates with the skills necessary to enter this industry, but these graduates often lack the library and information science background necessary for memory organizations. Likewise, Library and Information Schools (LIS) produce graduates with the theoretical and practice of the assessment, collection, evaluation, organization, and management of a broad-based range of knowledge. Additionally, LIS graduates are conversant with issues of diversity, equity and inclusion (DEI) related to what is collected, who collects it, and who has access. But these graduates often lack the technical expertise to be the developers and toolsmiths necessary to keep up with the rapidly changing web environment.

Lastly, we are focused not only on the technology and theory of web archives, but also on the increasing role of cybersecurity *of, about, and from web archives*. The web is beset by misinformation and disinformation, as well as cheap fakes and deep fakes [1], and web archives offer the ability to launder and backdate disinformation to make it appear as if content was crawled on well-known, legitimate sites and at an earlier time [2]. Furthermore, there is an emerging body of research showing that web archives are used for unintended purposes, such as bypassing censorship [3], denying pageviews and advertising revenue [4], practicing attacks on sites without alerting them [5], obfuscating original sources [6–8], and bypassing server/browser security conventions [9]. Our research has shown that increasingly web archives are not just a target for hackers to corrupt or knock off-line, but they can also serve as a launchpad and vector for attacks on other areas of the web [10, 11]. Furthermore, we have demonstrated that web archives continually modify their contents on replay, and thus conventional fixity analysis and audits are unsuitable for determining change – and authenticity – of web archives [12–20]. This is where graduates need experience in cybersecurity, to verify the web content not just at the time of ingest, but at the time of replay as well [21, 22].

The web touches all aspects of human discourse, and the PIs are interested in technology for its development and capture, theory of its organization and stewardship, and asserting its authenticity and integrity – the convergence

of the three disciplines as shown in Figure 1. With the unique faculty resources we have at ODU, we will innovate in both research and education in support of this vision. This project will harness the following synergies:

1. The Web Science and Digital Library Research Group (@WebSciDL¹) within the ODU Department of Computer Science, now at seven faculty and 18 PhD students, is internationally recognized for its contributions to web archiving as well as digital libraries.
2. The ODU Library and Information Studies program offers the only Master's degree in library and information studies in the state of Virginia² with the degree on the cusp of ALA accreditation. They welcomed their first cohort of students in fall 2019³ and continue to grow and expand.
3. ODU's Center for Cybersecurity Education and Research has just recently become the School of Cybersecurity.⁴ ODU began its cybersecurity program in 2015 with 11 students, and the new School of Cybersecurity now has more than 800 BS and MS students.

