

## White Paper on the IMLS Machine Learning Grant

Alex Papson, Anastasia Guimaraes, Christina Leblang, Daniel Johnson, Donald Brower, Eric Lease Morgan, Helen Hockx-Yu, John (Zheng) Wang, Laurie McGowan, Mark Dehmlow, Melissa Harden, Rebecca Leneway

### Publication Date

08-12-2023

### License

This work is made available under a Exclusive rights in copyrighted work license and should only be used in accordance with that license.

### Citation for this work (American Psychological Association 7th edition)

Papson, A., Guimaraes, A., Leblang, C., Johnson, D., Brower, D., Morgan, E. L., Hockx-Yu, H., Wang, J. (Zheng) ., McGowan, L., Dehmlow, M., Harden, M., & Leneway, R. (2020). *White Paper on the IMLS Machine Learning Grant* (Version 1). University of Notre Dame. <https://doi.org/10.7274/r0-320z-kn58>

This work was downloaded from CurateND, the University of Notre Dame's institutional repository.

For more information about this work, to report or an issue, or to preserve and share your original work, please contact the CurateND team for assistance at [curate@nd.edu](mailto:curate@nd.edu).

## Table of Contents

Appendices	1
Appendix A: Bibliography	3
Classification	3
Cross-disciplinarity	4
Discovery	5
Machine Learning	5
Natural Language Processing	6
Topic Modeling	6
Appendix B: Survey	8
Questions	8
Responses (Anonymized to protect respondent privacy.)	26
Appendix C: Workshops	107
Notre Dame Workshop	107
Agenda	107
Workshop Participants	108
Afternoon Discussion	109
Breakout Discussions	110
Palo Alto Workshop	113
Agenda	113
Workshop Participants	114
Brainstorming Activity	115
Breakout Discussions	119
New York Workshop	121
Agenda	121
Workshop Participants	122
Brainstorming Activity	123
Breakout Discussions	128
Washington D.C. Workshop	131
Agenda	131
Workshop Participants	132
Brainstorming Activity	133
Breakout Discussions	137
Writers' Workshop	139
Agenda	139
Workshop Participants	140

## Appendix A: Bibliography

### Classification

- Bao, Yang, Nigel Collier, and Anindya Datta. 2013. "A Partially Supervised Cross-Collection Topic Model for Cross-Domain Text Classification." In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*, 239–248. CIKM '13. New York, NY, USA: ACM. <https://doi.org/10.1145/2505515.2505556>.
- Beghtol, C. 19980101. "Knowledge Domains: Multidisciplinarity and Bibliographic Classification Systems." *Knowledge Organization* 25 (1).
- Bittermann, André, and Andreas Fischer. 2018. "How to Identify Hot Topics in Psychology Using Topic Modeling." *Zeitschrift Für Psychologie* 226 (1): 3–13. <https://doi.org/10.1027/2151-2604/a000318>.
- Brygfjeld, Svein Arne, Freddy Wetjen, and André WalsØe. n.d. "Machine Learning for Production of Dewey Decimal." In . <http://library.ifla.org/2216/1/115-brygfjeld-en.pdf>.
- Danilevsky, M., C. Wang, N. Desai, X. Ren, J. Guo, and J. Han. 2014. "Automatic Construction and Ranking of Topical Keyphrases on Collections of Short Documents." In *Proceedings of the 2014 SIAM International Conference on Data Mining*, 398–406. Proceedings. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611973440.46>.
- Dědek, Jan, Peter Vojtáš, and Marta Vomlelová. 2012. "Fuzzy ILP Classification of Web Reports after Linguistic Text Mining." *Information Processing and Management* 48 (3): 438–450. <https://doi.org/10.1016/j.ipm.2011.02.008>.
- Denda, Kayo. 2005. "Beyond Subject Headings." *Library Resources & Technical Services* 49 (4): 266–275. <https://doi.org/10.5860/lrts.49n4.266>.
- Friedman, Carol, Tara Borlawsky, Lyudmila Shagina, H Xing, and Yves Lussier. 2006. "Bio-Ontology and Text: Bridging the Modeling Gap." *Bioinformatics* 22 (19): 2421–29. <https://doi.org/10.1093/bioinformatics/btl405>.
- Gu, H, Y Perl, J Geller, M Halper, and M Singh. 1999. "A Methodology for Partitioning a Vocabulary Hierarchy into Trees." *Artificial Intelligence in Medicine* 15 (1): 77. [https://doi.org/10.1016/S0933-3657\(98\)00046-3](https://doi.org/10.1016/S0933-3657(98)00046-3).
- Hagedorn, Kat, Michael Kargela, Youn Noh, and David Newman. 2011. "A New Way to Find: Testing the Use of Clustering Topics in Digital Libraries." *D-Lib Magazine* 17 (9/10). <https://doi.org/10.1045/september2011-hagedorn>.
- Jian Qin, and Stephen Paling. 2001. "Converting a Controlled Vocabulary into an Ontology: The Case of GEM." *Information Research: An International Electronic Journal* 6 (2): <xocs:firstpage xmlns:xocs=""/>.
- Junger, Ulrike. n.d. "Automation First – the Subject Cataloguing Policy of the Deutsche Nationalbibliothek." In . <http://library.ifla.org/2213/1/115-junger-en.pdf>.
- "Knowledge Management Section Joint with Academic and Research Libraries Section and Rare Books and Special Collections Section." n.d. Google Docs. Accessed September 19, 2018. <https://docs.google.com/document/d/1klI0CqGITORXXaPHPHXM0ksLarCHPIacJ5yDiSeZJ7o/ed>

[it?usp=drive\\_web&oid=100690309179145782813&usp=embed\\_facebook](https://drive.google.com/document/d/1kII0CqGITORXXaPHPHXM0ksLarCHPIacJ5yDiSeZJ7o/ed).

- Lebow, David G. 2018. "A Social Machine for Transdisciplinary Research." *Informing Science* 21 (January): 201–17. <https://doi.org/10.28945/4025>.
- Mimno, David, and Andrew McCallum. 2007. "Organizing the OCA: Learning Faceted Subjects from a Library of Digital Books." In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, 376–385. JCDL '07. New York, NY, USA: ACM. <https://doi.org/10.1145/1255175.1255249>.
- Newman, David, Kat Hagedorn, Chaitanya Chemudugunta, and Padhraic Smyth. 2007. "Subject Metadata Enrichment Using Statistical Topic Models." In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, 366–375. JCDL '07. New York, NY, USA: ACM. <https://doi.org/10.1145/1255175.1255248>.
- Sabharwal, Arjun. 2011. "Knowledge Domain Navigation in Interdisciplinary Digital Landscapes." *Journal of Library Metadata* 11 (2): 63–82. <https://doi.org/10.1080/19386389.2011.570655>.
- Shu, Fei, Jesse David Dinneen, Banafsheh Asadi, and Charles-Antoine Julien. 2017. "Mapping Science Using Library of Congress Subject Headings." *Journal of Informetrics* 11 (4): 1080–94. <https://doi.org/10.1016/j.joi.2017.08.008>.
- Stein, Roger Alan, Patricia A. Jaques, and João Francisco Valiati. 2019. "An Analysis of Hierarchical Text Classification Using Word Embeddings." *Information Sciences* 471: 216–32. <https://doi.org/10.1016/j.ins.2018.09.001>.
- Vishwanath Bijalwan, Vinay Kumar, Pinki Kumari, and Jordan Pascual. 2014. "KNN Based Machine Learning Approach for Text and Document Mining." *International Journal of Database Theory and Application* 7 (1): 61–70. <https://doi.org/10.14257/ijdta.2014.7.1.06>.

## Cross-disciplinarity

- Bao, Yang, Nigel Collier, and Anindya Datta. 2013. "A Partially Supervised Cross-Collection Topic Model for Cross-Domain Text Classification." In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*, 239–248. CIKM '13. New York, NY, USA: ACM. <https://doi.org/10.1145/2505515.2505556>.
- Beghtol, C. 19980101. "Knowledge Domains: Multidisciplinarity and Bibliographic Classification Systems." *Knowledge Organization* 25 (1).
- Denda, Kayo. 2005. "Beyond Subject Headings." *Library Resources & Technical Services* 49 (4): 266–275. <https://doi.org/10.5860/lrts.49n4.266>.
- Jay, Caroline, Simon Harper, Ian Dunlop, Sam Smith, Shoaib Sufi, Carole Goble, and Iain Buchan. 2016. "Natural Language Search Interfaces: Health Data Needs Single-Field Variable Search." *Journal of Medical Internet Research* 18 (1): 1–21. <https://doi.org/10.2196/jmir.4912>.
- "Knowledge Management Section Joint with Academic and Research Libraries Section and Rare Books and Special Collections Section." n.d. Google Docs. Accessed September 19, 2018. <https://docs.google.com/document/d/1kII0CqGITORXXaPHPHXM0ksLarCHPIacJ5yDiSeZJ7o/ed>

[it?usp=drive\\_web&ouid=100690309179145782813&usp=embed\\_facebook](https://doi.org/10.28945/4025).

Lebow, David G. 2018. "A Social Machine for Transdisciplinary Research." *Informing Science* 21 (January): 201–17. <https://doi.org/10.28945/4025>.

Li, Lianghao, Xiaoming Jin, and Mingsheng Long. n.d. "Topic Correlation Analysis for Cross-Domain Text Classification," 7.

MLS, Linda G. Ackerson. 2001. "Challenges for Engineering Libraries." *Science & Technology Libraries* 21 (1–2): 43–52. [https://doi.org/10.1300/J122v21n01\\_05](https://doi.org/10.1300/J122v21n01_05).

Sabharwal, Arjun. 2011. "Knowledge Domain Navigation in Interdisciplinary Digital Landscapes." *Journal of Library Metadata* 11 (2): 63–82. <https://doi.org/10.1080/19386389.2011.570655>.

## Discovery

Campbell, D. Grant, and Karl V. Fast. 2004. "Academic Libraries and the Semantic Web: What the Future May Hold for Research-Supporting Library Catalogues." *The Journal of Academic Librarianship* 30 (5): 382–390. <https://doi.org/10.1016/j.acalib.2004.06.007>.

Gross, Tina, Arlene G. Taylor, and Daniel N. Joudrey. 2015. "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching." *Cataloging & Classification Quarterly* 53 (1): 1–39. <https://doi.org/10.1080/01639374.2014.917447>.

Jay, Caroline, Simon Harper, Ian Dunlop, Sam Smith, Shoaib Sufi, Carole Goble, and Iain Buchan. 2016. "Natural Language Search Interfaces: Health Data Needs Single-Field Variable Search." *Journal of Medical Internet Research* 18 (1): 1–21. <https://doi.org/10.2196/jmir.4912>.

Li, Ying, Dick R Miller, and Mary Buttner. 2002. "Bibliographic Data Mining: Automatically Building Component Part Records for e-Journal Articles on the Internet." *Journal of Internet Cataloging* 5 (1): 29–41.

## Machine Learning

Brygfjeld, Svein Arne, Freddy Wetjen, and André WalsØe. n.d. "Machine Learning for Production of Dewey Decimal." In . <http://library.ifla.org/2216/1/115-brygfjeld-en.pdf>.

Dědek, Jan, Peter Vojtáš, and Marta Vomlelová. 2012. "Fuzzy ILP Classification of Web Reports after Linguistic Text Mining." *Information Processing and Management* 48 (3): 438–450. <https://doi.org/10.1016/j.ipm.2011.02.008>.

"Knowledge Management Section Joint with Academic and Research Libraries Section and Rare Books and Special Collections Section." n.d. Google Docs. Accessed September 19, 2018. <https://docs.google.com/document/d/1kII0CqGITORXXaPHPHXM0ksLarCHPIacJ5yDiSeZJ7o/ed> [it?usp=drive\\_web&ouid=100690309179145782813&usp=embed\\_facebook](https://doi.org/10.28945/4025).

Stein, Roger Alan, Patricia A. Jaques, and João Francisco Valiati. 2019. "An Analysis of Hierarchical Text Classification Using Word Embeddings." *Information Sciences* 471: 216–32.

<https://doi.org/10.1016/j.ins.2018.09.001>.

Vishwanath Bijalwan, Vinay Kumar, Pinki Kumari, and Jordan Pascual. 2014. "KNN Based Machine Learning Approach for Text and Document Mining." *International Journal of Database Theory and Application* 7 (1): 61–70. <https://doi.org/10.14257/ijdta.2014.7.1.06>.

Zhiyuan Chen [Computer scientist. 2017. "Lifelong Machine Learning." San Rafael, California]: Morgan & Claypool.

<http://proxy.library.nd.edu/login?url=http://dx.doi.org/10.2200/S00737ED1V01Y201610AIM033>.

## Natural Language Processing

Bittermann, André, and Andreas Fischer. 2018. "How to Identify Hot Topics in Psychology Using Topic Modeling." *Zeitschrift Für Psychologie* 226 (1): 3–13. <https://doi.org/10.1027/2151-2604/a000318>.

Danilevsky, M., C. Wang, N. Desai, X. Ren, J. Guo, and J. Han. 2014. "Automatic Construction and Ranking of Topical Keyphrases on Collections of Short Documents." In *Proceedings of the 2014 SIAM International Conference on Data Mining*, 398–406. Proceedings. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611973440.46>.

Friedman, Carol, Tara Borlawsky, Lyudmila Shagina, H Xing, and Yves Lussier. 2006. "Bio-Ontology and Text: Bridging the Modeling Gap." *Bioinformatics* 22 (19): 2421–29. <https://doi.org/10.1093/bioinformatics/btl405>.

Hagedorn, Kat, Michael Kargela, Youn Noh, and David Newman. 2011. "A New Way to Find: Testing the Use of Clustering Topics in Digital Libraries." *D-Lib Magazine* 17 (9/10). <https://doi.org/10.1045/september2011-hagedorn>.

Joo, S, I Choi, and N Choi. 2018. "Topic Analysis of the Research Domain in Knowledge Organization: A Latent Dirichlet Allocation Approach." *Knowl. Organ.* 45 (2): 170–83. <https://doi.org/10.5771/0943-7444-2018-2-170>.

Mimno, David, and Andrew McCallum. 2007. "Organizing the OCA: Learning Faceted Subjects from a Library of Digital Books." In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, 376–385. JCDL '07. New York, NY, USA: ACM. <https://doi.org/10.1145/1255175.1255249>.

Rajagopal, Dheeraj, Daniel Olsher, Erik Cambria, and Kenneth Kwok. 2013. "Commonsense-Based Topic Modeling." In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, 6:1–6:8. WISDOM '13. New York, NY, USA: ACM. <https://doi.org/10.1145/2502069.2502075>.

## Topic Modeling

Bao, Yang, Nigel Collier, and Anindya Datta. 2013. "A Partially Supervised Cross-Collection Topic Model for Cross-Domain Text Classification." In *Proceedings of the 22Nd ACM International*



- Conference on Information & Knowledge Management, 239–248. CIKM '13. New York, NY, USA: ACM. <https://doi.org/10.1145/2505515.2505556>.
- Bittermann, André, and Andreas Fischer. 2018. "How to Identify Hot Topics in Psychology Using Topic Modeling." *Zeitschrift Für Psychologie* 226 (1): 3–13. <https://doi.org/10.1027/2151-2604/a000318>.
- Cain, Jonathan O. 2016. "Using Topic Modeling to Enhance Access to Library Digital Collections." *Journal of Web Librarianship* 10 (3): 210–25. <https://doi.org/10.1080/19322909.2016.1193455>.
- Danilevsky, M., C. Wang, N. Desai, X. Ren, J. Guo, and J. Han. 2014. "Automatic Construction and Ranking of Topical Keyphrases on Collections of Short Documents." In *Proceedings of the 2014 SIAM International Conference on Data Mining*, 398–406. Proceedings. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611973440.46>.
- Goldstone, Andrew. n.d. "PMLA Topic Model Browser." Accessed November 13, 2018. <http://agoldst.github.io/dfr-browser/demo/>.
- Hagedorn, Kat, Michael Kargela, Youn Noh, and David Newman. 2011. "A New Way to Find: Testing the Use of Clustering Topics in Digital Libraries." *D-Lib Magazine* 17 (9/10). <https://doi.org/10.1045/september2011-hagedorn>.
- Joo, S, I Choi, and N Choi. 2018. "Topic Analysis of the Research Domain in Knowledge Organization: A Latent Dirichlet Allocation Approach." *Knowl. Organ.* 45 (2): 170–83. <https://doi.org/10.5771/0943-7444-2018-2-170>.
- "Knowledge Management Section Joint with Academic and Research Libraries Section and Rare Books and Special Collections Section." n.d. Google Docs. Accessed September 19, 2018. [https://docs.google.com/document/d/1kII0CqGITORXXaPHPHXM0ksLarCHPIacJ5yDiSeZJ7o/edit?usp=drive\\_web&ouid=100690309179145782813&usp=embed\\_facebook](https://docs.google.com/document/d/1kII0CqGITORXXaPHPHXM0ksLarCHPIacJ5yDiSeZJ7o/edit?usp=drive_web&ouid=100690309179145782813&usp=embed_facebook).
- Li, Lianghao, Xiaoming Jin, and Mingsheng Long. n.d. "Topic Correlation Analysis for Cross-Domain Text Classification," 7.
- Mimno, David, and Andrew McCallum. 2007. "Organizing the OCA: Learning Faceted Subjects from a Library of Digital Books." In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, 376–385. JCDL '07. New York, NY, USA: ACM. <https://doi.org/10.1145/1255175.1255249>.
- Newman, David, Kat Hagedorn, Chaitanya Chemudugunta, and Padhraic Smyth. 2007. "Subject Metadata Enrichment Using Statistical Topic Models." In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, 366–375. JCDL '07. New York, NY, USA: ACM. <https://doi.org/10.1145/1255175.1255248>.
- Rajagopal, Dheeraj, Daniel Olsher, Erik Cambria, and Kenneth Kwok. 2013. "Commonsense-Based Topic Modeling." In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, 6:1–6:8. WISDOM '13. New York, NY, USA: ACM. <https://doi.org/10.1145/2502069.2502075>.
- Zhang, Xiang, Junbo Zhao, and Yann LeCun. n.d. "Character-Level Convolutional Networks for Text Classification." arXiv.org. <https://arxiv.org/pdf/1509.01626v3.pdf>.

## **Appendix B: Survey**

### **Questions**

*Start of Block: Introduction*

#### **Q1.1 IMLS/University of Notre Dame Survey: Machine Learning for Cross-Disciplinary Research**

##### **Background & Grant Overview**

The Institute of Museum and Library Services (IMLS) has awarded The University of Notre Dame a one-year planning grant to assess the need for library-based machine learning and natural language processing tools to facilitate automated metadata creation and classification in support of cross-disciplinary discovery and research.