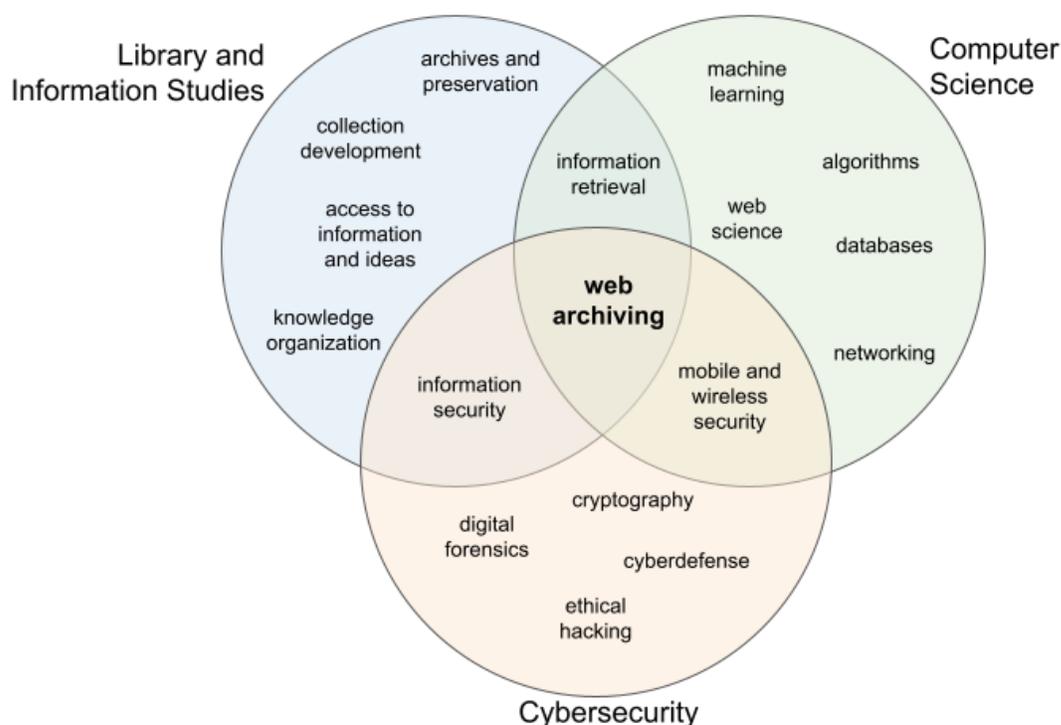


Figure 1: Our certificate and capstone course will be at the intersection of Computer Science, Library and Information Studies, and Cybersecurity.

The Harvard Library's Web Archiving Scan [23] recommended blending a traditional library focus on collection development with "smart technical development." Our certificate represents such a synergy with computer science (Nelson, Weigle), librarianship (Kimmel), archives (Ritchie), and security (Wu) – a synergy we believe is unique and will serve as a model for web archiving and general web science related collaboration elsewhere.

¹<https://twitter.com/WebSciDL>, <https://ws-dl.cs.odu.edu/>, <https://github.com/oduwsdl>, <https://ws-dl.blogspot.com>

²<https://www.odu.edu/stem/academics/library-science/about>

³We have every reason to assume a positive outcome for ALA accreditation, but in the unlikely event of not receiving accreditation in the current round, the MLIS will continue and we will reapply.

⁴https://www.odu.edu/news/2020/9/school_of_cybersecur

The National Digital Stewardship Alliance’s (NDSA) Web Archiving in the United States Survey [24] highlighted several areas in which the archiving field has evolved, including overall diversification of the field, maturation and expansion of programs, and technological advancements that have allowed for greater collaboration among institutions and communities. Our certificate represents a forward-thinking collaborative approach to training new professionals that will help broaden the field, as well as continue to diversify and expand web archives collections and outreach initiatives.

Noting a particular need for specialization in digital contexts, the US Bureau of Labor Statistics (BLS) projects “much faster than average” growth in the job outlook for archivists [25]. However, finding graduates that have both the technical web skills and the archival and information science background is rare. The Internet Archive, Library of Congress, and other employers have to compete to hire unicorns or train their existing staff in one area or the other. @WebSciDL is internationally recognized for its digital library and web archiving research as well as the quality of its graduates: in addition to conventional post-doc and tenure-track positions, our web archivists with PhD and MS degrees are finding employment at, among others: the Internet Archive, NASA, Library of Congress, Los Alamos National Laboratory, MITRE, Stanford, Berkeley, UCLA, Indiana University, and Emory University.

But despite our efforts, we cannot create enough MS and PhD graduates to fill the needs of these organizations. And for most existing staff at various libraries and archives, a complete graduate degree is also more than they need, or not a realistic life option.⁵ In this planning grant, we will address the post-graduate training needs of the organizations listed above by building on the materials we co-created in the CEDWARC⁶ project and establishing *an on-line, four course (12 hour) graduate certificate in web archiving*, drawing from the faculty from Computer Science, Library and Information Studies, University Libraries, and the School of Cybersecurity.

2 Project Design

Our ultimate goal is to develop a first-of-its-kind online, interdisciplinary graduate certificate to train professionals to better perform web archiving duties in their workplace. The primary results of this planning grant will be a market analysis for this certificate, a capstone course that incorporates input from our advisory committee and that integrates computer science, library science, and cybersecurity in an innovative and novel manner, and a recommended plan of study for the remaining three courses.

Because ODU is the institution offering this certificate, the nature of this proposal is about integrating disciplines from schools and departments within ODU. ODU has over 30 certificate programs⁷ at both the graduate and undergraduate levels, and the computer science department itself currently offers two graduate certificates: “Graduate Certificate Program in Cybersecurity (Online)” and “Modeling and Simulation Certificate in Computer and Informatics”.⁸ Our vision is to establish another certificate at ODU, focused on web archiving. To achieve organizational diversity, establish a national and international perspective, and to allow us to enter Phase One of the project in Fall 2021, we have already begun assembling an advisory board. The list below is the current *confirmed* members of our advisory board (a sampling of their emails committing to serving on the advisory board is included in the Supporting Documents section). We expect the list to continue to evolve, and we welcome further suggestions.

1. Amelia Acker, University of Texas,

⁵For example: <https://twitter.com/hrbrmstr/status/1268360642291011586>

⁶<https://cedwarc.github.io/>. Note that the CEDWARC project [26] is primarily about training modules for popular archiving software tools (e.g., ArchiveSpark [27], Archives Unleashed Toolkit [28], Social Feed Manager [29]) and purposefully does not address theory, ethics, security, and other topics appropriate for a graduate course of study in web archiving.

⁷<https://online.odu.edu/academics/degree/certificates>

⁸<https://www.odu.edu/compsci/academics/grad-cert>

- <https://www.ischool.utexas.edu/people/people-details?PersonID=302>
2. Sawood Alam, Internet Archive, <https://www.linkedin.com/in/ibnesayeed/>
 3. Justin Brunelle, MITRE, <https://www.linkedin.com/in/justinfbrunelle/>
 4. Sumitra Duncan, The Frick Collection, <https://www.linkedin.com/in/sumitraduncan/>
 5. Dragan Espenschied, Rhizome, <https://rhizome.org/profile/despens/>
 6. Mark Graham, Internet Archive, <https://www.linkedin.com/in/markjohngraham>
 7. Pamela Graham, Columbia University, https://guides.library.columbia.edu/prf.php?account_id=27393
 8. Jane Greenberg, Drexel University, <https://drexel.edu/cci/about/directory/G/Greenberg-Jane/>
 9. Olga Holownia, The British Library, <https://netpreserveblog.wordpress.com/2016/06/06/iipc-meet-the-officers-2016/#390>
 10. Leslie Johnston, NARA, <https://www.linkedin.com/in/lesliejohnston/>
 11. Martin Klein, Los Alamos National Laboratory, <https://www.linkedin.com/in/martinklein0815>
 12. Ginger Garrison, NASA Langley Research Center, <https://www.linkedin.com/in/gingergarrison/>
 13. Tori Maches, University of California, San Diego, <https://www2.archivists.org/groups/web-archiving-section/2020-2021-web-archiving-section-leadership-announcement>
 14. Marcia Mardis, Florida State University, <https://directory.cci.fsu.edu/marcia-mardis/>
 15. Ian Milligan, University of Waterloo, <https://uwaterloo.ca/history/people-profiles/ian-milligan>
 16. Christie Moffat, National Library of Medicine, <https://www.linkedin.com/in/christie-moffatt-7a771866/>
 17. Abigail Potter, Library of Congress, <https://www.linkedin.com/in/abigailpotter/>
 18. Elizabeth (Lily) Pregill, J. Paul Getty Trust, <https://www.linkedin.com/in/lilypregill/>
 19. Robert Sanderson, Yale University, <https://www.linkedin.com/in/robert-sanderson/>
 20. Vanessa Thaxton-Ward, Hampton University, <https://www.linkedin.com/in/vanessa-thaxton-ward-ph-d-17265126/>
 21. Helen Tibbo, University of North Carolina, <https://sil.s.unc.edu/people/faculty/profiles/Helen-Tibbo>

We have assembled an extensive advisory board in order to have a range of input from different organizations, disciplines, and societies. Many of the organizations (e.g., Internet Archive, Library of Congress, Rhizome, LANL, NASA) have hired our graduates in the past and will be especially well-positioned to provide feedback for the proposed graduate certificate. We have a mix of researchers, developers, educators, practitioners, and managers

with hiring authority. The organizations represent a broad cross section of significant people and organizations involved in web archiving:

- iSchool faculty: Drexel University, University of North Carolina-Chapel Hill, University of Texas-Austin, Florida State University
- Other faculty: University of Waterloo
- University libraries: Hampton University, University of California-San Diego, Columbia University
- Government: Library of Congress, National Archives and Records Administration, National Library of Medicine, Los Alamos National Laboratory, NASA, British Library
- Non-profit: Internet Archive, MITRE
- Museums: Rhizome, The Frick Collection, J. Paul Getty Trust, Yale University
- Past and present leadership in the following societies, organizations, and projects: Society for American Archivists (SAA), American Library Association (ALA), Association for Computing Machinery (ACM), International Internet Preservation Consortium (IIPC), Documenting the Now (DocNow) [30, 31], National Forum on Ethics and Archiving the Web (EAW) [32].