##### **Project Outcomes**

During 2019, the Hesburgh Libraries grant team will develop several in-person and online opportunities designed to stimulate national engagement with and assess stakeholder demand for these proposed tools. The grant team intends to: facilitate conversation among scholars, computer scientists, and librarians reconcile domain-specific, best approaches to support cross-disciplinary discovery determine how to provide a framework for librarians to engage in innovative approaches support research to ascertain the national need for library-based automated tools that support cross-disciplinary research identify and solidify partnerships for tool development in support of this need.

##### **Purpose of This Survey**

The purpose of this initial survey is to conduct an environmental scan of the academic landscape and to identify Higher Ed community members interested in a nation-wide conversation on machine learning in cross-disciplinary discovery. We hope to learn if and how others are harnessing concepts of machine learning to aid in cross-disciplinary research. We encourage you to respond, even if your thoughts about a more efficient way to encounter disparate bodies of knowledge for research are still in the formative stage. Diversity is important to our outcome. We welcome responses from institutions of all sizes and resource levels, regardless of your experience with topic modeling or machine learning.



### **How We Will Use Your Information**

We will use personally identifiable information only in the selections/recruitment of participants in the workshops/research.

### **Other survey responses may be identifiable by demographics and will be used to:**

stimulate workshop discussions to inform a white paper that will be published as part of summative evaluation of the workshops help us determine the need for machine learning and topic modeling tools for cross-disciplinary research.

Non-personally identifiable information will be stored in our institutional repository as an artifact of our research. Personally identifiable information will be stored in a secure location for the project duration + 12 months, after which it will be permanently deleted.

### **How Long Will This Take?**

Survey completion time will vary depending on individual experience but should take no longer than 10 minutes. Please respond to each question to the best of your ability.

Thank You

Your input is crucial to the success of this project. We appreciate your cooperation. For more information, please see [the project website](#).

### **Q1.2 What is your primary affiliation with your university/institution?**

- ☐ Faculty (teaching and/or research) (1)
- ☐ Librarian (2)
- ☐ Computer scientist/engineer (3)
- ☐ Other (please specify) (4) \_\_\_\_\_

*End of Block: Introduction*

*Start of Block: Questions for All*

**Q2.1 For this survey, cross-disciplinary research is any research or work that involves more than one field of knowledge.**

Examples of cross-disciplinary research are papers written jointly by biology and anthropology professors, grants that involve computer scientists, psychologists, and statisticians, or studies that include both psychologists and political scientists.

**Q2.2 In this survey, machine learning refers to any data analysis method where a computer builds a model and continually improves the performance of the model by optimizing for a specific task.**

For example, linear or logistic regression are not considered machine learning, but neural networks, CART (C4.5, trees), and SVM (support vector machines) are considered machine learning.

**Q2.3 Have you used machine learning in your research?**

- ☐ Yes (1)
- ☐ No (2)

*Display This Question: If Have you used machine learning in your research? = Yes*

**Q2.4 For what purposes have you used machine learning?**

- ☐ Supervised Learning (classification or identification) (1)
- ☐ Unsupervised Learning (topic modeling) (2)
- ☐ Reinforcement Learning (neural networks for image analysis or parts-of-speech analysis, etc.) (3)
- ☐ Other (please specify) (4) \_\_\_\_\_

**Q2.5 How familiar are you with machine learning?**

- ☐ Extremely familiar (1)
- ☐ Very familiar (2)
- ☐ Moderately familiar (3)
- ☐ Slightly familiar (4)
- ☐ Not familiar at all (5)

**Q2.6 Do you think machine learning can be used to enhance cross-disciplinary research?**

- ☐ Yes (1)
- ☐ No (2)
- ☐ Maybe / Not sure (3)

*Display This Question: If Do you think machine learning can be used to enhance cross-disciplinary research? = Yes*

**Q2.7 Please describe a potential use case for applying machine learning to facilitate or enhance cross-disciplinary research.**

---

---

---

---

*Display This Question: If Do you think machine learning can be used to enhance cross-disciplinary research? = No*

**Q2.8 Why do you think machine learning won't enhance cross-disciplinary research?**

---

---

---

---

*End of Block: Questions for All*

*Start of Block: Questions for Researchers*

Display This Question: If What is your primary affiliation with your university/institution? = Faculty (teaching and/or research)

**Q3.1 Do you collaborate with scholars from other disciplines in your research?**

- ☐ Yes (1)
- ☐ No (2)

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q3.2 For what disciplines are you producing cross-disciplinary research? (Hold Ctrl or Cmd key to make multiple selections.)**

- ☐ Humanities: Fine Arts (1)
- ☐ Humanities: History (11)
- ☐ Humanities: Languages & Literature (12)
- ☐ Humanities: Philosophy/Theology (13)
- ☐ Humanities: Other (14)
- ☐ Social Sciences: Sociology/Psychology (15)
- ☐ Social Sciences: Business/Economics (16)
- ☐ Social Sciences: Law (17)
- ☐ Social Sciences: Political Science (18)
- ☐ Social Sciences: Other (19)
- ☐ Science: Physics/Astronomy (20)
- ☐ Science: Chemistry (21)
- ☐ Science: Engineering (22)
- ☐ Science: Life Sciences/Medicine (23)
- ☐ Science: Other (24)

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q3.3 In two or three sentences, please describe your cross-disciplinary research.**

---

---

---

---

---

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q3.4 When doing cross-disciplinary research, what obstacles do you usually encounter?**

---

---

---

---

---

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q3.5 How do you currently address these obstacles?**

---

---

---

---

---

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q3.6 Does your university/institution have resources to handle/address these issues?**

- ☐ Yes (1)
- ☐ Maybe (2)
- ☐ No (3)

*Display This Question: If What is your primary affiliation with your university/institution? = Faculty (teaching and/or research)*

**Q3.7 Which of the following resources do you have available to you for use in your research?**

**Please select all that apply.**

- ☐ Grant Writing Support (1)
- ☐ High Performance Computing / Computer Lab(s) (2)
- ☐ Statistical Support (3)
- ☐ None of the above (6)
- ☐ Other (4) \_\_\_\_\_

Display This Question: If What is your primary affiliation with your university/institution? = Faculty (teaching and/or research)

**Q3.8 To what degree do you agree that terminology (jargon) is an obstacle to cross-disciplinary research?**

- ☐ Strongly agree (1)
- ☐ Somewhat agree (2)
- ☐ Neither agree nor disagree (3)
- ☐ Somewhat disagree (4)
- ☐ Strongly disagree (5)

Display This Question: If What is your primary affiliation with your university/institution? = Faculty (teaching and/or research)

**Q3.9 How likely are you to collaborate with people from each of the following areas:**

	Extremely likely (1)	Somewhat likely (2)	Neither likely nor unlikely (3)	Somewhat unlikely (4)	Extremely unlikely (5)
Librarians (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Computer Scientists/engineer s (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Humanists / Social scientists / Other sciences (13)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other types of scholars (15)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: Questions for Researchers

*Start of Block: Questions for Librarians*

*Display This Question: If What is your primary affiliation with your university/institution? = Librarian*

**Q4.1 Do you collaborate with scholars from other disciplines in your research?**

- ☐ Yes (1)
- ☐ No (2)

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q4.2 For what disciplines are you producing cross-disciplinary research? (Hold Ctrl or Cmd key to make multiple selections.)**

- ☐ Humanities: Fine Arts (1)
- ☐ Humanities: History (11)
- ☐ Humanities: Languages & Literature (12)
- ☐ Humanities: Philosophy/Theology (13)
- ☐ Humanities: Other (14)
- ☐ Social Sciences: Sociology/Psychology (15)
- ☐ Social Sciences: Business/Economics (16)
- ☐ Social Sciences: Law (17)
- ☐ Social Sciences: Political Science (18)
- ☐ Social Sciences: Other (19)
- ☐ Science: Physics/Astronomy (20)
- ☐ Science: Chemistry (21)
- ☐ Science: Engineering (22)
- ☐ Science: Life Sciences/Medicine (23)
- ☐ Science: Other (24)

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q4.3 In two or three sentences, please describe your cross-disciplinary research.**

---

---

---

---

---



*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q4.4 What is your role in facilitating cross-disciplinary research/scholarship? Please select all that apply.**

- ☐ Identify research content (1)
- ☐ Describe/catalog/tag (2)
- ☐ Recommend metadata (3)
- ☐ Curate collection (4)
- ☐ Write programs/applications (5)
- ☐ Provide subject expertise (6)
- ☐ Connect people/programs/centers (7)
- ☐ Other (please specify) (8) \_\_\_\_\_

*Display This Question: If What is your primary affiliation with your university/institution? = Librarian*

**Q4.5 To what degree do you agree that terminology (jargon) is an obstacle to cross-disciplinary research?**

- ☐ Strongly agree (1)
- ☐ Somewhat agree (2)
- ☐ Neither agree nor disagree (3)
- ☐ Somewhat disagree (4)
- ☐ Strongly disagree (5)

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q4.6 When doing cross-disciplinary research/scholarship, what obstacles do you usually encounter?**

---

---

---

---

---

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q4.7 How do you currently address these obstacles?**

---

---

---

---

---

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q4.8 Does your library have resources to handle/address these obstacles?**

- ☐ Yes (1)
- ☐ Maybe (2)
- ☐ No (3)

*Display This Question: If What is your primary affiliation with your university/institution? = Librarian*

**Q4.9 Which of the following resources do you have available to you for use in your research?  
Please select all that apply.**

- ☐ Grant Writing Support (1)
- ☐ High Powered Computer(s)/ Computer Lab(s) (2)
- ☐ Statistical Support for Research (3)
- ☐ None of the above (11)
- ☐ Other (please specify) (4) \_\_\_\_\_

*Display This Question: If What is your primary affiliation with your university/institution? = Librarian*

**Q4.10 How likely are you to collaborate with people from each of the following areas:**

	Extremely likely (1)	Somewhat likely (2)	Neither likely nor unlikely (3)	Somewhat unlikely (4)	Extremely unlikely (5)
Librarians (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Computer Scientists/engineer s (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Humanists / Social scientists / Other sciences (13)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (11)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*End of Block: Questions for Librarians*

*Start of Block: Questions for Computer Scientists/Engineers*

*Display This Question: If What is your primary affiliation with your university/institution? = Computer scientist/engineer*

**Q5.1 Do you collaborate with scholars from other disciplines in your research?**

- ☐ Yes (1)
- ☐ No (2)

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q5.2 For what disciplines are you producing cross-disciplinary research?**

- ☐ Humanities: Fine Arts (1)
- ☐ Humanities: History (11)
- ☐ Humanities: Languages & Literature (12)
- ☐ Humanities: Philosophy/Theology (13)
- ☐ Humanities: Other (14)
- ☐ Social Sciences: Sociology/Psychology (15)
- ☐ Social Sciences: Business/Economics (16)
- ☐ Social Sciences: Law (17)
- ☐ Social Sciences: Political Science (18)
- ☐ Social Sciences: Other (19)
- ☐ Science: Physics/Astronomy (20)
- ☐ Science: Chemistry (21)
- ☐ Science: Engineering (22)
- ☐ Science: Life Sciences/Medicine (23)
- ☐ Science: Other (24)

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q5.3 In two or three sentences, please describe your cross-disciplinary research.**

---

---

---

---

---

*Display This Question: If What is your primary affiliation with your university/institution? = Computer scientist/engineer*

**Q5.4 What is your role in facilitating cross-disciplinary research/scholarship? Please select all that apply. Please select all that apply.**

- ☐ Theory (1)
- ☐ Modeling/visualization (2)
- ☐ Software development (3)
- ☐ Consulting (4)
- ☐ Training (5)
- ☐ Other (please specify) (6) \_\_\_\_\_

*Display This Question: If What is your primary affiliation with your university/institution? = Computer scientist/engineer*

**Q5.5 To what degree do you agree that terminology (jargon) is an obstacle to cross-disciplinary research?**

- ☐ Strongly agree (1)
- ☐ Somewhat agree (2)
- ☐ Neither agree nor disagree (3)
- ☐ Somewhat disagree (4)
- ☐ Strongly disagree (5)

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q5.6 When doing cross-disciplinary research/scholarship, what obstacles do you usually encounter?**

---

---

---

---

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q5.7 How do you currently address these obstacles?**

---

---

---

---

*Display This Question: If Do you collaborate with scholars from other disciplines in your research? = Yes*

**Q5.8 Does your university/institution have resources to handle/address these issues?**

- ☐ Yes (1)
- ☐ Maybe (2)
- ☐ No (3)

*Display This Question: If What is your primary affiliation with your university/institution? = Computer scientist/engineer*

**Q5.9 Which of the following resources do you have available to you for use in your research?  
Please select all that apply.**

- ☐ Grant Writing Support (1)
- ☐ High Powered Computer(s)/ Computer Lab(s) (2)
- ☐ Statistical Support for Research (3)
- ☐ None of the above (5)
- ☐ Other (please specify) (4) \_\_\_\_\_

*Display This Question: If What is your primary affiliation with your university/institution? = Computer scientist/engineer*

**Q5.10 How likely are you to collaborate with people from each of the following areas:**

	Extremely likely (1)	Somewhat likely (2)	Neither likely nor unlikely (3)	Somewhat unlikely (4)	Extremely unlikely (5)
Librarians (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Computer Scientists/engineer s (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Humanists / Social scientists / Other sciences (13)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (11)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*End of Block: Questions for Computer Scientists/Engineers*



*Start of Block: Demographics*

Display *This Question*: If invisible != great

**Q6.1 Please disregard this question because the workshops have concluded, but if we here at Notre Dame were to delete this question altogether, then our data analysis process would break. Would you be interested in participating in a workshop that will discuss machine learning and/or topic modeling? (Workshop registrations are limited.)**

- ☐ Yes (1)
- ☐ No (2)

Display *This Question*: If invisible != great

**Q6.2 Please disregard this question because the workshops have concluded, but if we here at Notre Dame were to delete this question altogether, then our data analysis process would break. Would you be interested in participating in an online discussion group that will discuss machine learning and/or topic modeling?**

- ☐ Yes (1)
- ☐ No (2)

Display *This Question*: If invisible != great

**Q6.3 Please disregard this question because the workshops have concluded, but if we here at Notre Dame were to delete this question altogether, then our data analysis process would break. How would you like to participate?**

- ☐ I am interested in presenting (1)
- ☐ I am interested in attending and contributing to conversations (3)

Display *This Question*: If invisible != great

**Q6.4 Please disregard this question because the workshops have concluded, but if we here at Notre Dame were to delete this question altogether, then our data analysis process would break. Please provide a brief (3-4 sentence) description of a proposed presentation**

---

---

---

---

---

*Display This Question: If invisible != great*

**Q6.5 Please disregard this question because the workshops have concluded, but if we here at Notre Dame were to delete this question altogether, then our data analysis process would break. Please select the workshop(s) you are interested in attending. (Please note: Notre Dame workshop is full)**

- ☐ FULL - Palo Alto, CA (April 1, 2019) (2)
- ☐ New York City, NY (April 25, 2019) (3)
- ☐ Washington, DC (May 31, 2019) (4)

**Q6.6 What is your name?**

---

**Q6.7 For what institution do you work?**

---

---

---

---

---

**Q6.8 What is your preferred email address?**

---

**Q6.9 What is your job title?**

---

*Display This Question: If invisible != great*

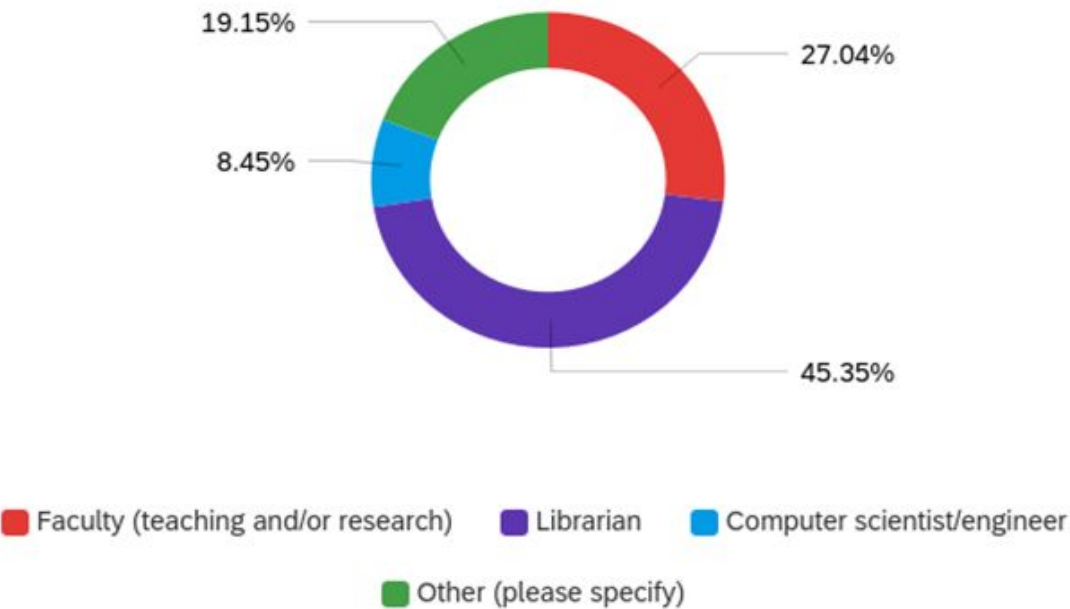
**Q6.10** If you would like to recommend a colleague to participate in the workshop, please fill in their information below.

	Name (1)	Email (2)	Affiliation (3)
Colleague 1 (8)			
Colleague 2 (9)			
Colleague 3 (10)			

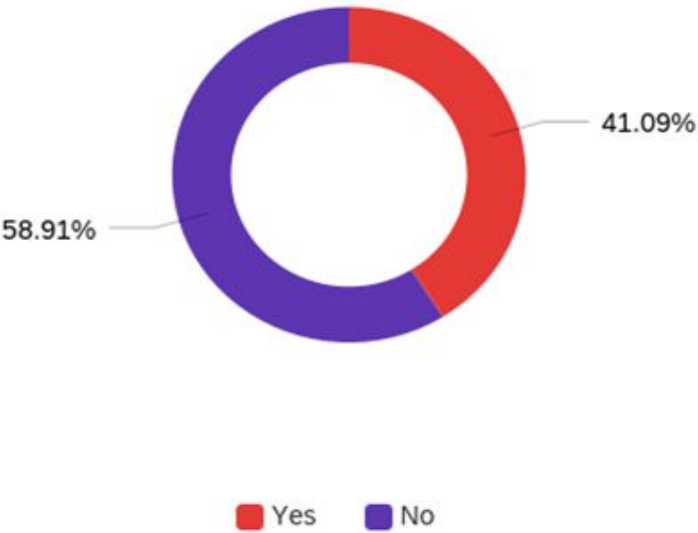
*End of Block: Demographics*

Responses (Anonymized to protect respondent privacy.)

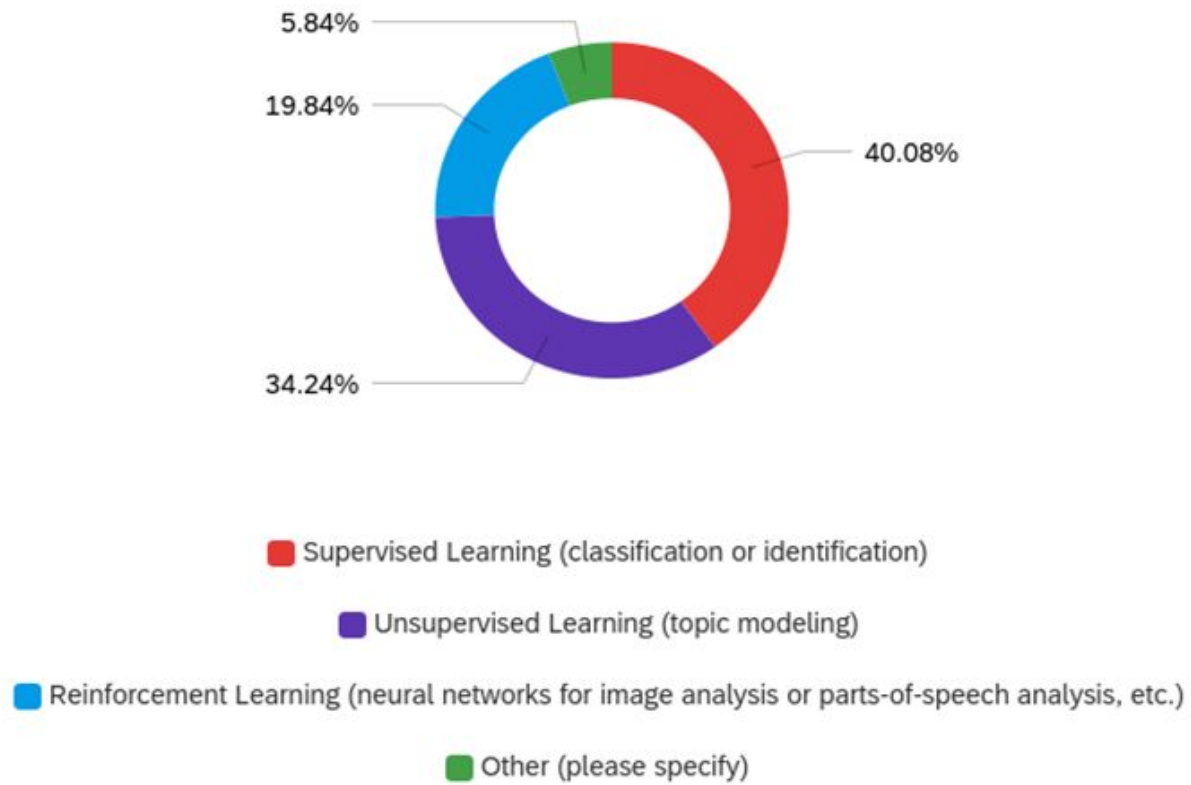
Everybody - What is your primary affiliation with your university/institution?



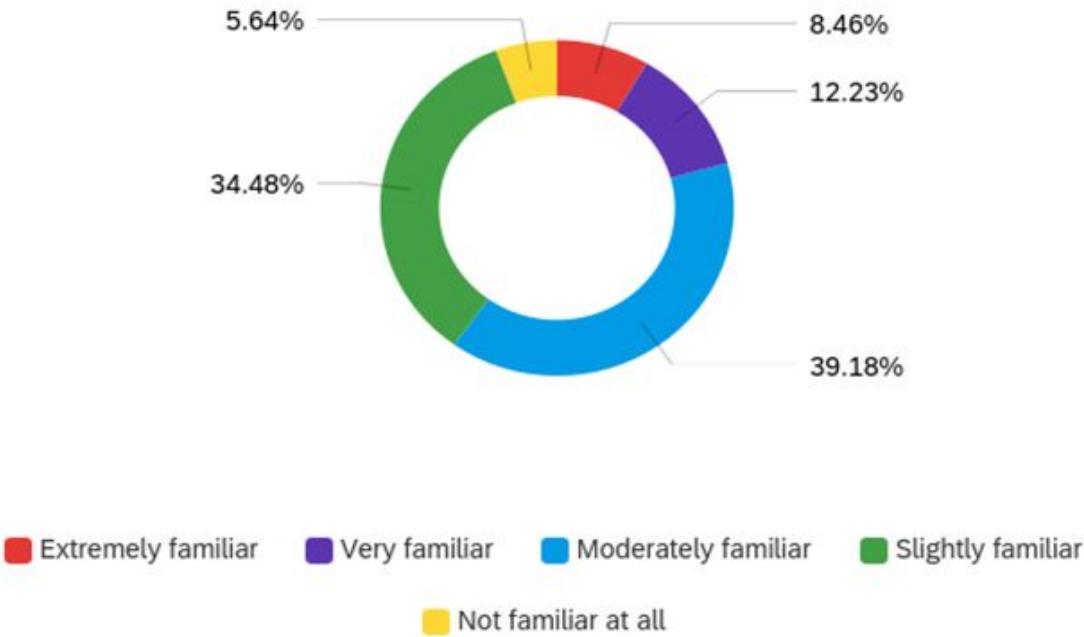
Everybody - Have you used machine learning in your research?



**Everybody - For what purposes have you used machine learning?**

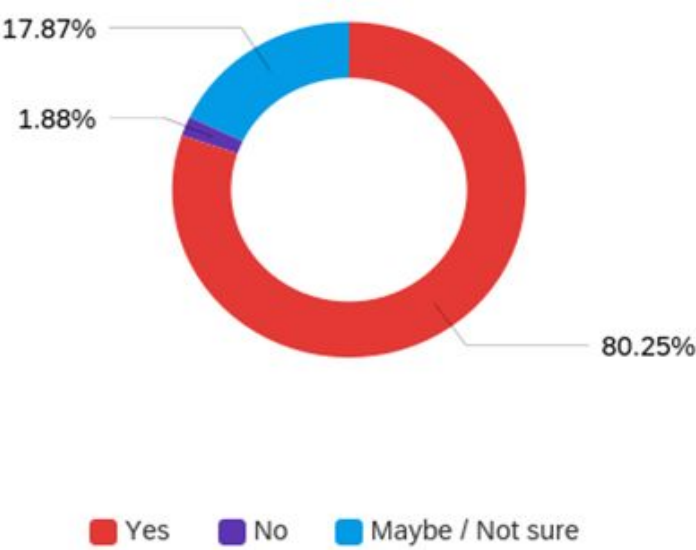


Everybody - How familiar are you with machine learning?





Everybody - Do you think machine learning can be used to enhance cross-disciplinary research?



**Everybody - Please describe a potential use case for applying machine learning to facilitate or enhance cross-disciplinary research. (With personally identifiable information removed for respondent privacy.)**

I wrote [a web based tool], which lets people explore graduate theses by conceptual similarity. The algorithm can support the following cross-disciplinary possibilities (not all implemented): \* if the "most similar" theses to a given thesis happen to be in a different department, they will be surfaced -- this is not possible to do with the existing metadata for the underlying collection \* data visualization of the clusters of theses would let people see where there are clusters which contain different departments (e.g. with color-coding) \* researchers can find other researchers whose work, or whose supervisees' work, is "near" theirs -- identifying possible collaborators \* it's possible to look at the words "near" a given word in meaning in different disciplines, and thereby get a sense of how those disciplines might use the word to mean different things (see <https://mitlibraries.github.io/ml2s/2017/07/06/six-ways-of-looking-at-oxygen.html> )

---

I work at a large museum with a large amount of digitized material, catalog data, and name data extracted from historical records. I can think of many opportunities for machine learning. There are many use cases that I can imagine now and surely many that have not been imagined by anyone yet. Some sets of areas relate to computer vision. e.g. for use against digitized historical records; speech to text relating to oral histories or other audio or video sources; textual analysis related to catalog record or text extracted from historical records, or name related data extracted from historical records. My institution has had some visiting scholars and/or digital humanists who have done some experiments, which has stimulated greater interest among people here. We also now have scholars outside history to include sociologists, anthropologists, ethnologists, and the interactions and collaborations among computer scientists and collections specialists with computer learning tools, I firmly believe will expand. It's very early and the opportunities are just beginning to become more widely understood.

---

I do not believe I'm your target audience for this survey. The use case i am primarily interested in is regarding cultural heritage metadata. We are investigating opportunities to use ML to make smarter suggestions for values in metadata fields; we are investigating incorporating this into batch ingest workflows. I suppose I could make an argument for how this would inform cross-disciplinary research, but our primary concern is ML assisted workflows for knowledge workers. A large consortial repository system would be a good place to do this work.

---

---

Cultural studies is a highly interdisciplinary field which examines problematics (or 'social crises of sorts') across the social formation, examining the articulation of social and material forces in the realms of the economic, technological, juridico-political, etc. Such mapping work could be enhanced by building computer models to study trends of adoption of certain policies or ideologies, say, to verify the veracity of truth claims made by those researchers who heretofore only used qualitative and lightly quantitative models.

---

duplicate record detection

---

Data modeling and entity extraction from full text documents, followed by mapping terms to a cross-disciplinary controlled vocabulary..

---

Supervised machine learning can be used to assign subject headings and other controlled vocabulary within library catalog and metadata workflows. While these tend to be tricky multi-label classification problems, with enough data and thoughtful preprocessing, there is much potential here. Meanwhile, unsupervised learning probably should be an essential part of actually creating such vocabularies, especially when dealing with larger, time-sensitive, text corpora. By automating catalog, repository, and even data asset processes we are presumably facilitating cross-disciplinary research through enhanced discoverability and access.

---

There are many different areas. Training a network to identify inputs for flight control in order to quickly compensate for failing subsystems. Networks can be used for diagnosing ailments with inputs from a series of tests, or to recommend additional tests in order to create a better diagnosis. I can envision using this system to ensure that doctors are consistent in their diagnosis and not necessary to replace them. Networks used to classify the kind of customer and to anticipate the products they would be interested in, maybe as suggestions to helpers in a brick and mortar store. There are possible potentials in education in order to tailor the learning experience to optimize for a specific learning style. Maybe can pair advisors/mentors to specific students based on matching learning and teaching styles. Each of these directions depend on a wealth of information that is used to train a network to identify

---

ML for ontology generation

---

---

A potential use case: improving text classification using fiction One could take a diverse fiction corpus from the Romantic period and use human labelers to determine instances in the texts where nationalism is being addressed as either positive, negative, or ambiguous. A support vector machine could then be trained on examples of positive, negative, and inconclusive sentiments about nationalism. The supervised model could then be applied to a larger set of texts from the early 19th century, such as periodicals, to assist with research that seeks examples of the complex origins of nationalistic concepts in a dynamic political economy, with attention to its reliance on colonialism. This type of research could potentially take the insights a literary scholar might have in identifying ambiguity and irony in texts, and generate new results and insights that would be helpful to historians or post-colonial theorists.

---

ML applied to legal text can enhance analytical approaches to law often seen in the law & economics literature

---

Use of machine learning to classify large amounts of documents that will help librarians to understand better the collection and will improve knowledge extraction and data organization.

---

My interest is currently focused on the application of ML and AI to support archival workflows, including the identification of "sensitive" information via classifiers (as screening for this information often comprises a key aspect of processing those collections), and the development of the new models of discovery and access by cultural memory organizations. I would love to test the potential for developing training sets to support description of archival materials based on work already accomplished by metadata experts. For instance, it would be great to provide a service to archivists that would attempt to provide an approximate date and identify photographic processes applied to uploaded digitized photographs, studio stamps etc., as well as link to similar photographs held by other organizations, etc.

---

Sentiment analysis of historical documents, which can be used both for insights into historical processes and to improve the algorithms of the sentiment analysis software for computer science.

---

One of my projects is using it for this purpose at the moment. We have a project that involves linguists, literature specialists, archaeologists, anthropologists and computer scientists. One of the literature guys is training a model to automatically identify data that may be of interest to the linguists from among documents he is working with for other purposes.

---

---

any research program involving scholars+digital corpora needs digital tools to facilitate corpora building, analysis of such corporas, etc. Machine learning can help to such tasks

---

Machine learning (ML) techniques might be utilized with Natural Language Processing (NLP) to identify correlation between currently unconnected bodies of research. For example, digital humanities and economics are both engaging in digital scholarship by creating online resources their audiences, occasionally with GIS output. Despite the similarities in project scope--form and content--the two domains are entirely disconnected. NLP might create a foundation for communication between the research domains, in an effort to build upon and reuse projects.

---

study of large data that could then be visualized to tell a story

---

Implementing natural language processing with machine learning for the purposes of systematic reviews of medical research within and beyond the most obviously relevant sub-disciplines.

---

My expertise in machine learning lies in the use of topic modeling in understanding large collection of unstructured text, such as a collection of Shakespeare plays, specifically combining sentiment analysis in the modeling process of topic models thus factoring sentiment values in the semantic relatedness of words. Another huge part of my research is in topic model diagnostics and visualization, where we are hoping to create visualizations and metrics to help inter-disciplinary researchers to better use topic models and understand the results.

---

---

In [an article I wrote], I worked with a colleague with a background in linguistics, mathematics, and digital humanities to develop an example of how distributional semantic modelling (a machine learning technique for building semantic representations based on analysis of large-scale corpora) could be applied to the extraction of implied references to a target concept from a historical corpus. The work was conceived as an early-stage prototype for how this type of modelling could aid a historian searching a very large collection of domain-specific but also heterogeneous textual data. More generally, I would like to explore ways to apply context-specific lexical semantic representations generated using machine learning techniques to the analysis of phenomena such as semantic change (the way word meaning develops historically across a diachronic corpus), and to investigate how (and even if) this type of analysis can help researchers develop critical perspectives on large-scale historical and cultural heritage data. The hypothesis I would like to test is the idea that projections from high-dimensional representational spaces can be mapped to critical perspectives, allowing researchers to interact with large-scale data in more nuanced ways.

---

Machine learning is useful for modeling and analyzing data from many different sources, including the humanities, social sciences, and physical sciences. In some cases, the machine learning principles can be understood and applied without any interdisciplinary collaboration, but in others, the machine learning methods can be refined and improved through collaboration with computer science.

---

Bioinformatics could use machine learning to predict the likelihood of disease.

---

Social anthropologist who's a domain expert in a locale (Hangzhou) wants to understand economic theory (from agrarian to industrial, as we're referring to mainland China here) and apply it to their research. AI suggests an interview panel (for firsthand research). AI suggests high quality interviewees—these interviewees would be outside social anthropology. They'd be cross-disciplinary, and enable broad understanding of economic theory in mainland China over time. This expands learning, reducing silo-ization of thought

---

I am interested in making my textual language archive in an indigenous language more accessible to my community and to scholars interested in doing more work in this archive who may need assistance with translation. I am working [a particular scholarly model], moving the oral and textual into different digital formats. I foresee collaboration between myself as a historian and professors in law and governance, environment and climate change, linguistics, literature, ocean science. The archive I work in is the largest indigenous language textual archive in native North America and the Polynesian Pacific.

---

---

Several groups are using machine learning to classify data from neural recordings (usually invasive brain-surface or spinal-cord recordings) for: (1) identifying neuronal activity to use as a control signal for brain-computer interfaces, (2) identifying biomarkers of a disorder, (3) investigating how neural circuits work. As such, the use of machine learning in neural engineering helps bring together electrical engineers, computer scientists, neuroscientists, neurosurgeons, applied mathematicians, and so on. I work primarily as a moral philosopher in such multidisciplinary settings, and this is another avenue for machine learning to influence cross-disciplinary research. Of course moral philosophy is tasked with determining the complications that arise from the use of machine learning. The converse is also becoming more common: more social scientists and philosopher using machine learning algorithms to answer empirical questions about moral problems. For example, researchers who used natural language processing to create models of respectful language to determine if people of color are treated disrespectfully by law enforcement officers.

---

Perhaps one way to do cross-disciplinary research would be to have a project where machines assess academic language for the similarities and differences among disciplines. Perhaps this has already been done, but I wonder what do philosophers mean when they say feminist methodology as oppose to what historians mean and what similar or different language each discipline uses to discuss these topics. The machine could scan all the journal articles in a series of disciplines and try to learn key words and do analysis. This would be more advanced than just textual analysis because ideally the machine could also analyze images/graphs/charts/and so on, as well as ideally teach itself the new language and measure that against another discipline. Granted, this is an idea from a person who indicated they didn't know very much about machine learning!

---

Machine learning can certainly enhance and advance our learning of cultures, languages and even human behaviour. Deep learning data analysts have shown promising results. So certainly machine learning can help cross disciplinary study and research.

---

A machine to crawl university-level, local, regional, and (inter)national archives that could facilitate research and aggregate personalized research inquiries and endeavors. Unlike resources such as ArchiveGrid or Europeana (though the expectation would be that the machine would tap into these resources and OCLC functions), the purpose of this learning machine would be to assist individual researchers by training themselves on the researchers' actual needs based on their queries, their previous scholarship, their location, and digital/material records.

---

---

I'm a literary scholar with a background in economics interested in using machine learning techniques for textual analysis. I know that people are working on this (across disciplines actually: we have sociologists, anthropologists, linguists and economists, as far as I am aware) but I would like to (a) know more and (b) see in what ways different disciplines could work together on analyzing text.

---

I have primarily worked with scholars to use topic modelling to examine large corpora of texts. In one case, I worked with a faculty member in the business school to examine what topics emerge in a corpus of journal articles about innovation and creativity. This faculty member was interested in drawing from other disciplines to structure his study, so we looked at literature on topic modeling from history and English.

---

### Social Media Sentiment Analysis using Twitter Dataset

---

A digital humanities project in a museum that conveys deeper meaning for a work of art that you may otherwise overlook or miss, if not informed earlier.

---

We have worked with the application of AI and ML to dynamically organize and tag medical school curriculum allowing the broad content areas to be organized by concept without reliance on medical jargon. This makes the curriculum more accessible, brings peer reviewed resources from the library into the conversation alongside faculty authored content and allows alignment and reporting to multiple standards. The AI and ML effectively acts like a virtual librarian who has no memory capacity but is expert in everything.

---

We use NLP and deep learning for applications as wide as health and psychology (sentiment, emotion) and linguistics (dialects, machine translation).

---

Machine learning applied to cultural studies is per se a cross-disciplinary activity, as it needs, to a certain degree, a background in computer sciences, either acquired through additional courses and seminars or by practice. Machine learning is needed when we want to account for both economical and cultural dynamics, for example when we need to examine two overlapping systems such as the television industry production and grassroots cultural manifestations. In a review of literature, might be helpful to adopt topic modeling to identify topics that intersect in research from different fields. Neural network analysis might also be useful to identify the connections between academic articles to enhance a cross-disciplinary research and go outside each discipline's boundaries.