We will enlist the advisory board for input on all three of the phases of this project:

Phase One (months 1–4): 1) All PIs will survey (led by Weigle) existing USA graduate programs and certificate opportunities in Computer Science, iSchools, and Cybersecurity, and 2) perform (led by Ritchie) a hiring needs survey of GLAM institutions as well as other institutions with memory organization requirements at the federal, state, and university level. This will include the skills they wish to hire (new staff), as well as grow (existing staff). Both surveys will include questions about efforts to recruit from under-represented populations and to promote diversity. Many in the advisory board will represent the needs of their own institutions, as well as recommend additional organizations that we should survey (i.e., snowball sampling). We will also use professional meetings, email lists, and social media to solicit feedback.

Phase One will culminate in *Deliverable 1*, an environmental scan regarding the needs of web archiving expertise, how organizations are currently meeting these needs, and what they would like to see in the near- and long-term future. Anecdotally, we know there is a need for these skills, but this is our chance to formally gather, analyze, and publish these results. The findings from the scan will be shared with the Advisory Board for their advice moving into Phase Two.

Phase Two (months 4–12): Building upon Phase One, we will develop a cross-listed capstone web archiving course (led by Nelson, Kimmel, Wu): CS/CYSE/LIBS 433/533 - Web Archiving Theory, Practice, and Implications. Although the exact course content will be determined during the course of this project, Table 1 shows our current thinking about what such a course should cover. We will also identify possible additional courses (Table 2) from the existing course offerings in Computer Science, Library and Information Studies, and Cybersecurity (led by Nelson, Kimmel, Weigle, Wu), as well as propose possible new electives in order to create a certificate in Web Archiving. The new LIS program has a new course: Seminar in Archives and Special Collections, with proposed additional coursework in access and preservation to meet student and employer interests. Digital materials management will be addressed in each course. Final recommendations from existing courses and/or proposed new courses will be informed by Deliverable 1.

Phase Two will produce *Deliverable 2*, course materials for the cross-listed capstone course. The course will incorporate established and emerging theories related to the development of equitable, diverse, inclusive collections

| Week | Topics |
|---------|--|
| Week 1 | Review of basics: command line, Unix, Python, shell scripting. |
| Week 2 | History of the Web, HTTP Mechanics, W3C Web Architecture, Memento, curl, wget, WARC. |
| Week 3 | Web Crawling and Replay, Heritrix, Open Wayback, pywb, Webrecorder/Conifer. |
| Week 4 | Collection analysis, ArchiveSpark, Archives Unleashed Toolkit. |
| Week 5 | Collection development policies for web archives, collection building, collection evaluation, seed selection. Issues of DEI related to collections. |
| Week 6 | Web archive security, Web archive integrity, authenticity and temporal violations, archival illusions. |
| Week 7 | Archival Appraisal and Selection, Provenance, Collaborative theory of appraisal, community collecting in the digital age. Developing equitable, diverse, and inclusive collections through collaborative archival practices. |
| Week 8 | Descriptive Metadata for Web Archiving, extracting metadata, deriving metadata, creating inclusive and reparative metadata. |
| Week 9 | Ethics of Web archiving, Right to be Forgotten, Right to be Remembered, DEI, reimagining and expanding the archive to include marginalized voices communities. |
| Week 10 | Threats to archives, archives as threats, counter-archiving. |
| Week 11 | Collaborative collecting harvesting social media. Who archives, what, and for whom? Diversity and inclusion as reflected in possible topics, themes, tools, and formats. |
| Week 12 | Collection summarization, collection visualization. |
| Week 13 | Storytelling, Dark and Stormy Archives, How Narratives Shape Archives. |
| Week 14 | Web data privacy, Digital Forensics. |
| Week 15 | Intellectual property, copyright, Fair Use. |

Table 1: A strawman proposal for the proposed CS/CYSE/LIBS 433/533 - Web Archiving Theory, Practice, and Implications. The actual content will evolve during the period of the planning grant. Suggestions and feedback about course content are welcome at any time.

and practices. By centering the importance of collaborative practices, the capstone course will instill the importance of identifying, preserving, and working with communities to actively document those whose voices have been overlooked or marginalized. We also anticipate identifying which other courses will fill out the 12 credit hours for the proposed certificate, possibly including:⁹

- CS 418/518 Web Programming
- CS 432/532 Web Science
- CS 495/595 Web Client Security
- CS 751/851 Digital Libraries
- CS 795/895 Web Archiving Forensics
- LIBS 608 Foundations of Libraries and Information

⁹A sampling of some of these courses as they exist online: <https://github.com/phonedude/cs595-s21>, <https://www.cs.odu.edu/~jbrunelle/cs518/>, <https://www.cs.odu.edu/~mweigle/CS432-S21/>, <https://github.com/phonedude/cs895-f20>.