---



---

Pattern/entity recognition in large digital image corpora to 1) filter image collections, 2) amass reference images, 3) enhance research capabilities on that entity (be it a building, a person, an artifact, etc.) 3) test our ability to retroactively apply photogrammetric techniques to create 3D models

---

Use machine learning to handle open source data.

---

I am exploring (and designing and using) facial recognition software to explore the history of governmentality, biopower, and the changing cultural understandings of the face in Western history. I plan on creating a media archeology of the technology and to build and use the software to make videoart pieces that make arguments about the technology and its histories.

---

Machine learning could be applied to archaeological data that would also be relevant to the study of climate

---

I conduct research on Wikipedia, and we are seeing that features in editor self-presentations on their user pages is associated with greater or less conflict in deliberation. Machine Learning would assist us greatly in expanding our dataset. I have also used LDA analysis (am I right in considering this to be a type of machine learning? topic analysis?) to explore how to identify influence, or rather affiliation, in poetry. This has not gone very well.

---

It could reduce the number of hours spent analyzing interview transcripts.

---

Process the research papers of many disciplines, identify language that asks questions or indicates unresolved issues, do the same with documented findings, characterize both in non-jargon synonyms, and look for apparent meaning overlap between fields using related concepts or words, then notify designated experts in each field where questions were documented that have overlapping synonyms with other fields where findings have been documented.

---

I am a [organization name] Humanities PhD currently teaching in a Computer Science Department in which we did a cross-disciplinary exploration of computational thinking in the disciplines with Biology, Music, Sociology Kinesiology faculty. While the topic wasn't machine learning, it was about affordances and applications of concepts to research methods in other fields. I can imagine it being applied to data collected in clinical or field work (crowd sourced or otherwise); lab data collected; musical patterns to do comparative analyses of genres and sound simulations; in literature, drawing connections among authors, styles, thinkers.

---

---

I would be especially interested in ways that machine learning could be used to support transcribing of old hand-written documents. A second potential case would be for supporting the study of old photos--grouping by similar scenes or similar groups of (potentially unidentified) persons.

---

Using digitised periodicals (the one I used had a 30-year run, with monthly issues) to train a word embedding model in order to explore how discourses of space and place were being used in this periodical. This project brings together discourse analysis with periodical studies and geography (in addition to the machine learning component).

---

I work as the Digital Learning and Scholarship Librarian [at my organization]. My main job is to facilitate the use of technology for interdisciplinary work and to partner with faculty to perform this research as a technology expert / humanities specialist. I'm increasingly working to incorporate ML into projects. Recently, for example, I have used unsupervised learning on a project with the Department of English to study a corpus of around 3000 essays to see how undergraduate students reflect on their own writing ability. The insights gleaned from this project will be used to refine curriculum and essay prompts in the future. Additionally, I'm very interested in the use of ML for predictive analytics across fields as well as its overall applications within the field of cultural humanities and linguistic (in which I have a long academic background).

---

## Sentiment analysis

---

I am interested in investigating machine learning further for its applicability in processing born-digital archival collections of heterogeneous file types.

---

I've explored machine learning a very little bit with other librarians who are much more knowledgeable of it than I am. I'd be interested in projects related to text mining that either contribute to conceptual history/historical semantics, or to topic modelling as a way to identify research trends.

---

At [my organization] we are applying machine learning to historical documents and seeking to allow researchers to examine them using current vocabulary that does not exist in the historical time they are studying. This will open the information in the documents to the researcher before they learn a new or unfamiliar vocabulary. This can also be applied to cross disciplinary research as well, but we are more interested in historical documents which are a form of cross disciplinary research.

---

---

As the Technical Director of the [an archive] amongst other manuscript editorial projects, I am interested in applying machine learning to explore handwriting recognition in manuscripts that are far too irregular or damaged to be interpreted by existing tools, and far too labor-intensive to be handled by human transcribers without machine assistance. The [archive] has formed a collaboration between scholars in English, Computer Science and Engineering, and [a center at another organization], with the aim of creating a complete editorial workflow solution for transcription and annotation of manuscript collections that could work within a browser.

---

I saw a presentation at Social Science History Association about spatial humanities that used text mining in historical newspaper articles about poverty to locate place names and then mapped those places, to see if the poverty discussed was more urban or rural. A parallel project might identify articles about crime, then locate place names within those articles, as well as words indicating types of crimes, and then map those. Cross-disciplinary and cross-archival research might compare this to police records to see how the media coverage of crime in a neighborhood compared to its relative level.

---

Connecting the scholars doing the similar or related research using the library's user record (from the pattern of checked out items/accessed articles).

---

If you have a problem that you are working on and if you can identify corpora of texts that analyzes this problem, you can tag this corpora for example, noun and verb senses, particular types of information, sentiment and then you can train the model using this tagged corpora and eventually apply it to a previously unseen document to find new instances of what you are searching for.

---

To automate tasks

---

I think machine learning would help researchers at my university find other researchers find funding bodies relevant to their work. Machine learning could use the researcher's current publications and grants to make recommendations for alternative funding sources as well as potential partners for interdisciplinary research. The Dimensions database does some of this work, but machine learning could make this more customized for researchers.

---

Connecting metadata between performing arts practices and archives so that people will be able to understand and find related interests. Or neural networks around food purchase, food consumption, and health issues.

---

---

NER is useful but does not represent the richness of the documents with which we work. For example, consider this fragment from a nineteenth-century French chronicle of Ottoman Algerian history: “To further attach himself [to his ally], Pasha Hassan married his [ally’s] daughter, then he launched troops against the rebel...” In just this short passage, which is not even a full sentence, we find several people referenced who are unnamed. If we look back at the text, we see that Pasha Hassan’s ally is Ben-El-Kadi of Kuku, Algeria, and the rebel is Abd-El-Aziz, but Ben-El-Kadi’s daughter is never named. This occurs frequently in historical source material. Therefore, I am currently working to develop an information extraction system that will pull named individuals, relationships, as well as unnamed persons and format this information in nodes and edges lists for network visualization and analysis.

---

Professional practice research for Social Workers to find policy and legislation pertaining to agency policy and procedures evidence-driven best practices at all levels--local, state, federal.

---

The Museum is looking at opportunities to use machine learning and other techniques to connect many social science practices with Holocaust history.

---

In our case we deploy machine learning to a broad class of interdisciplinary problems involved in recognizing features of sound, from bioscience to music and linguistics.

---

Training the machine to recognize movement from the body moving in space AND the notation systems being used. History, Dance, notation and graphical systems, architecture, all being supported by the machine. Actually, there are many dance and performing arts examples.

---

Use of machine learning-based tools for identification of people, places, topics. genres. form/structure, etc. in audiovisual content can assist in making collections more discoverable, usable, and useful across multiple disciplines. Development of such tools and systems with which to apply them supports cross-disciplinary research intersecting library science, computer science, computer vision, music information retrieval, computational linguistics, and other areas.

---

Building customized object classifiers for non-contemporary science and cultural heritage objects.

---

Information literacy and general education

---

---

I am an archaeologist, and I believe machine learning could be applied to a range of archaeological and/or anthropological problems to enhance cross-disciplinary research.

---

To aid in archival processing for a large corpus of born-digital materials - this could include named entity recognition and/or topic modeling, for instance, and could be applied to materials from a variety of disciplines. It could also aid in digital humanities research projects.

---

The main area that I've been working is to apply machine learning to collection management and collection building in libraries. I have focused mainly on digital library collections and have tried to use machine learning to assist with tasks that are commonly needed. I think that libraries have great potential in using machine learning tools and services to better serve our users.

---

Based on my work in digital curation, I am working on machine learning to increase the visibility of archival collections. At this time, I work with students whose background is in computer science and computer engineering. Using our current framework, we have started to add general metadata tags to grow our data and to give more context to the works. With that in mind, a potential use case for machine learning could boost cross-disciplinary research with respect to art history, political science, and other research areas where archives could play an important role.

---

Literature & Arts: A collaboration with History. We apply machine learning to extract person/location entities, relations, social networks, and many other types of information from Ancient Chinese literature. This brings a better understanding of the government mechanism more than 2000 years ago. Chemistry: We apply representation learning to describe atoms/chemical compounds with numerical features and feed into (deep) predictive models for materials discovery. Education: We apply machine learning to construct concept hierarchies from scientific literature so that the knowledge could be up-to-date in the field, far beyond the textbooks can go. Biomedical research: We apply machine learning to construct biomedical knowledge graph from more than 15 million articles for knowledge discovery, inference, and reasoning. And many many ...

---

The machine has been trained to keep track of a researcher's research activities including search behavior, depositing research papers and then analyze the pattern of needs, automatically generate results relevant to the next step of research and or relevant information (articles, domain expertise) from other interdisciplinary areas, after interaction from the research, the narrow-down information/directions even decisions will be produced

---

---

Machine learning could potentially be used to develop chat bots that assist with research discovery phase, directing researchers to common or uncommon research paths across disciplines, or alternately, in disciplines new to the researcher. Chatbots could be informed both through networked analysis of bibliographies as well as ongoing chat interaction with researchers engaging with other researchers, librarians, and chatbots in a curated online chat.

---

Our curriculum supports cross-disciplinary courses, students, learning. On our library website, it would be helpful if our "cross-reg" students could rank search results in the SERP as relevant or not to their query, so that the search engine could begin learning what is most relevant for our cross-disciplinary students (perhaps as opposed to straight-up business students).

---

I'm working with a student now to use interdisciplinary citation networks to predict or identify productive new avenues of research.

---

We have a collection of [city permits] on microfilm that is being digitized. The microfilm collection is heavily used by genealogists, architecture researchers, researchers of Chicago's built environment, environmental research firms, and interested citizens. We know that publicly available data set of this same information from 2006 to the present is heavily used by Computer Science students and students in our Urban Planning program. The digitized images of the 1872-1954 data will be of limited use: these are handwritten ledger entries and about 700K typed and handwritten index cards that lead the user to the correct ledger entry. Crowdsourcing is an option for transcribing and coding the ledgers so that they are an actionable data set, but the 700K index cards? I have seen presentations on machine learning for transcribing and coding handwritten material and the presenter used a set of images very similar to the cards we have.

---

enhance discovery and data linking under catalog or library discovery platform

---

Automatic indexing of documents; extraction of text from specific section in a document for analysis or enhancement

---

Given a particular problem where the search space is exponentially large, it might be possible to reduce the space by first applying a ML algorithm to it. For example, finding the local minimum energy states of a folded protein.

---

---

We conducted a very rudimentary test of neural network technology with one of our large collection of digitized archival images (over 32000 images from a single archival collection) - it seems this might be a way to facilitate research that incorporates archival research, information literacy, visual literacy, and possibly research into those resources that takes into account photography techniques, facial recognition, etc. I'm not entirely sure where to take it, but it was certainly a new way of looking at that collection and suggested new forms of research.

---

Machine learning for booksprints, where the machine learning is used both to deal with large amounts of data early in the process, but also to surface key themes, concepts and ideas as the process proceeds.

---

Machine learning models are useful for synthesis, prediction, and classification. Any research endeavor that involves one of these three canonical tasks may be of interest to machine learning practitioners.

---

Each area has its particular terminology, and it takes time for a researcher to learn it. I believe ML can help associating related areas that are "separated" by technical jargon.

---

---

First, my background: I'm a computer scientist whose research subspecialty is computer vision and robotics. Second, my soapbox: computing (that is creating computational artifacts, not only using them skillfully...) should be considered a professional literacy for more fields. Third, my insight from the past ten years of trying to make it so: the key to encouraging computing is very, very strong protections for the ACADEMIC IDENTITY of the fields in which it might be considered a useful tool. Though not as often acknowledged as such, computing is an identity threat to many. Computing's skillset will yield benefits (note: benefits, not breakthroughs!!) more widely if we (honestly) address that threat, ensure it's not a threat, and find the overlaps by which computing can contribute to the central challenges of other fields -- without foregrounding the computing per se. ML is the best computational subfield through which to encourage non-CS researchers to embrace of computing as a professional literacy and valuable skillset. First, its hype opens doors. Second, its many failures appeal to the (entirely justifiable) academic skepticism of what are really just shiny repackagings of data-driven function-approximation. But ML's true strength is neither its over-touted successes nor its (as yet) underappreciated limitations. Rather, ML is DISENGAGED from the (vastly) more important part of the fields it seeks to help (or to demonstrate its inability to help). It's nothing more nor less than model-building. The models may be useful (yay!) They may be distracting or worse (it's real life, after all). But the models are a means to an end. The end is, as it always has been, the field's motivating set of questions. The hard-won insight needed to claim insight from a model, however constructed, has not been diminished one iota! Indeed, models are now a dime a dozen (or a dime for several million dozen). Their ubiquity all the more starkly highlights the centrality of the humanistic and/or scientific sensibility, the philosophical priorities of practitioners and critics, and the aesthetics for creators and consumers like. These have always and will forever drive each individual field of inquiry.

---

I'm interested in how machine learning can improve our ability to navigate large amounts of text data specifically to help examine the contributions of under cited authors. My interest in this comes from the perspective of genre fiction with applications for LIS and HCI to examine what interfaces best allow users to understand and interact with the results of machine learning.

---

We have a project to digitize 700,000 objects in our photo archive, which would yield 1.4 million digital images. we are exploring computer vision for metadata generation, visual search in the resulting user interface, and as a tool for research/analysis. We are interested in identifying in particular the CV applications that might relate to areas of innovation or experimentation within the field of CV. (Knowing that some of CV applications that would be useful to us might be "easy" or well understood enough not to require collaboration with CV experts.) This project is just one example of one in which we want to explore CV.

---



---

I work in the humanities -- literature and cultural studies -- and I think machine learning can come into play in looking at social media analysis or text mining

---

We are currently working on a few different machine learning applications, but arguably the most interesting cross-disciplinary research project involves using SVM to classify sound effects in cartoons from the 1950s and 1960s. The project is a collaboration between Visual & Cultural Studies, Data Science, and the Digital Scholarship Lab. We are also working on an annotation tool for time-based media that integrates some machine learning for automatic shot detection, segmentation, etc. and can also be used to produce labeled training data. A third project, Architectural Biometrics, involves the classification and comparison of laser scan data to model changes between different architectural structures built from the same blueprints and is a collaboration between Art & Art History, Electrical & Computer Engineering, and the Digital Scholarship Lab.

---

I'm a metadata librarian and I am interested in computer vision to apply keyword/terms to images that could then be reconciled with controlled vocabularies. I'm aware that University of Utah Libraries is working on this, and I'm aware of tools like Google Cloud Vision.

---

I was involved in a cross-disciplinary project involving the classification of sound using machine/deep learning. I wanted to see how these techniques could be used to process large audio archives in libraries. I wished I could have participated more but lack the comp sci knowledge. You can find more information about the project here: [at a URL]

---

Creating and standardizing metadata and records creation for historical archive digitization and collections management.

---

Discovery of the authorship of anonymous or disputed texts

---

We have used machine learning to train custom entities for automated authority record creation with Spanish-language names. This work led to a paper co-authored by [a couple of people].

---

Applying natural language processing techniques for (1) digital humanities; (2) automated analysis of historical, sociological, or economic documents; or (3) domain-specific information extraction.

---

---

While there are several areas where ML has already carved a space for cross-disciplinary research (economics, law, biology, etc), from a library administrator viewpoint, I think library's collections and data could be a big part of these efforts. For example, following are a couple of projects where I see good collaboration across multiple disciplines: 1. Using ML to process huge backlogs of archival/special collection materials. Facing recognition to identify people in donated photos; improving OCR of digitized material; extracting names of people, places, events, historical dates, etc from collections; - partnership between computer scientists, social scientists, and librarians will be very valuable in solving these issues. 2. Longitudinal studies in biology and other fields will benefit from wide partnerships. At [my institution], librarians are currently working with a biologist who has over 40 years of [animal] population data that includes images to identify individual [animals]. Bringing along computer scientists in this project will help in developing automatic whale identification systems, based on current images. Librarians role in this project is to organize, digitize and manage the variety of data in this project effectively (images, data in MS-Access database, field notes, slides, etc).

---

Apply machine learning to remotely-sensed images and 3D data to locate archaeological sites. It would be interesting to apply the image recognition abilities of machine learning to classified remote sensing imagery to identify common patterns of specific site types in certain environmental conditions. It would also be extremely useful to apply machine learning to airborne lidar data to develop better post-processing algorithms for identifying archaeological features in diverse topological and environmental conditions.

---

My research team's current work is using machine learning to identify content and genre-types in mixed collections of digitized cultural heritage materials. The outputs of the classification itself--and the ability to connect researchers from many disciplines with materials for study--is one area where the project can ultimately facilitate cross-disciplinary research. The research team's own efforts are also cross-disciplinary, as I'm working as a researcher in the humanities as well as information science with computer scientists.

---

When working on large corpora in Cultural Studies, unsupervised machine learning such as topic modeling is essential to catalog recurring topics in a collection of texts and to ultimately create clusters on which scholars can work efficiently, minimizing the risk of biases. Additionally, neural network can help researchers defining networked structures and dynamics within a set of data or information (i.e. how institutional networks are built, or how social networks are generated following recurring schemes)

---

A dancer wishes to use neural networks to generate new dance choreographies given dances she herself has choreographed...

---

---

The easiest place to start would be Literary Theory and Anthropology because historically they have been allied and there are shared vocabularies such as “catechetical” “mimesis” as well as unshared vocabularies “virilocal”

---

How about two? Inferring genre metadata in the HathiTrust collections (or at first, perhaps, in the HathiTrust Research Center collection), and inferring discipline in a new project to create a national platform for sharing learning materials that have been made accessible for people with print disabilities. We have three participating repositories (HathiTrust, Internet Archive, and Bookshare), one of which (Bookshare) has human-generated metadata about discipline, while the others do not.

---

Using machine learning to determine prosody of 18th-century poetry. (English literature, linguistics, computer science.)

---

semantic web

---

I see a potential use of the ML in Peace Accords Matrix research in collecting data from web crawl and process the data to predict peace implementation trajectory.

---

The creation of OCR files from handwritten texts would have the potential to help researchers in multiple disciplines. Also, machine learning could be used to annotate datasets making them more useful across disciplines.

---

Intelligent application of taxonomy for auto-classification etc. could improve retrieval of research resources, both traditional (papers, data) and non-traditional (current projects, working papers, notes, experts).

---

Our [grant supported] platform allows faculty from eight colleges at the [my institution] to analyze text and image datasets (PubMed, HathiTrust, LexisNexis, etc) using a combination of ML methods.

---

I don't know if this applies but I would love to use machine learning to classify our search terms and use this information to better inform metadata creation across disciplines.

---

Employing machine learning to pattern recognition on images of handwritten documents to allow for virtual OCR across many disciplines.

---

---

Based on readers' checkout pattern, recommend what other materials they might want to use.

---

Machine learning potentially could analyze a corpus of articles from Field A and another corpus of articles from Field B, both corpora having been curated because they deal with some similar subject matter (e.g., family, politics). The machine learning could recognize that Field A talks about politics using these terms, but Field B talks about politics using these other terms. The machine learning could then build a crosswalk between the two fields' terminologies, which would help library discovery tools pull up relevant results based on the searcher's stated discipline.

---

Neural networks could be used for the analysis of digitized photographs from library special collections in order to classify "pictorial tropes" that may be of interest to scholars in art, history, and literary or cultural studies. This use case is borrowed from the Yale Digital Humanities Lab project Neural Neighbors ([http://dhlab.yale.edu/projects/neural\\_neighbors.html](http://dhlab.yale.edu/projects/neural_neighbors.html)).

---

authorship attribution and data mining in ancient sources.

---

One possible way could be topic classifications from raw texts (such as articles) with unsupervised learning to group them in a cluster. This could enable researchers to discover resources related to its topic, not based on a disciplinary context. In addition, the New York University tries to use learning machine techniques to facilitate data search and uncover relationships between datasets and articles. Its goal is to enable researchers to identify who work with the same datasets, on what topics and what final outcomes. Lastly, it could be interesting to compare results between unsupervised learning and supervised learning in which metadata is created by professionals.

---

One possible way could be topic classifications from raw texts (such as articles) with unsupervised learning to group them in a cluster. This could enable researchers to discover resources related to its topic, not based on a disciplinary context. In addition, the New York University has tried to use learning machine techniques to facilitate data search and uncover relationships between datasets and articles. I've analyzed datasets from different sources and tried to connect them by pattern. I haven't applied yet machine learning to try this as I'm learning ML via Coursera, but I do believe that there could be potential use of machine learning to link different datasets by patterns which could enable researchers to discover that they are not aware.

---

---

Political Science, [particular school and particular institute]: We can use news stories about elected heads of state to predict the trajectory of democracies in the following years. Law and Political Science: We can use the text of reparation measures to victims ordered by the Inter-American Court of Human Rights to predict the probability of compliance with the Court's rulings by states that have violated human rights.

---

(1) the ability to learn to recognize patterns from training data without knowing all the disciplinary specifics that lead to classification of that training data is very powerful. (2) sometimes pattern recognition can produce results that entrenched disciplinary analytic and simulation frameworks cannot. We've already seen some interesting examples of this, for example in materials science and physics.

---

Machine learning is applied statistics and problem solving. I honestly cannot see a field in which Statistics and Problem Solving are not relevant.

---

We often pretend that "data analytics" or "data science" is something new ... however, science without data isn't science. What changed is that we use computers to assist us in collecting, archiving, and analyzing data. As such ML as a subset of data analytics is an integral part of any science, and often these tools are what provides new insights or the ground to collaborations: "I can do something you can't in a domain I know nothing about, please let's work together?" -> synergy

---

Personally, I am really interested in automating the cataloging process with machine learning for libraries. I think there is a great potential here if libraries have access to full-text content and cover images, and a machine learning frame can be developed by librarians working along side with computer scientist and domain experts to extract cataloging information for both bib records and item records. Another more ambitious use case is to develop a better knowledge base for the whole internet. Online information and knowledge is quite scattered and unorganized. As a result, we rely on search engines like Google to discover specific information. But if we form a group of librarians, domain experts and computer scientist, it is possible to start a better data structure of machine verified online information. On top of that better knowledge base, people can also run natural language queries to get exactly the information they are looking for.

---

reducing noise in collections by producing a classifier to assess relevance

---

---

Machine learning could be applied to improving (or making possible) OCR of hand-written documents from historical periods (medieval, early modern, modern). Having OCR'ed copies of historical handwritten documents would be of great use to historians and literary scholars. Another use might be automated resolution of subject headings across multiple metadata schema. So for example, automatically associating the subject "Cranial Neoplasms" from the MESH headings with "Skull--Diseases" from the LCSH headings.

---

I'm interested in machine learning to train computers to better read and ocr manuscripts and handwriting, which I think could be a cross-disciplinary project linking computer science and humanities.

---

author name disambiguation, citation classification

---

In the past, I was a software engineer working in research on speech technologies (speaker recognition, automatic speech recognition, language detection, etc.). Libraries could help machine learning researchers by providing data (maybe even labeled data) in exchange for transcription metadata to improve search and discovery. This could also apply to video and image processing, but speech is how I think of things. In training speech systems, having the appropriate data across different domains (e.g., medical vs. military, US vs EN, etc.) was essential for training language and acoustic models. Having the ability to gain access to data and be able to use it to either train or test systems was often difficult and costly. If libraries provided audio and video data along with metadata to researchers in speech technologies, this could be useful for testing and training systems, ultimately resulting in improved systems across different domains. Then, if machine learning systems could provide back to the libraries any resulting transcriptions from audio or videos, then it could be useful to the library as additional metadata for domain knowledge seekers to search and discover information better.

---

Identify similar topic, better categorizations

---

I am collaborating with two research groups in Computer Science department and applying machine learning to reveal the hidden patterns in student learning behavior, discover the strongest predictor for successful learning outcomes, and identify and boost students who are not thriving academically.

---

---

This is [name], and I contributed an example in the project proposal. I will use the same example. I collect data on free trade agreements (PDFs of the treaties). I measure the impact of FTAs on trade flows, and then ultimately on economic welfare. Historically, I have used a dummy variable in linear regressions to estimate the impact of an FTA on country-pairs' trade flows, and then use the trade flow changes to help estimate the welfare effects (via a theoretical model). However, this method is crude. Two issues come to mind. First, to the extent we could use machine learning to better interpret the degree of economic liberalization in the FTA treaty and distinguish it from other FTA treaties, we could more finely measure the liberalization and use this instead of a dummy variable in a linear regression. Second, it would be useful if we could -- in an interdisciplinary way -- see how economic and political factors interact to see if there are additional or separate benefits from an FTA attributable to political or governance benefits relative to economic benefits.

---

I'd love to examine the potential of machine learning to interpret natural language search queries when presented with structured metadata for library records; to look at searches based on sample records or full-text documents to suggest similar results; and I'd like to look into topic modeling or sentiment analysis of early modern titles as a way of grouping / ranking search results

---

1. Automated tools that can suggest possibilities of research areas with relation to need or funding available. 2. Tools that provide network diagram to collaborate across disciplines or subject areas.

---

Using a trained text mining algorithm to identify the mathematical theorems and lemmas which are applied in scientific disciplines to better understand the role mathematics plays as the foundation of science.

---

We're using machine learning to extract from the corpus of open access research publications in our repository to do concept extraction and faculty mapping in order to help facilitate forming inter-disciplinary teams take on new research challenges.

---

The Library of Congress has billions of files in its web archives. The files have a range of technical characteristics as well as technical metadata related to where they were archived from the web. This seems like a domain where machine learning could be useful in enabling access and discovery of this content to support research in a wide range of domains.

---

---

We would like to explore machine learning in automated description of image and textual objects within special and digital collections and manuscripts. There are tens of thousands of objects in our collections that are currently inaccessible to scholars, researchers, and librarians.

---

A kind of meta-example I have in mind (and partially underway) is to run a clustering algorithm on the vectorized text of a given institution's collection of theses and dissertations (ETDs). The development of the clustering procedure in itself might involve computer scientists and (e.g.) librarians, but also potentially any other faculty conducting meta-disciplinary analysis, perhaps especially historians of science and their ilk. The resulting clusters would also serve as a dataset that was \*about\* cross-disciplinary research, by revealing which discipline's ETDs were most closely related, to then be compared to other more familiar measures of similarity (librarian subject headings, administrative structure of the institution, etc).

---

1) Recommendation systems could be tailored to encourage cross-disciplinary suggestions. 2) OCR and proper digitization of books so that they are searchable and machine-readable would allow for the knowledge to be catalogued and accessed for further use by any computational discipline

---

The most exciting use case to me is the potential to use machine learning to allow libraries to exchange metadata, and especially in adapting metadata for use as linked data.

---

I am interested in the applications of machine learning and natural language processing in advancing scholarly research in a library setting.

---

Machine learning could allow for description at a larger scale than can be resourced with people.

---

Analyzing historical school district data to reflect demographic transit and local development

---

I manage [my institution's] discovery interface. Results (particularly from our licensed content provider, but also from our catalog) are returned through keyword matching, with the ability to narrow among pre-coordinated (and often non-controlled) subject headings. Being able to expose items from across subject domains based on their "aboutness", through machine learning techniques, would be an incredibly powerful tool for exposing otherwise hidden gems of our collections.

---



---

We have a large digital postcard collection (approximately 3000 images). Faculty collaborators have thrown around some ideas for creating a model that would allow for facial recognition and the identification of other visual patterns across the collection. I find this to be a interesting concept. I'm compelled by possibilities for analyzing digital archival cultural heritage collections as data.

---

Library with the expertise to bridge cross-disciplinary research and teaching. One example could be segmenting videos from a library collection and delivering an algorithm to find themes that might be relevant to several disciplines or research groups.

---

Machine learning training via Library Carpentry to better prepare the library community to have conversations with researchers and support the research enterprise further.

---

It's useful in pan-humanities research, for sure. Manpower for quantitative analysis of source material is scarce: almost entirely lacking. So, this analysis will certainly rely on machine learning in the future, even for simple OCR. Personally, I tend to include techniques like crowdsourcing, which although relying ultimately on humans for analysis, nicely complements machine learning, and can provide similar interfaces/results for researchers. Once Digital Humanities corpora are analyzed, they provide rich datasets for other researchers within the humanities, but also for science and engineering fields. For a practical example, Google has leaned heavily on its Google Books analysis for its excellent translation tools.

---

Machine learning of controlled vocabulary; classifications, etc

---

Use machine learning in combination with natural language processing and Knowledge representation(graphs) to ingest text and analyze text without building ontologies or taxonomies.

---

Using a deep neural network to train word embeddings on a corpus of founding era legal documents that can be used to further legal and corpus linguistics research.

---

---

The [organization] recently developed a Digital Humanities Fellowship and our first Fellow used machine learning to explore methods of locating images with certain visual characteristics. He used OpenCV and some machine learning techniques to identify approximately 40,000 images out of about 40 million images in a particular collection. This was not my own research but I played a supporting role in the project. I believe there is tremendous potential in using machine learning to support research in our historical collection, including oral history audio and video and archival paper collections. The general method described above has garnered a lot of interest. Our historical collection comprises some 85 million page images. If machine learning could be used to identify certain types of materials, it could support research by providing efficient research access to that type of materials. For example, if images relating to correspondence could be identified, using machine learning, out of large collections of which correspondence comprises but a very small percentage of images, this may be of great interest to certain researchers. Also in our collection there are sometimes scattered lists of persons, which may comprise again a very small percentage of the entire collection. I think this scratches the surface. This type of inquiry is by nature completely interdisciplinary, and various use cases could be useful to catalogers and archivists and/or to historians and/or genealogists and computer scientists.

---

Hypothetically , lets say a law professor would like to study the outcomes of particular set of data privacy laws in the EU. This research could cut across multiple domains, such as economics, social sciences, health sciences, public policy. In order to carry out such research the same professor may need to elicit help for Machine Learning aspects from computer scientists and or data scientists. Further , if a predictive module of future outcomes were the goal, then even more emphasis would be placed on the computational aspects of this research.

---

I have participated in grants where the PI is proposing use of machine learning to simultaneously search PubMed, PMC, and multiple other databases to extract all mentions of a single specific gene. I'm not the primary researcher - I just provided the RDM plan - but this is increasingly a common type of work at my institution.

---

Using machine learning to dynamically create evolving search terms. For example, starting with search terms that are broad and seemingly unrelated, then as results are selected as of interest, incorporating into the search to create possible new search terms. Also opportunity to introduce accidental discovery by having system choose a near-related topic to include.

---

---

ML could be used on the full text of articles to extract entities--people, places, organizations, chemical compounds, "subjects", et al.--in a much more scalable and rigorous way than human generated metadata. These entities could then be linked across different domains that typically have different descriptive practices and indexing services. We've been exploring AI/ML as a potentially valuable complement to linked data in the Mellon-funded LD4P2 grant--with the former giving good entities, and the latter serving as a way to stitch resources together across institutions and domains. Further, ML on articles could be used to provide or augment description of datasets, which might have great cross-disciplinary reusability, but typically come with little to no description of their own.

---

Categorizing historical photos of the region with different sets of categories that apply to different domains. Forestry researchers may be interested in the types of vegetation, artists may be interested in the colors and shapes and geography researchers may be interested in the difference between historical and modern images of the same location.

---

Analysis of historical special collection pieces in our Library.

---

I think the argument in your proposal is compelling. Also: in emerging areas of science, where the terminology is still fluid or is tied to a particular school of thought or lab... seems ML could help in retrieval by adjacent fields? Like when nanotech was new - literature from chem, phys, engineering mostly separate.

---

I'm interested in exploring machine learning to facilitate an understanding of how K-12 teachers can find and apply educational standards to their supplemental classroom materials.

---

Digitized newspapers and local history collections are my #1 right now for ML. I'm from a public library so learning how to make these tools more broadly available to researchers by making our collections more accessible to the bots is where I'm going with this.

---

Based on the search parameters given and a history of those parameters over a session the machine learning algorithm could suggest more relevant hits based on this.

---

Ability to speed up metadata creation, detect outliers in data sets, and other tools that can enhance the data and it's use.

---

---

From the NEP: New Economics Papers and the bims: Biomed news project, it would be possible to build reports that are between the area of economics and medicine. In fact, we have a nep-hea report, but it does not at this time take in papers from PubMed. In general machine learning has great use in bringing in documents that relevant but outside the discipline of the user.

**Everybody - Why do you think machine learning won't enhance cross-disciplinary research?**

Why do you think machine learning won't enhance cross-disciplinary research?

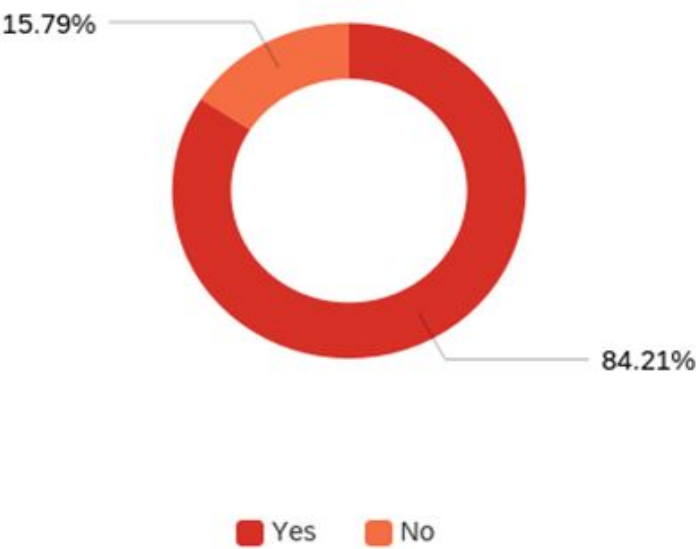
---

I can see a point in opening, amassing, and categorizing research, and make it more accessible to the layman and to people outside the field. This would in itself make research more widely useful and applicable, but also more searchable. However I don't see why machine learning specifically needs to be at the center of this. In my view, standard keyword-based search would do wonder over such content, the roadblocks lie in acquiring in.

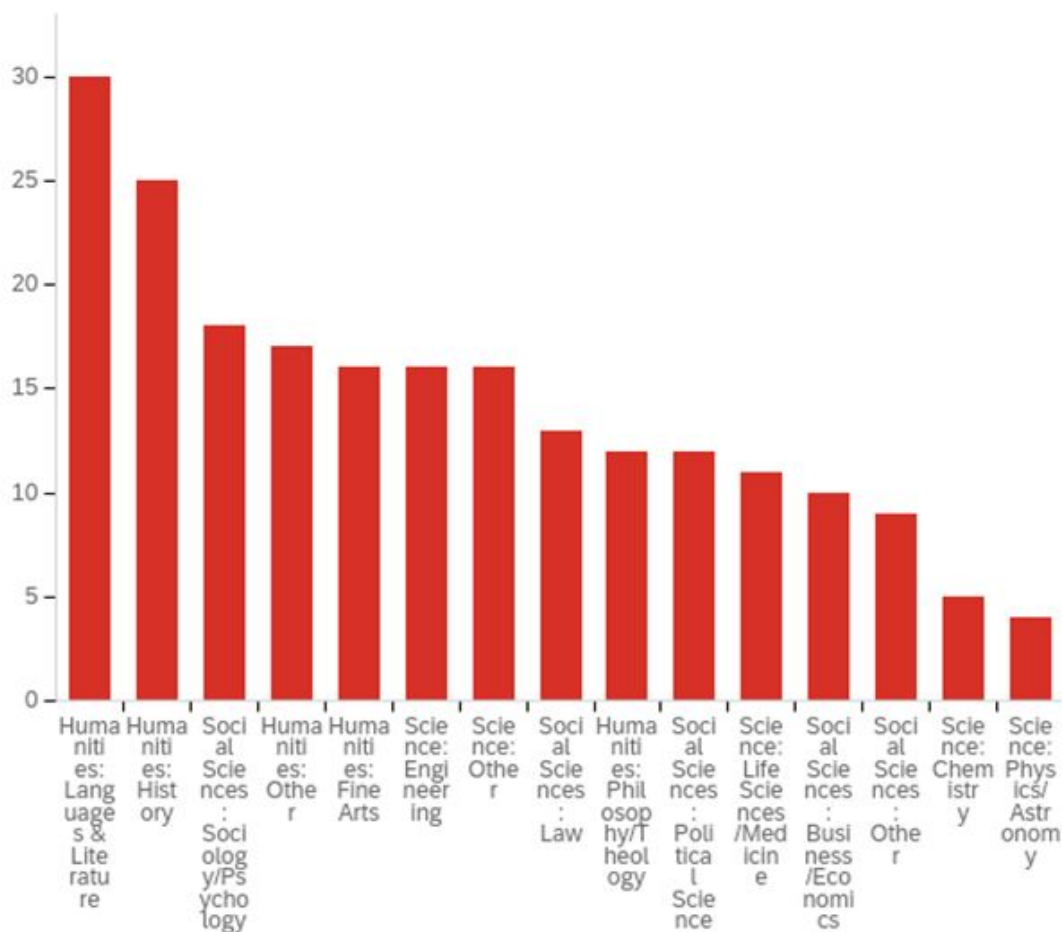
---

Its a tool, agnostic to employed disciplinary methods.

Scholars - Do you collaborate with scholars from other disciplines in your research?



**Scholars - For what disciplines are you producing cross-disciplinary research?**



**Scholars - In two or three sentences, please describe your cross-disciplinary research.**

In two or three sentences, please describe your cross-disciplinary research.