- LIBS 612 Research Methods in Library and Information Studies
- LIBS 658 Knowledge Management: Planning, Selecting and Managing Collections
- LIBS 677 Knowledge Organization and Access
- LIBS 681 Assessment and Evaluation in Library and Information Science
- LIBS 693 Seminar in Archives and Special Collections
- CYSE 465/565 Information Assurance for Cybersecurity
- CYSE 464/564 Networked Systems Security
- CYSE 607 Advanced Digital Forensics

At this crucial development stage, we anticipate frequent input from the advisory board on both the capstone course and the certificate. As has been our idiom of late, we anticipate making the course materials available as a GitHub repository, since it facilitates both making copies and collaborative editing. While this certificate will be offered by ODU, we will share the course materials so that they may be adapted in part or whole for education by other institutions and organizations. Our current plan is to offer the capstone course in Fall 2022, after the period of the planning grant is over. While ODU is currently uniquely positioned to develop and deliver this highly specialized, interdisciplinary certificate, we would welcome competition from other universities as validation of the need for this kind of training. Our goal is not to be the only school with a curriculum focused on web archiving, but ODU will take the lead in order to show that it can and should be done.

Phase Three (months 11+): Throughout Phases One and Two we will be seeking input from the advisory board as well as self-evaluating. After the planning grant ends, all PIs will review their lessons learned about both the certificate content and planning, as well as the *meta-level interdisciplinary collaboration itself*. This final analysis will also be shared with the advisory board for their input and advice moving forward.

Phase Three will produce *Deliverable 3*, a lessons learned document that we will both formally publish as well as make available as an arXiv eprint, in order to facilitate projects with similar interdisciplinary goals. Since Phase Three will begin near the end of the project, Deliverable 3 will be published after the initial year is complete.

Depending on the impact of Covid-19 on travel, we hope to physically travel to places such as the Internet Archive, Library of Congress, and National Archives and Records Administration, to support Phase One. However, if that is not possible, then we will conduct our business over Zoom and other teleconferencing platforms as we have done since March 2020. Given the size of the advisory board, we will not attempt to schedule an in-person meeting with all members, but will rely on teleconferencing for meetings. Private documents will be shared between PIs and the advisory board using either Google Drive, Microsoft Teams, or MediaWiki. We will establish an email list and a public-facing web site to communicate our goals and status to the larger community. In addition to the eprints and formal publication of Deliverables 1 and 3, we will also plan to discuss the project at panels and presentations at professional meetings, such as the ACM/IEEE Joint Conference on Digital Libraries, IIPC Web Archiving Conference, iPres, etc.

The first indicator for this project's success has already been established via the enthusiastic replies from the people participating in the advisory board. If we had not identified a significant need, we would not have enjoyed this level of uptake. Certainly the continued positive feedback of the advisory board (and the additional people and organizations they recommend we survey) will continue to be a primary indicator of success throughout this project. Furthermore, arriving at a comprehensive capstone course and corresponding plan of study that all PIs are enthusiastic about is the other primary indicator of success. If the PIs and advisory board are unwilling

to recommend the classes and certificate program to employees, students, and friends, then we have not been successful.

In addition to the advisory board feedback, we will evaluate the project based on external interest in Deliverables 1 (environmental scan) and 3 (lessons learned). Deliverable 2 (the capstone course) will be evaluated based on initial enrollment numbers and ODU's existing student opinion surveys¹⁰ at the conclusion of the course. Regarding the new IMLS reporting criteria, we imagine:

- **Effectiveness:** The project will have been effective if Deliverable 1 (environmental scan) is met with enthusiastic response from the advisory board, Deliverable 2 (capstone course) is set to be taught at ODU, and Deliverable 3 (lessons learned) has been published as an eprint.
- **Efficiency:** Efficiency opportunities will be scoped according to the limited budget of a planning grant, but the primary goal is to complete all deliverables in the first 12 months so we can begin offering the certificate as soon as possible.
- **Quality:** During the 12 months of the planning grant, quality will be measured qualitatively by the feedback from the advisory board. After the capstone course is offered, continued enrollment numbers will establish if the needs of the target audience are being met.
- **Timeliness:** Similar to efficiency, the 12 months of a planning grant give us a tight window of opportunity. We will have to adhere to the time table in the schedule of completion to complete our deliverables in the time allotted.