---

biomedical informatics; communication/argumentation

---

I have collaborated with others in the fields of computer science and fine arts both. Though digital humanities does not describe my work well, in this case, it's the closest bucket to put my work in besides also instructional design, curriculum design, communication design, pedagogy, poverty studies, cultural studies, etc.

---

Data visualization of other knowledge domains

---

---

Previous research included working with dance in generating ballroom dance choreography, using AI and game theory.

---

I apply ML techniques to unstructured policy text, which I call "policy analytics." The dimensions of text that I quantify depend on the research application in mind. Sometimes, it's econ; sometimes, poli sci or law.

---

Almost all my research is crossdisciplinary or interdisciplinary. One current project is about understanding book trade networks in the enlightenment - this involves historians, linguists, computer scientists and literature specialists. Another involves understanding pre-historic contact networks in Australia and the Pacific, and includes linguists, anthropologists, archeologists, literature people, design specialists and computer scientists. Another is a collaboration between historians, linguists and anthropologists to digitise and repatriate important 19th century anthropological manuscripts.

---

I am working in transmediation of mo'olelo (history, story) from oral and textual forms into digital formats. I have worked with Aboriginal Territories in Cyberspace (ABTECH) and the Institute for Indigenous Futures as a cultural consultant helping students make Hawaiian language video games which incorporate traditional knowledge and stories into the gaming experience. I am interested in how engagement with the digital can maintain indigenous knowledge systems---through language or translation. I am called to consult on many different projects for the community, in public history and programs for museums, for scientists studying climate change, etc.,

---

AI and inclusion

---

I have a PhD in history, am employed as an Assistant Professor in Communication Studies, and contribute to research projects on women in science.

---

I am trying to understand heritage languages and culture and the changes they are undergoing in light of globalization

---

My scholarship is in digital humanities, games studies, and teaching and learning. My DH research engages with the sciences (linguistics and engineering), and teaching/learning scholarship engages creative making curriculum design. I collaborate primarily with other DH scholars, archivists, and librarians.

---

I'm a computer scientist developing workflows, tools and methods for scholars in the humanities and social sciences to do their research.

---

---

I work with Cultural experts from Native American groups, early American historians and historians of the American West, art historians, computer scientists, and archaeologists.

---

My own research uses feminist and critical race approaches to the epistemology of technoscience to analyze social media. Part of this research interest is using social media to study transdisciplinarity. I also collaborate with colleagues from various fields on educational research projects, including research on teaching digital literacy, and researching faculty development around interdisciplinarity.

---

I study conflict in deliberations in Wikipedia.

---

I analyse city dashboards and real time data users for the development of user centered design paradigms.

---

Interested in issues tied to pedagogical approaches and how tools for thinking model aspects of computational methods for analysis and thinking through a dynamic or issue.

---

I have developed [app name], a tone training app, in collaboration with game developers, programmers, sound engineers, and graphic artists. We modified [this app] to use it as an experimental instrument for an empirical second language studies and used the audio assets to build an interactive audio database in collaboration with digital humanities librarians,

---

I use word embedding models to better understand the workings of large textual corpora for the purposes of literary analysis (this is work that is cross-disciplinary but not undertaken with any collaborators). I have also collaborated with a database expert to use topic models to explore a corpus of book reviews. In my non-machine-learning work, I have used stylometric analysis to explore the authorship of late C19th texts, and have a collaboration in process with a statistician.

---

I am the Visiting Assistant Professor in English and Digital Scholarship [at my visiting university], where I serve as the Technical Director of [an archive collection], contribute to the development of digital humanities curriculum, and teach classes on modernist literature, media studies, and the digital humanities.

---

I'm a historian currently working with a sociologist colleague on a project about historic preservation and segregation. We are taking a model developed to study these trends in

---



---

New York and applying it to Boston and other cities, seeking the ways in which historic preservation has shaped segregation by race and class.

---

historical development of credit markets and institutions, effect of technology on conceptualization of humanities, public humanities and data visualization

---

While my work focuses on historical source materials, I collaborate with librarians, linguists, scholars of literature and cultural analytics, mathematicians, and programmers.

---

As above

---

WE use motion capture to study intersections of language and movement. Statistical analysis of empirical data, abstract movement 2d and c3d data, video/audio generation and analysis. Biomechanical engineering is a large part (see ARTeFACT and Schrifftanz Zwei projects)

---

History of anthropology. I am trained as an anthropologist but it seems to be about split on people who were trained in history or anthro. I also do a lot of work in museum studies, which draws folks from anthro, history, art, sometimes education/psychology.

---

Literature & Arts: A collaboration with History. We apply machine learning to extract person/location entities, relations, social networks, and many other types of information from Ancient Chinese literature. This brings a better understanding of the government mechanism more than 2000 years ago. Chemistry: We apply representation learning to describe atoms/chemical compounds with numerical features and feed into (deep) predictive models for materials discovery. Education: We apply machine learning to construct concept hierarchies from scientific literature so that the knowledge could be up-to-date in the field, far beyond the textbooks can go. Biomedical research: We apply machine learning to construct biomedical knowledge graph from more than 15 million articles for knowledge discovery, inference, and reasoning. And many many ...

---

Working with others to think critically about the concepts and ideas that are foundational to, but not necessarily explicit in machine learning and related areas.

---

Understanding videos of humans conducting music ensembles. (project has been dormant a couple of years but I want to restart it).

---

---

Computing -- that is, creating, not consuming, computational artifacts -- is most assuredly not like writing. Yet computing, like writing, affords many disciplines (and their practitioners) the opportunity to build insights of value to their paths of inquiry. My research is developing those paths.

---

Historical archive / digital surrogate creation and dissemination

---

We use techniques from Information Theory to capture various aspects of literary style in order to engage with authorship attribution problems in early modern drama

---

I collaborate with computer scientists and engineers, specifically working in geoinformatics and 3D optical metrology, as well as with art historians and soil scientists on using and developing geospatial and 3D tools to understand processes of change in ancient Maya landscapes. I also work with colleagues in animation on data-driven virtual reality projects to investigate how VR might be useful in archaeological research (beyond public engagement).

---

Published in Electrical Engineering, Literary Theory, Mathematical & Historical Cartography, Statistics, History. International prizes in software engineering for game design and Cartography. National prizes for history.

---

In the past, I collaborated with the iCeNSA in Colombia project. The project was to monitor and analyze the polarization of people's perception of the peace agreement and how that impacted the referendum outcome.

---

Our Digital Scholarship Center is the lead on several cross-disciplinary projects using topic modeling, neural networks, and computer vision on datasets across disciplines.

---

I am PhD both in Physics (chaos theory) and Theology and I have applied some statistical methods to authorship attribution for texts of the 4th century

---

---

My recent cross-disciplinary work includes projects like: (1) creating and implementing technology-enriched formal and informal learning programs to empower female students in Grades 7-12, focusing on health-related STEM areas; (2) research on cancer survivors' needs and preferences for health information management and create a new technology for cancer survivors for health information management and for life style changes and management; (3) creating and implementing a comprehensive blended program for healthcare professionals, medical students, children with special needs, families and community leaders on leadership in developmental disability; (4) equity and healthcare disparity; (5) mobile training for bare-footed doctors in remote areas on HIV prevention and patient communications; (6) interactive e-learning systems on computational thinking and more.

---

I work with a sociologist to analyze appointment of women justices in Latin American supreme courts. I am developing a project with the Inter-American Court of Human Rights to assess compliance with the Court's rulings.

---

I have projects with Sociologists, College of Global Affairs, Ecologists, The Vatican Secret Archives and others.

---

I study how brains evolved to evolve AI. I work with behavioral biologists cognitive scientists, psychologist, engineers, evolutionary biologist, computer scientists.

---

My work focuses on the analysis of historical library collections, which involves reconstructing lost libraries and analyzing their collections as a means to understanding their social impact on historic users.

---

collecting tweets and webpages to aid various researchers, and helping them with subsequent analysis and visualization

---

Large-scale literary history via text mining. Integrate social, economic, and literary data sets.

---

Studying how archives construct our understandings of the past

---

I work in Library Science, web and visual resources archiving, and data visualization. I'm currently working on a project analyzing change in web homepage form/structure over time.

---

---

data-centric science & technology studies (e.g., scientometrics)

---

[redacted for respondent privacy]

**Scholars - When doing cross-disciplinary research, what obstacles do you usually encounter?**

When doing cross-disciplinary research, what obstacles do you usually encounter?

---

Norms are different in different fields. Lots of proprietary software. Challenging to publish in journals.

---

Benchmarking models of accepted rigor and justifying their use in your work (since I don't qualify as a 'sociologist' but if their assessment metrics fit--I'm gonna use them--just have to justify it formally)

---

Lack of understanding/support in reviewers from each discipline (for journals or conferences)

---

Communication. The terms and way of thinking are very different, and sometimes, collaborators have trouble understanding how much time AI needs to work. The black box nature of the machine learning algorithm is also difficult to explain.

---

jargon; using different terms for the same thing

---

Lack of shared vocabulary, leading to misunderstandings. Differences in expectations about research processes and workflows. Differences in data management and preferred tools.

---

One of the obstacles I encounter is the lack of a shared lexicon, since language requires context and the contexts we are encouraged to build ourselves into are often not indigenous. I require funding and support for the students who work with me to build up some of this scaffolding to the point where it is shareable and accessible.

---

Institutional obstacles

---

---

Academic gate keeping. Also, having difficulty speaking the jargon across disciplines. When I say "big data" in a humanities setting it doesn't mean the same thing as in a social science setting.

---

Just started

---

Understanding keywords or the complex histories of disciplinary discourse; similarly, adapting keywords to appropriate contexts in rhetorical situations outside of those disciplines.

---

Successful cross-disciplinary research needs both sides to be genuinely interested in the viewpoint of the other, as well as a literacy of the interests and possibilities the other side has to offer.

---

Identifying shared research priorities, epistemologies, and methodologies.

---

Different scholarly communities mean that we have different audiences in mind for eventual publication, and so there is not always agreement on what is most exciting or worth further pursuit within our preliminary findings.

---

I often wrestle with my weak skill set.

---

Communicating in the same language.

---

Translating terms and finding ways to articulate the limits of what things like machine learning can do, which turn into interesting moments for how to think around and with the tech.

---

the need to translate a discipline specific notions and language. For example, the term "tone" means a wide variety of meaning to my interdisciplinary team members.

---

Different vocabularies and critical emphases (though this is very often a productive as well as a problematic difference). More prosaically, there are very different expectations around authorship, credit, timelines for publication and suchlike between humanities scholars and those in the STEM disciplines.

---

Funding shortages and poorly structured incentives.

---

---

Often, we will study the concepts, but with different lenses and methods. Studying racial segregation, for instance, crosses boundaries from History often with Sociology and Economics. We often have differing ideas about evidence--what constitutes evidence, what constitutes proof, etc. We also often have read different foundational texts. Our differences can be both micro-level and macro-level in this way, even when we are working on the same fundamental topic.

---

lack of support within specific field, lack of financial support to take on research that has uncertain outcome or is rooted in more than one field than the one the granting agency is interested in

---

Currently, I have no funding to support my research, so collaborations happen through colleagues' goodwill and willingness to share their time and expertise.

---

Some researchers deploy machine learning in search of tools and others for deeper understanding. This distinction is compounded with the different disciplinary approaches. Another issue is funding more available in stem and which tends to drive the research trajectory.

---

specialist language challenges, incorrect estimates as to time it takes to complete, lack of understanding regarding bodies (e.g. dancers), finding equipment (generally \$ and others not willing to share equipment (eg mocap systems). lack of understanding of complex movement.

---

Different theoretical orientations.

---

Domain knowledge (language) gap between collaborators

---

Conceptual confusions between disciplines.

---

Data collection - obtaining an adequate amount of data with adequate quality. The omnipresent "barriers-between-disciplines" obstacles also exist but are not distinctive problems with machine learning.

---

The primary obstacles are ones of academic identity, e.g., "these tools were not part of the way in which I absorbed my discipline -- as a result, using them may threaten the core principles and practices of the discipline."

---

---

Limits of my own machine learning and other technical limitations

---

Scholars in the Humanities are sceptical of computational methods and have trouble understanding them

---

Applying for Humanities grants that provide smaller funding amounts that are not sufficient to support collaborators in engineering and computer science. Funding challenges also exist with international partners; however, often it is necessary and important to collaborate outside the US. At times, there is a steep learning curve of the technical side to effectively lead coordinate/lead projects even if I am very familiar with applying the tools, sometimes it is challenging to acquire the knowledge and terminology to "translate" my ideas directly to others in different disciplines; however, I have become better equipped to deal with this challenges through experience.

---

Learning to speak another disciplines lingo.

---

Nope.

---

Mainly cultural and "dialect" differences between fields.

---

extension of bibliography and need for technical abilities

---

Obstacles are mostly due to the lack of communications with and understandings of other disciplines

---

Different disciplines emphasize different themes

---

Other disciplines usually think that we can do a thing without much better. ML seems magical to them; they don't understand the research involved.

---

The key is to find a common language and to identify each other's strength and weaknesses, also often the question about where we can find synergy is important.

---

Siloization of resources within traditional departments necessitates separate places within the university to work across disciplines.

---

---

takes a little time to learn the vocabulary, ways of thinking, and problems of interest

---

Access to data, knowledge of best practices across disciplines.

---

just getting up to speed in the other field's literature and their technical vocabulary

---

Mismatched jargon. Too much reliance on specialist technologists, or fear of engaging with technology.

---

people from different disciplines are hard to talk to each other (e.g., taste, what counts as a contribution, etc.)

---

The largest roadblock is funding.

### **Scholars - How do you currently address these obstacles?**

How do you currently address these obstacles?

---

Do due diligence in understanding the tools of many fields and using them responsibly, acknowledging why and how they were developed and why and how they apply to my work.

---

Specialized and interdisciplinary venues

---

Analogies.

---

I learn the terminology of both disciplines.

---

Lots of meetings, lots of talking, and willingness to experiment with different ways of doing things.

---



---

By working with partners from outside of [state] from [other institutions], who have a lot of resources to assist us and particularly institutions outside of the U.S.

---

Personal networks

---

In initial stage

---

Reading more, talking with colleagues in other departments, revising writing.

---

In a cross-disciplinary project, there needs to be constant tight interaction between the people from the different disciplines. Intermittent workshops are not enough, the work truly needs to be integrated.

---

It varies by project. In some cases it involves collaboratively defining a project from the outset. In other cases it involves remaining open to an organic research process that may lead in new, unanticipated directions. In all cases, it requires trust and rapport among the collaborators.

---

Try to be flexible and frame the research question in ways that could be of interest to many audiences.

---

I have taken some courses. I get some assistance from my research colleague who is in business. Sometimes, I don't overcome them.

---

Time. We spend a lot of time on comms.

---

Collaborative sharing of resources and trying out lesson plans that share common learning objectives from differing disciplinary vantage points.

---

Through multiple open conversations until we find a shared understanding

---

A lot of talking and negotiating. I do my best to learn the unspoken social norms of other fields, esp. machine learning and NLP, before blundering into them. (Eg. conference papers are \*always\* dense and convoluted and very much written to be read aloud at literature conferences, but to deliver a paper in this way at a computer science conference would be frowned upon.)

---

---

By building strong core teams with shared motivations to be able to weather the inevitable lack or cessation of resources and to provide incentives based on participation such as personal and professional development that our institutions may not proactively offer.

---

Collaboration is very time-consuming and cannot be done at the pace of a traditional solo research project. A great deal of patience and discussion are required in order to proceed with meaningful work.

---

i ignore them and persevere

---

I am early in this project, so I am still building my cross-disciplinary team. The next step will be to apply for grant funding.

---

Discussion. Not always possible to reconcile.

---

continue searching out willing collaborators, and demonstrating the value of the projects

---

Cite everything? Read outside my wheelhouse?

---

Frequent meetings

---

Formally through many meetings, but increasingly the sprint methods (e.g. booksprints) are very productive.

---

Be creative - collect data from multiple sources.

---

Ensuring that the computing occupies its rightful place as a toolset and no more. Today's "conventional wisdoms" actively work against keeping computing in proper perspective, alas. Yet, if any institutions are able to maintain perspective, our colleges and universities are they.

---

Working with a team, and of course developing my areas of expertise

---

We try to write extremely clearly and use analogies that make sense to Humanities scholars

---

---

We look for complementary grants in our respective fields to acquire sufficient funds. I find colleagues that are truly excited about cultural heritage and that excitement shines through in their engagement and devotion to projects. I also find a research aspect that satisfies their interests and not simply consider them as a partner to apply the requirements I need--we are both carrying out this research to address research in our individual fields, but we can do it much better by working together and pursue new lines of inquiry.

---

Humility.

---

Not Applicable

---

friends

---

reaching out, networking, and lots of communications

---

Sometimes it is just luck to find people who work on similar topics across disciplines

---

Downplay the AI hype, I need to thoroughly understand their problem before I can begin to help.

---

We talk, talk, talk, and at some point try to explain what the others do with our own words. Also, we often just make experiments or let ideas clash and sort out the results after the fact.

---

Seek out multidisciplinary spaces

---

meetings, working on papers together

---

Slowly! Conversations with collaborators, reading journals beyond my own field.

---

conversations with scholars, reading

---

Education, outreach, patience.

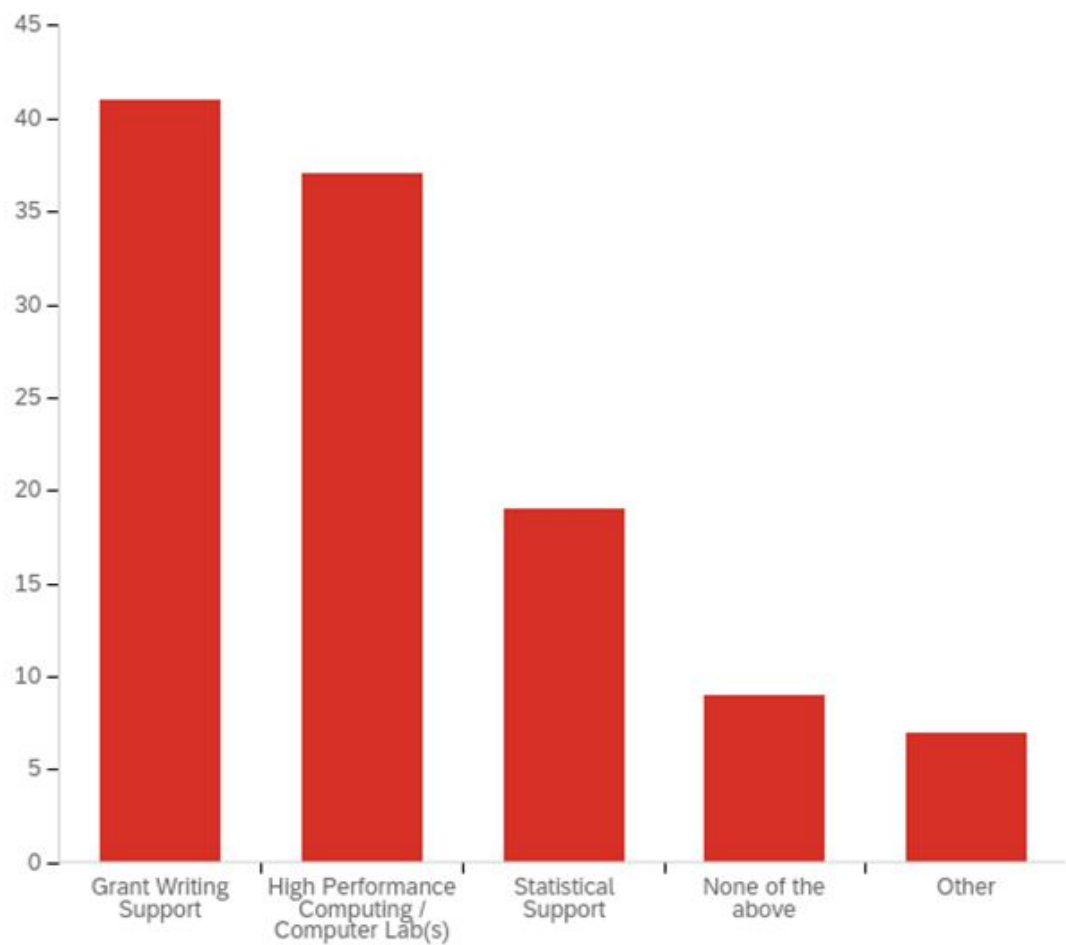
---

talk :D

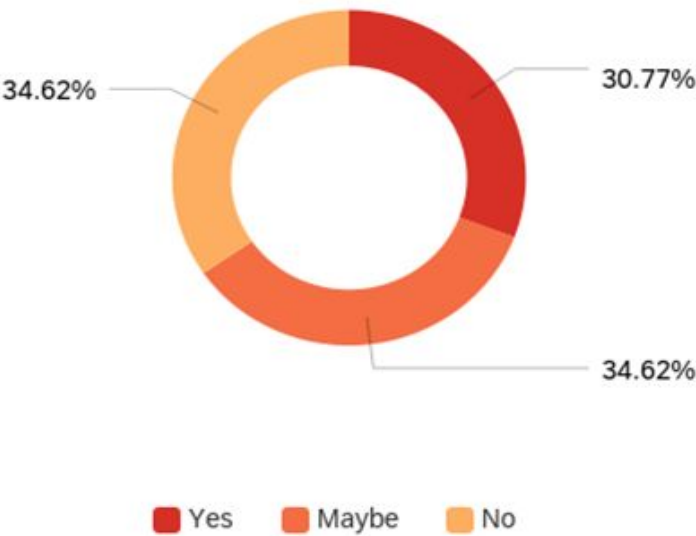
---

By writing many proposals.

**Scholars - Which of the following resources do you have available to you for use in your research?**



**Scholars - Does your university/institution have resources to handle/address these issues?**



Other

Other - Text

---

Research and travel support

---

A network of people concerned about the same issues, even if they don't have a good way to address them.

---

Peers in my own discipline to review my grants (though the quality is variable)

---

Colleagues willing to fashion new, collaborative paths

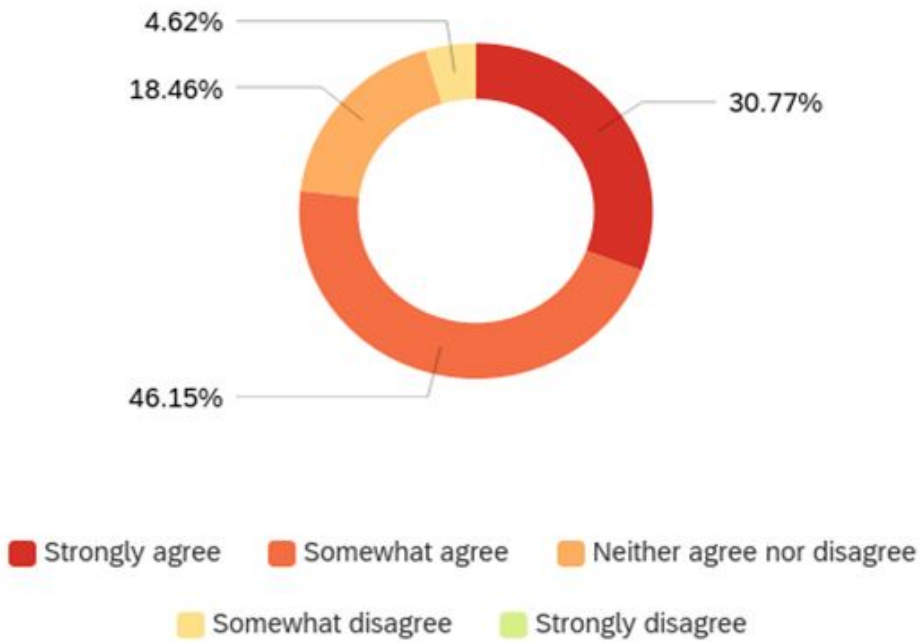
---

BEACON center for the study of evolution in action

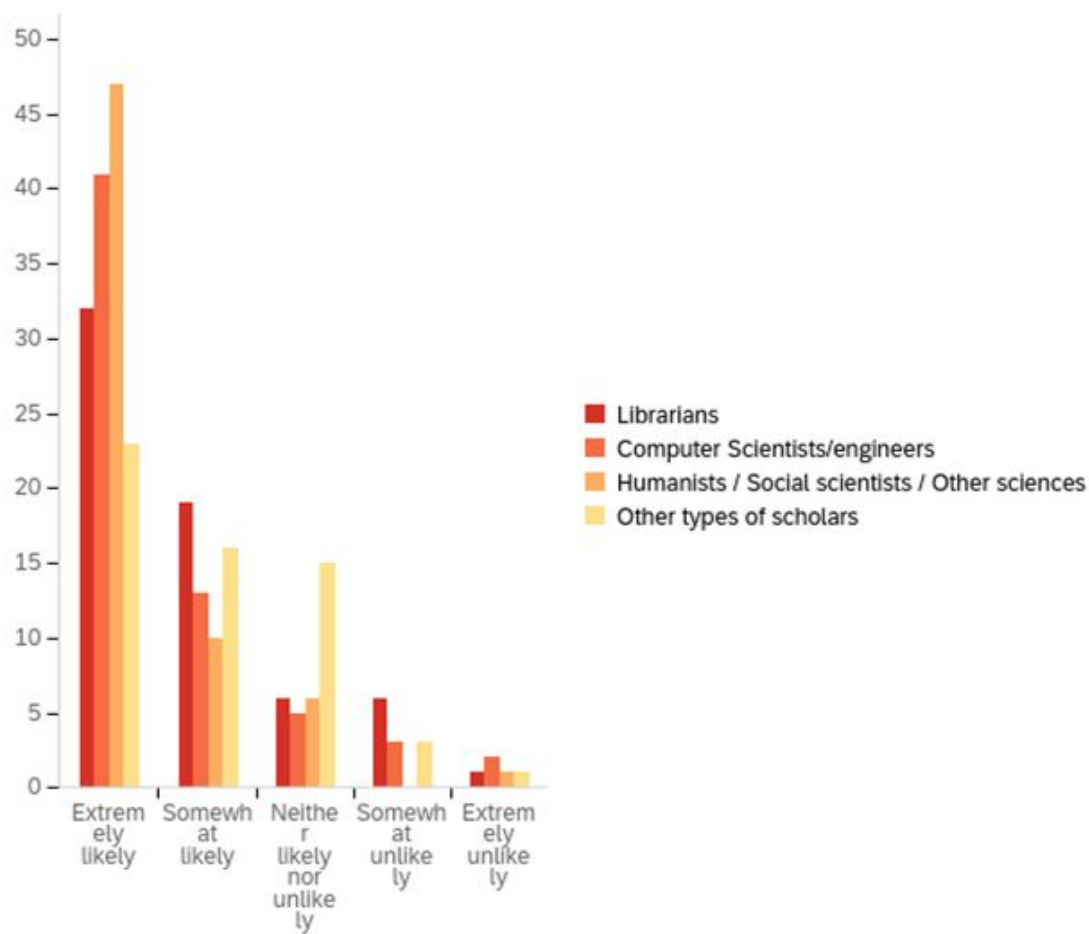
---

Soft support: encouragement, a welcoming culture

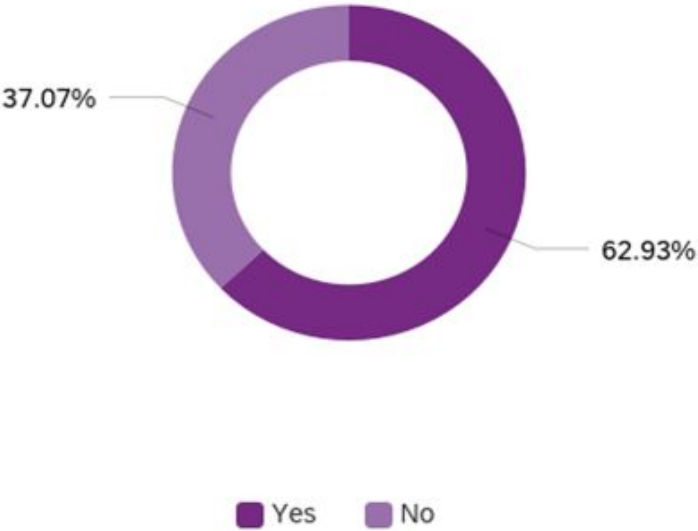
**Scholars - To what degree do you agree that terminology (jargon) is an obstacle to cross-disciplinary research?**



Scholars - How likely are you to collaborate with people from each of the following areas:

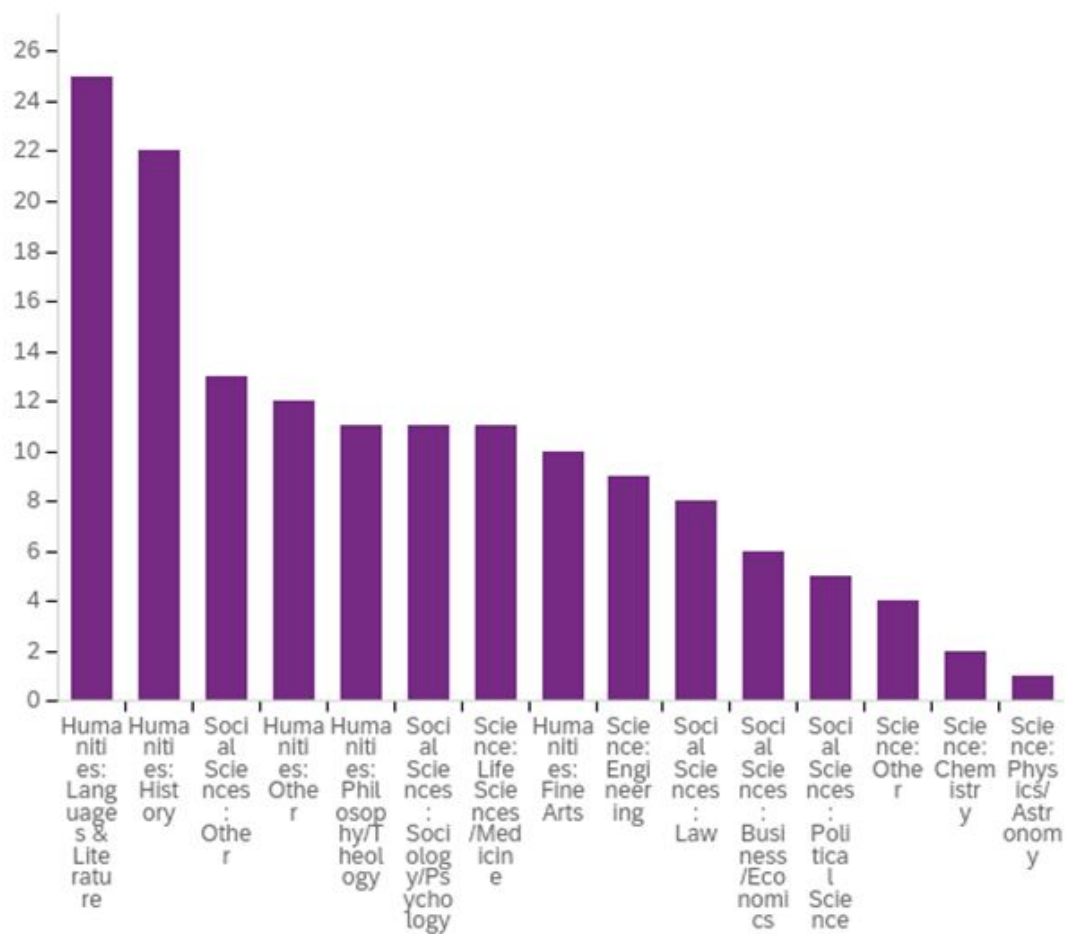


**Librarians - Do you collaborate with scholars from other disciplines in your research?**





**Librarians - For what disciplines are you producing cross-disciplinary research? (Hold Ctrl or Cmd key to make multiple selections.)**



**Librarians - In two or three sentences, please describe your cross-disciplinary research.**

In two or three sentences, please describe your cross-disciplinary research.

---

Supporting faculty and student cross-disciplinary research as DH librarian

---

Information science

---

---

As the librarian responsible for scholarly communications and library automation at the Federal Reserve Board, my primary concern is enhancing system interoperability, and much of my time is spent writing python applications for improved efficiency. One of my current projects involves collaborating with various stakeholders from across the Federal Reserve system to standardize and convert vocabularies / name authorities into RDF.

---

cds projects in text mining, topic modeling for history of science and for biology, grant projects in math and physical sciences(chem, materials science, physics) with an emphasis on HEP, computer science, informatics, digital preservation and metadata.

---

The work we do on ePADD puts us in touch with scholars from many disciplines who have different needs and expectations for a AI-facilitated tool focused on email, and involves collaboration with 3 CS PhDs that do not have background in librarianship or archival science. Many of the tools incorporating similar technologies like LTA are heavily used in the legal profession to support discovery. Likewise, the collections we bring in and make available in the University Archives cover a broad swatch of interest areas and are important to a variety of researchers.

---

I work as a librarian, but I have a PhD in history and continue to do my own research. In my own work, I have used topic modeling to perform a distant reading on a large corpus of documents from truth and reconciliation commissions. I was interested in seeing if topics emerged across these commissions, and what those topics might tell us about how we understand international law and human rights violations.

---

In archives and special collections, we support researchers primarily in history but also to some extent in literary studies and theology. Also sometimes in social sciences.

---

I work with anyone at the University, regardless of discipline. My current work involves mapping sexual assault within the Cleveland area with sociologist, examining hospital transfer networks and their inefficiencies with nursing, and generating 3D models of local historical resources.

---

I've done some work on what could maybe be called a sociology of theological knowledge... describing the task of theology with reference to Luhmann's systems theory, and Joas's social theory. I've also done some work on possibilities for indexing emerging fields like political theology research using topic modelling. And most of the theology work I do uses philosophy and, to a lesser extent, history.

---

---

Itemizing Japanese Military Maps from WWII period - the procedure crosses map making history, geography, GIS, linguistics (local area names vs how foreign troops named the areas).

---

I am currently a Co-PI on the project that examines the Chicago Public Library One Book One Chicago (OBOC) program that has been going on since 2001. We use circulation data obtained from Chicago Public Library, text features extracted from featured and recommended works, geo-locations extracted from the books and social media postings about the OBOC. A team of researchers works on this project, a GIS specialist, a computer scientist, a text mining specialist, and an English Book History specialist.

---

I am researching digital humanities with my sociologist friend. I think that counts even though he's a librarian too.

---

I work with education researchers to determine how graduate students learn best from research consultations with librarians.

---

Work with Social Work prof. to find best ways to use library resources to teach undergrads how to find law and legislation for their agency placements and areas of interest. Government documents are difficult to find and use, the topic is complex, and Social Workers at all levels need to be informed on laws that impact their practice. As a librarian, I'm always trying to find ways to help students, who will become practitioners, to find current policies and laws and use them to advocate for their clients and to influence policy-makers.

---

I support and collect materials and documentation related to human rights, which potentially falls within all of the categories you have listed.

---

We have worked between the libraries and the History department to use NLP and machine learning to analyze and visualize large collections of newspaper data. We are currently working with computer scientists to mine web archives for meaningful documents to add to the digital library. Finally we are working with our linguists to use machine learning for low-resource languages.

---

I have worked with the Sociology and Anthropology department on a collection of artifacts from around the world at a previous academic library. Having being 11 months in my current position, I am growing collaborative partnerships for digital curation and digital humanities work.

---

---

My research is in the humanities and the ethics of data science. As a digital scholarship librarian, I advise on research projects across the humanities and computational sciences.

---

I've served as the Library contact for grant proposals where the PI was in the History or English Department. The English dept project was a metadata standard and the History dept project was a transcription project. I've also worked with faculty in the Computer Science department to make Library data/imagery more available for data visualization projects.

---

Work with faculty in Architecture, Dance, History, and Africology to develop new narratives and evidence about living in the city/choosing the city; work with faculty in Languages/Literature and Theater to develop digital archive of Yiddish theater synopses and encode and provide new ways to search, visualize and otherwise discover relationships between plays, people, places, etc.

---

My research program focuses on the intersections between sound, listening, and knowledge-making in the humanities. I am interested broadly in how sound and listening are currently used by humanities researchers to create knowledge and am especially interested in ways that virtual research environments (VREs) and digital humanities tools in general support, supplement, transform, or interrupt research.

---

See earlier response.

---

As noted in the previous steps, one project we are currently working with a biology professor who has over 40 years of Right Whale data. Another project involves, developing ML models to generate automatic captions/labels for images in library archives.

---

We are currently developing and testing machine learning systems and applications for the analysis of document-based cultural heritage materials, where we analyze textual materials as images for the purposes of content-identification as well as artifact-level analysis. Currently, we are focused on historic newspapers, and have an IMLS grant to explore the generalizability and scalability of our methods and approach.

---

English prof and librarian looking at the rhetoric involved in data visualization creation

---

My research is in building research infrastructure (like the HathiTrust Research Center) that has broad applicability across the disciplines.

---

---

Developing taxonomy with input from economics subject matter experts

---

Big sensor data processing, esp identify patterns.

---

Working with education faculty to look into the community health problems, find out what factors contribute to the risk and how to educate them.

---

I currently am involved in an NEH-funded project investigating demographic change in northwest Indiana. The project involves historians, computer programmers, English faculty, librarians, and others.

---

My cross-disciplinary research has been mostly in the social professions (education, counseling), finding literature and using it to align the methods and data collection techniques to match with existing best practices in analysis in that field. I also facilitate research planning and content discovery, which may or may not become a co-authoring role.