3 Diversity Plan

In designing this project, we are guided by the Society of American Archivists' Core Values Statement and Code of Ethics.¹¹ We believe that it is essential for archival education programs, professional organizations, and hiring institutions to work together to develop practices and policies that support the recruitment, retention, and ongoing support of inclusive communities of practice. By embracing collaborative practices, our project will actively work to achieve a diversified and representative membership in the profession.

The synergy among the various partners developing this certificate is enhanced by the differences among library science, computer science, and cybersecurity. While it is too soon to know the demographics of those who enroll in the certificate program, Table 2 shows the demographics of the constituent disciplines as of Fall 2019 (i.e., pre-Covid-19), for BS and MS students since they are our target audience. Some of the highlights include an aggregate student body of 28% female, 56% minorities, 27% Pell Grant recipients, and 23% first generation students.

| | Total enrollment | Minority | Female | Pell Grant | Veteran | First Generation | US Citizen |
|--------------------------------------|------------------|------------------|------------------|------------------|------------------|------------------|-------------------|
| Computer Science BS + MS | 688 | 357 | 138 | 164 | 53 | 129 | 637 |
| Cybersecurity BS + MS | 659 | 430 | 186 | 225 | 65 | 203 | 657 |
| Library and Information Studies MLIS | 83 | 13 | 76 | 0 | 2 | 2 | 83 |
| totals | 1430 | 800 (56%) | 400 (28%) | 389 (27%) | 220 (15%) | 334 (23%) | 1377 (96%) |

Table 2: Demographics as of Fall 2019.

¹⁰<https://www.odu.edu/academics/courses-registration/studentopinionsurvey>

¹¹<https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics>

In the needs assessment first phase of the project, the team will include inquiries regarding diversity. We assume we know what kinds of diversity employers are seeking, but survey questions could probe those assumptions. The inclusion of diverse participants in each survey might further expand our understandings. Are there existing certificates in Computer Science, iSchools, and Cybersecurity in HBCUs or other successful minority-serving universities? What can libraries and archives with a focus on diverse populations tell us about attracting more diverse students to these workplaces? Members of the HBCU Library Alliance¹² will be intentionally included in the needs assessment.

In the design of new courses and a Web Archiving certificate in the second phase, we anticipate that team members will continue the work developing a channel that attracts students to cross over from both fields. While much of web archiving is technical work, the content of web archives and the various users who might benefit from access to those archives should reflect the kinds of diversity found in communities and workplaces. Unearthing those human and service connections may attract librarians to the technical work and offer grounding to those from the more technical computer science field.

By planning to offer this certificate online, we make it easier for remote and local working students to attend, including students who otherwise can not take time off from their jobs or families (some of ODU's courses are asynchronous, and the synchronous elective graduate courses are frequently offered in the evening). Distance coursework increases geographic diversity and may also attract more non-traditional students, many with diverse life and educational experience. Online teaching, as we have all recently discovered due to Covid-19, requires thoughtful planning, such as that enabled by this grant.

While not an HBCU, ODU is a "minority serving institution"¹³ and has an enrollment that reflects the region. Expanding from the three disciplines highlighted in Table 2, half of ODU's first-time freshmen are "first generation" students, 42% of our undergraduates receive Pell Grants,¹⁴ 25% of our students have some affiliation with the military,¹⁵ and 31% of our students, graduate and undergraduate, are "part-time".¹⁶ All of this is to say that ODU has demonstrated expertise in servicing what used to be called "non-traditional" students; in fact, our very existence depends on it.

In support of "non-traditional" students, Old Dominion University has a well established history of distance education that is pedagogically and technologically supported by ODUOnline¹⁷ and the Center for Learning and Teaching.¹⁸ The University allows a student to enroll in up to 12 credit hours (generally 4 courses) as a non-degree seeking student, providing a pathway for those who already have a graduate degree, or who want to explore graduate coursework, to participate in graduate coursework and earn credits that might be applied to an eventual degree. Graduate certificates provide a more formalized non-degree path for a student who wishes to demonstrate a defined area of competence. Generally, students apply to enter a certificate program, may be eligible for financial aid based on enrollment in a certificate program, and upon completion, have the certificate noted on their transcript. Certificates undergo review through the Office of Institutional Effectiveness and Assessment at ODU.¹⁹

¹²<http://www.hbculibraries.org/>

¹³<https://odu.edu/facultystaff/research2/funding/msi>

¹⁴<https://www.pilotonline.com/opinion/columns/vp-ed-column-stansberry-0510-20200510-n7nqykhqqjenlktwp3ktvzhpq-story.html>

¹⁵<https://www.odu.edu/admission/military-admission>

¹⁶https://datausa.io/profile/university/old-dominion-university#enrollment_status

¹⁷<https://online.odu.edu/>

¹⁸<https://clt.odu.edu/>

¹⁹<https://www.odu.edu/assessment>

4 Broad Impact

Web archiving straddles multiple disciplines, and the staff at galleries, libraries, archives, museums, and other memory organizations rarely have the full range of skills necessary to perform these tasks. For example, organizations like the Internet Archive and Library of Congress have to internally train their staff, and lack a uniform curriculum to assist in this task. Based on our record of success and close ties with many of these organizations, we will incorporate the feedback from our advisory board and combine it with the breadth of expertise of the PIs to provide the necessary training for federal, state, and university libraries and archives. In doing so, we will establish the first steps for web archiving as an emerging discipline unto itself.

We will formally publish our results in order to enjoy the benefit of peer review, but we will (as is our norm) also publish as eprints so they may be widely disseminated. The course materials and course plan will be made in a GitHub repository under an MIT License²⁰ so that other institutions may emulate our success. Because we have a history of making our course materials available online for web crawling, many of our syllabi are already indexed in Open Syllabus.²¹

We have been fortunate enough to place many of our MS and PhD graduates at memory organizations like the Internet Archive, Library of Congress, Los Alamos National Laboratory, NASA, Rhizome, as well as numerous university libraries. However, we cannot produce enough conventional graduates to fill the web archiving needs of such organizations, and for existing staff, many of them mid-career, an additional graduate degree is simply not feasible. Anecdotally, we know that there is a demand for this kind of training, but this planning grant will enable us to formally describe and measure the demand, as well as provide a template for other organizations to follow in similar and related interdisciplinary educational endeavors.

²⁰For example: <https://github.com/phonedude/cs595-s21/blob/main/LICENSE>

²¹A list of materials assigned in ODU CS syllabi can be found at: https://opensyllabus.org/result/school-field?id_field=Computer+Science&id_school=1030792151044, but we have just recently learned that the full syllabi themselves are not shared unless the instructors opt in, as per: <https://blog.opensyllabus.org/share-syllabi/>. We will ensure that these materials are available for distribution from Open Syllabus and similar projects as we discover them.

| August 1, 2021-July 31, 2022 | | | | | | | | | | | | | | | |
|--------------------------------------|--------|------|--------|---------------------|-------------|-----|-----------------------|-----|-----|-----|-----|--------|--------|--------|--------|
| | AUG | SEP | OCT | NOV | DEC | JAN | FEB | MAR | APR | MAY | JUN | JUL | | | |
| Phase One: Environmental Scan | | | | | | | | | | | | | | | |
| PH 1 | | | | | | | | | | | | | | | |
| 1.1 Program survey | Yellow | Blue | Yellow | Yellow | | | | | | | | Yellow | Blue | | |
| 1.2 Hiring survey | Yellow | Blue | Yellow | Yellow | | | | | | | | Yellow | Blue | | |
| Phase Two: Develop Coursework | | | | | PH 2 | | | | | | | | | | |
| 2.1 Capstone course | | | Yellow | Blue | | | | | | | | Yellow | Yellow | Yellow | Blue |
| 2.2 Identify electives | | | | | | | | | | | | Yellow | Yellow | Blue | Yellow |
| Phase Three: Evaluation | | | | | PH 3 | | | | | | | | | | |
| | | | | Yellow | | | | | | | | Blue | Yellow | Blue | |
| Participants | | | | Project Team | | | Advisory Board | | | | | | | | |



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.

Rights in all digital products created for this project will follow Old Dominion University's Intellectual Property Policy and vest either with their creator or joint creators, which will then be relicensed for free reuse under the MIT License. The MIT License grants limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the digital objects. We anticipate making the course materials available as a GitHub repository, since it facilitates both making copies and collaborative editing. We do not anticipate any prohibitive restrictions.

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.

As stated above, for this grant, rights remain with individual creators subject to the MIT License. We will impose no restrictions on reuse beyond those included in the MIT License.

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.

We do not anticipate creating any digital products that raise privacy concerns or cultural sensitivities. Because we anticipate creating course materials that may be reused, our instructions to course instructors will be clear that any use of third-party materials must either include permission or otherwise be appropriate for reuse online under U.S. law, including fair use.

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.

The primary digital content we anticipate creating are course materials and the related package of course content that we will publish with the hopes of making the model of education as transparent and easily duplicable by others. We also plan to collect feedback from the project's advisory board and survey participants, and will produce an environmental scan report.

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.

All course material will be created on standard office productivity tools -- Microsoft Office, Google Drive, Microsoft Teams- as well as MediaWiki. We will establish an email list and a public-facing web site (odu.edu and/or github.com) to communicate our goals and status to the larger community.

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.

.pdf, .tex, .docx, .txt, .md (markdown) and possibly Python programs (.py).

Workflow and Asset Maintenance/Preservation

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

Quality of feedback and course materials will be ensured by regular review and refinement by project PIs. The project PIs will be seeking input from the advisory board as well as self-evaluating. After the planning grant ends, all PIs will review their lessons learned about both the certificate content and planning, as well as the meta-level interdisciplinary collaboration itself. This final analysis will also be shared with the advisory board for their input and advice moving forward.

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).

Digital course materials and final reports will be preserved through the ODU Digital Commons Institutional Repository, which is a service the ODU Libraries that supports the activities of ODU's faculty, researchers, students, and library staff by preserving, securing, and providing access to digital resources. We will also utilize GitHub.com, arXiv.org, and opensyllabus.org for sharing course materials.

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).

After ingest into the ODU Digital Commons repository, our materials will be assigned standardized Dublin Core descriptive metadata in consultation with repository staff. All other repositories will apply their own metadata standards.

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

Metadata will be preserved with the deposited content in accordance with ODU repository policies.

<https://guides.lib.odu.edu/c.php?g=502940&p=5839940>

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).

Because our materials are designed for use, we anticipate communicating digital content and targeting relevant audiences through listservs, social media, blogging, etc. We will also expose digital materials openly to a variety of aggregators and encourage the project's Advisory Board to participate in promoting the discovery and use of content.

Since the primary products will be reports and course materials, APIs, etc. are N/A.

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).

All digital content will be made openly available online through the ODU Digital Commons repository and through the project website, both accessible via standard web browsers. We will formally publish our results in order to enjoy the benefit of peer review, but we will also publish as eprints so they may be widely disseminated. The course materials and course plan will be made in a GitHub repository under an MIT License and indexed in Open Syllabus so that other institutions may emulate our success.

Example courses available through github.com:

<https://github.com/phonedude/cs595-s21>
<https://github.com/phonedude/cs895-f20>
<https://github.com/phonedude/cs532-s17>

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.

<https://digitalcommons.odu.edu/>
<https://dc.lib.odu.edu/digital/>

Examples:

<https://doi.org/10.25777/pc8d-y213>
https://digitalcommons.odu.edu/computerscience_fac_pubs/172/
https://digitalcommons.odu.edu/stemps_fac_pubs/85/

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.

It's possible that some software will be developed in the context of the capstone course, but otherwise software is not a deliverable for this project. So we are considering this as "N/A".

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.

N/A

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.

N/A

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

N/A

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.

N/A

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

This is "N/A" for this project, but our other software and data can be found at:

<https://github.com/oduwsdl>

Access and Use

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

N/A

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

Name of publicly accessible source code repository:

N/A

URL:

N/A

SECTION IV: RESEARCH DATA

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.

N/A

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?

N/A

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.

N/A

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

N/A

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?

N/A

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

N/A

A.7 Identify where you will deposit the data:

Name of repository:

N/A

URL:

N/A

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

This project is 12 months long, with three deliverables. Each deliverable is a good opportunity to review the efficacy of the DMP:

- * Deliverable 1 (environmental scan) - month 4
- * Deliverable 2 (capstone course) - month 12
- * Deliverable 3 (lessons learned) - month 12+