---

One of the research is to discover how a complex object of social interest is constructed/represented in different disciplines by analyzing their specific communications (e.g., scientific papers, news articles, speeches, and etc.) using text mining techniques in the philosophy area. Our goal is to develop R and Python packages. The other study was to identify trust factors in data reuse.

---

My research interests are in the visual literacy competency in cataloging staff and implicit bias awareness training in cataloging and metadata creation work. As manager of cataloging and metadata creation work, I'm very interested in the automation of "routine" description work that happens in libraries, freeing up humans for "heavier" intellectual work.

---

I am a librarian working to better understand the nature of mathematics research. This is primarily focused at the moment on scientometric impact work but I hope to move into more textual and thematic analysis in the future. For this work I collaborate with mathematicians and information scientists.

---

Scholarship on preserving and making use of digital content for libraries, archives and museums by a wide range of researchers in various disciplines.

---

---

I am involved in research projects developing archiving and research environments for 3D scans of fossils and other morphological specimens. Our goal is to make accessible fragile, rare physical specimens in digital research environments.

---

My most significant foray into cross-disciplinary research has been an ongoing study using bioinformatic methods to better understand "novelty" (or newness) in literary texts. I have also collaborated with other researchers on campus, typically with non-data scientists who are seeking a collaborator to aid with data and computational text analysis.

---

Librarian working with faculty on digital humanities or library-owned projects.

---

I am currently working with a Computer Science student, a Digital Initiatives Librarian, and community partners at the Texas After Violence Project, an oral history non-profit focused on how criminal justice policies impact families and communities across generations. We are setting up a text analysis project using R to analyze a data set consisting of transcripts of all of the oral histories the organization has ever conducted. We are using an index of psychological terms in order to explore patterns in the language of trauma that emerge across the collection.

---

virtual reality, augmented reality, and other emerging technologies enhance learning, teaching and research in libraries.

---

Research in Library Carpentry focuses on library and information science but this community provides services to a number of research disciplines. This also applies to The Carpentries with our primary audience being early career researchers.

---

It's mostly me facilitating the research of other faculty. They have done cross-disciplinary research ranging from data science & aviation, to English & computer science, to medicine & computer science, and more.

---

Generally use of Digital Humanities as relates to library and archives holdings.

---

My research crosses women studies, business, education and library science. Specifically I research gender and organizational citizenship in library technology..

---

I combine library and information science research and elements of computer science, statistics, and began work on biology inspired work such as evolutionary algorithms more recently.

---

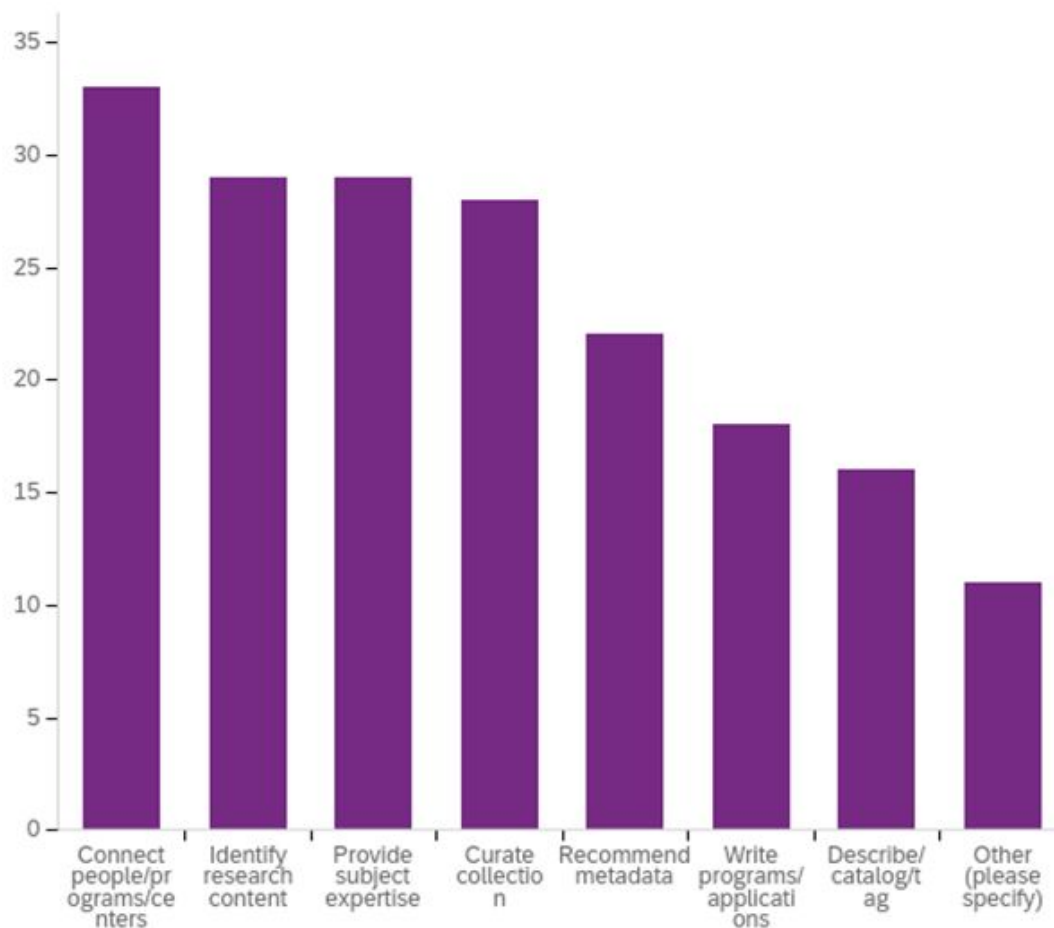
---

As a librarian and technologist, I do not conduct research directly. Rather, I and my group work with researchers across Stanford on providing discovery and other information services that meet their particular domain needs. This includes services such as combining and providing access to specialist areas like gov docs, web archives, CJK materials, geospatial assets, image sets, tabular data, full text articles, digitized books, et al.

---

My own research is on scholarly communication in science and the role of social media. I also study bibliometrics. I've worked on teams doing knowledge systems for counter wmd, doing mixed initiative systems for supporting analysts, and a bunch of other things.

**Librarians - What is your role in facilitating cross-disciplinary research/scholarship? Please select all that apply.**



Other (please specify)

Other (please specify) - Text

---

visualize/analyze/communicate results

---

I also work as a co-investigator on projects

---

Digitize content for use in project

---

Assist with technology work plans for grants, assist with infrastructure issues, such as platforms for digital exhibits

---

Serve as principal investigator on cross-disciplinary research team

---

Build research infrastructure

---

build cyberinfrastructure

---

Suggest resources or tools

---

Technical leadership

---

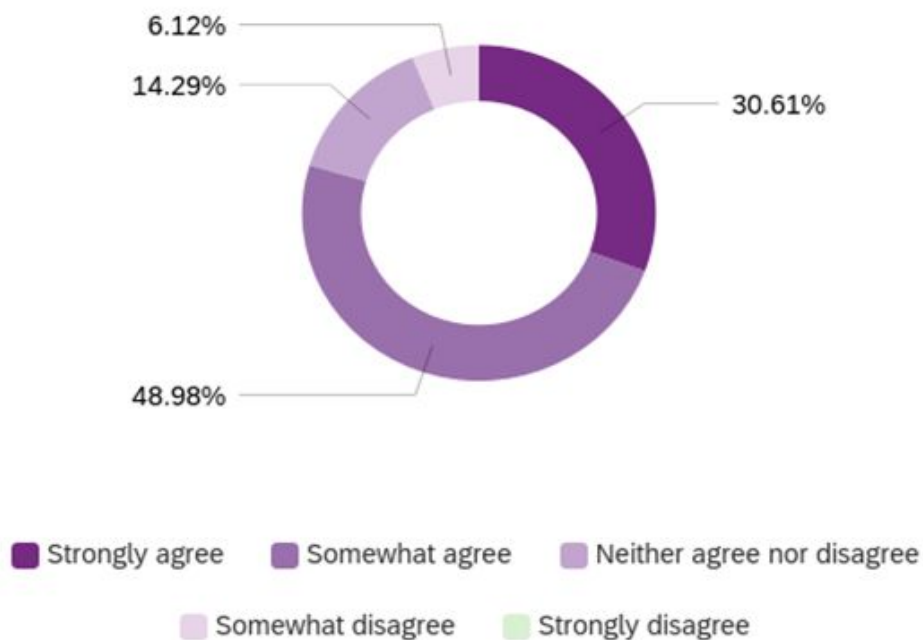
Instructional

---

Technology strategy, interoperability, information service development



**Librarians - To what degree do you agree that terminology (jargon) is an obstacle to cross-disciplinary research?**



**Librarians - When doing cross-disciplinary research/scholarship, what obstacles do you usually encounter?**

When doing cross-disciplinary research/scholarship, what obstacles do you usually encounter?

- 
- Institution/department says it supports, but promotion & tenure process says otherwise
  - Finding the right technical support on campus (eg who should by the 3d camera - grant money, IT dept, library?)
- 

time and coordination of efforts between faculty with different schedules and additional research agenda.

---

The biggest obstacles tend to be related to expectations and communication.

---

---

disciplinary scientists' lack of familiarity with standards, standards based approaches, mistaken assumption in disciplinary scientists who unaware of an ontology, tool or method in their discipline assume it doesn't exist/needs to be built (and that they are the \*right\* ones to build it) even if it already exists in another domain. Scholarly naiveté when it comes to how much time it takes to train/test and build up to usefulness machine learning or cognitive computing solutions so that they are more thorough, accurate and/or more efficient than other techniques. disciplinary scientists' lack of familiarity with standards, standards based approaches, mistaken assumption in disciplinary scientists who unaware of an ontology, tool or method in their discipline assume it doesn't exist/needs to be built (and that they are the \*right\* ones to build it) even if it already exists in another domain. Scholarly and informatics naiveté when it comes to efficiently using ontologies, taxonomies and digital thesauri in information systems for search expansion, for synonymization, for facilitating cross-disciplinary research.

---

Differences in goals, familiarity with the technologies, resources

---

Disciplines may use terminology in different ways. There may also be assumptions about methodologies that not all disciplines use.

---

Primarily financial, and lack of time.

---

There is strong siloing of disciplines as well as long-standing political barriers. Additionally, discipline specific demands are placed on faculty that can often steer them away from interdisciplinary work that has no guarantee of success. Finally, digital projects that are heavily interdisciplinary often don't meet the criteria for tenure, which favors traditional publications at the individual level.

---

Especially within the humanities or crossing over between humanities and social sciences, one real difficulty is that in investigating the same phenomena, different things are interesting/relevant to research. This especially plagues the discourse between theology and sociology... [two authors] are very good about making these obstacles explicit. This isn't necessarily about the problem of jargon, although jargon is a relatively superficial version of this problem. The problem is most basically one of an underlying logic, underlying assumptions, then the methodology that follows from these assumptions and logics, and then, finally, the jargon.

---

Resources (cross reference that I want to use) not found.

---

---

Cross-disciplinary research implies a team of people working on a common research problem. Everybody on the team is likely to have other priorities and engagements and it is not possible to find time for a meeting that is suitable for everybody. Not having enough resources to obtain support for certain tasks.

---

controlled vocabulary, reusable data, identifying potential collaborators.

---

terminology in Social Work and law; government documents complexity

---

Differences in how information and knowledge is produced means that there is no single discovery mechanism.

---

Having some groups understand the possibilities, while having the other groups understand some of the limitations of either the data or the cultures that are involved.

---

To be cognizant of each other field and to clearly identify one's responsibility and role in that collaborative partnership is critical.

---

Quickly immersing myself in the discourse, values, methods, and conventions of a new discipline.

---

People in other disciplines tend not to recognize how much work/time the Library invests. "Digitization" is seen as a service like photocopying, without regard to the time it takes to create identifiers and metadata, track how much of a collection has been reformatted, preserve the reformatted masters etc.

---

Lack of resources in terms of time and people; uncertainty about technology needed and where to secure the right infrastructure to support a project, along with skill sets.

---

Not all humanists think that data applies to their fields

---

Lack of computer science background/programming skills, familiarity with tools

---

We need to deliberately make space for research in various fields and to give project members the leeway to pursue their own disciplinary-relevant work as part of a multi-disciplinary project.

---

---

Sometimes there is no clear agenda. Different groups seem to be talking about different things, and talking at each other. Vocabulary is definitely an issue.

---

Making assumptions about one another's methods and/or assuming a shared base of knowledge. I think terminology may be a symptom of this larger obstacle but is not the obstacle itself.

---

not being familiar with other departments, e.g. who to ask about things, any political issues, funding situations

---

Different research-cycle periods (time to publication), different notions of research, different goals.

---

Organizational barriers.

---

Especially in INFORMATION TECHNOLOGY, ironically, related disciplines are stove-piped in different buzz words--this allows a researcher to claim that something is novel when really all that is new is the name. There are solid industry-specific incentives for this, but it holds the field (and other fields with similar situations) back.

---

narrow-mindedness, too quick to judge

---

Domain knowledge Tools for the discipline

---

Differing mental models for what we mean by "research." Because my abovementioned project is a digital humanities project, that terminology has confused some involved.

---

major differences in jargon, especially where two disciplines use the same term differently. I've never co-authored with a group that crosses science/social science/humanities boundaries, but I've been a facilitator with such broad groups. They can easily break down in the differences in priorities of what counts as an appropriate research goal, as well as getting tangled up where we misunderstand what each other are saying and have to backtrack to untangle the conversation.

---

Either as a librarian or data analyst, I always need a bit of time to fully understand subjects or topic at the beginning.

---

---

Data are described very differently across knowledge domains. Considerable effort is needed to understand not just the dataset itself but the ontologies, vocabularies, and conceptual models that impose structure on a dataset from a specific domain.

---

Mathematics has a very steep language learning curve which makes it seem that they are often speaking a different language than other people.

---

Aligning incentives and resources to support collaborations between different kinds of researchers and organizations.

---

Terminology is indeed one, but priorities of research output varies from discipline to discipline. Metadata schemas and organization of data are other obstacles.

---

I think the main obstacles in my experience may be anxiety in the form, say, of imposter syndrome, or an arrogance about the importance or centrality of one's own discipline, maybe often some mix of both in a single person. In any case, anxieties can limit dialogue in cases where the fear of being wrong is too high.

---

agreement upon level of detail for description

---

Acquiring and organizing pertinent data sets. Finding time to build relevant skills, for instance in R.

---

There is the language barrier like the jargon question before (we have an activity called jargon busting that responds to this) but having a shared experience to also help with the language barrier is something we try to improve with hands-on training and activities.

---

Problematic terminology, data fields that don't have the same meaning across disciplines

---

Different foundations

---

The ontology is often different.

---

Funding issues to get started on new projects.

---

---

Optimization for one domain's assumptions, metadata or practices can reduce accessibility by others; finding a sweet spot for "good enough", where the perfect doesn't become the enemy of the good; lengthy discussions in the abstract about potential approaches and hiccups with approaches (analysis paralysis)

---

Sometimes lack of acceptance of work coming from a different tradition.

---

Experience and expertise

### **Librarians - How do you currently address these obstacles?**

How do you currently address these obstacles?

---

not well; scholar's own social network determines outcomes not institutional support

---

I end up being the coordinator or the project falls through.

---

I think it is best to assess the skill sets of those involved in the research and strategically divide the work. Taking some time early on to set expectations and establish a common vocabulary for the project is also helpful.

---

By extending basic schema with disciplinary schema, by doing cost benefit analysis, ROI calculations, tests and prototypes to help researchers understand the long on-ramps to creating efficiency or realizing the accuracy or scale that machine learning make so tantalizing.

---

Identifying shared goals, establishing buy-in and valuing contributions from developers, subject experts, and others

---

It's important to come to a shared understanding of vocabulary at the beginning of a project. It's also crucial that project roles are well-defined.

---

Basically by not pursuing any projects that are beyond our resources.

---

---

I handle these on a case-by-case basis, but have no overarching solution. I am most successful when I can bring a few like-minded people together and get them talking, but just getting them in the same room is a challenge!

---

I find it takes a lot of basic reading in the non-native discipline to understand what sorts of questions are most relevant for that discipline. There is an ongoing conversation in the literature and it feels deceptively as if you can just jump into it because you're all talking about, e.g., "religion". But without going back and starting from scratch with the discipline you're less familiar with, you can be talking past each other.

---

Ask around.

---

Giving ourselves time to work on research. Acknowledgement that research takes time. Trying to apply for grants to try to provide more resources.

---

I use a variety of discovery systems to help people find funding and connect to other researchers.

---

help students with gov doc tools and how to construct search terms, including controlled vocabulary

---

One must search many different sources using several techniques and methods.

---

Lots of discussion between groups, examples, demonstrations and so on.

---

To have a clear definition of the process, dialogue, transparency, and deadlines.

---

Reading disciplinary literature.

---

I don't think we've addressed them sufficiently. Even if I bring a particular faculty member up to speed, the next one will approach us with the same misunderstandings.

---

Work across the library and campus to identify skilled staff who can help with aspects of projects; identify appropriate resources for infrastructure and work with IT staff both in the Library and at the campus and college level to ensure support.

---

---

Relying on examples of finished projects

---

Will soon be pursuing a computer science degree

---

Our project management methods give project members significant space to define the project requirements and their role in the project given their research needs.

---

Writing research statements, and focusing on simple goals helps. Clarifying and re-inforcing the meaning of key words also helps.

---

Asking lots of questions, explaining where I'm coming from from my own disciplinary backgrounds, writing about the work from our various disciplinary frames

---

try to ask who I do know for support

---

Directly.

---

Talking with individuals in each department.

---

Taxonomy, subject matter expertise. These issues are not nearly as pronounced in economics.

---

trial-and-error

---

Figure out on my own

---

Unnecessary repetition and clarification

---

Try to bring the conversation to basic topics and terms where possible ; Define shared goals and clarify what those goals look like ; break down segments of the problem for assigning leadership

---

I start with Wikipedia articles and then quick literature review. I also read my collaborators' previous publications.

---



---

I usually need to understand the metadata schema (if any) that is being used to describe data, so I start there. I've sometimes found that I don't have the expertise in certain software or subject areas (e.g., creating metadata descriptions of GIS data to increase access).

---

I have done a large amount of work as a mathematics communicator so I function as a translator when the language becomes an issue.

---

Work to understand motivations of different kinds of stakeholders.

---

Many hours of human engagement, learning about context, and working to solve problems.

---

I think that courtesy, even excessive politeness and respect for each other's work can go a long way.

---

only offer what we can resource

---

By ensuring that we have the right expertise and exposure distributed across our research team.

---

Mentioned before but here it is again: jargon busting, hands-on training with lessons, in-class activities

---

A LOT of custom programming

---

Usually a lot of discussion

---

I learn the different ontology in each subject dictionary.

---

Writing grants or finding other funding sources of seed-funding.

---

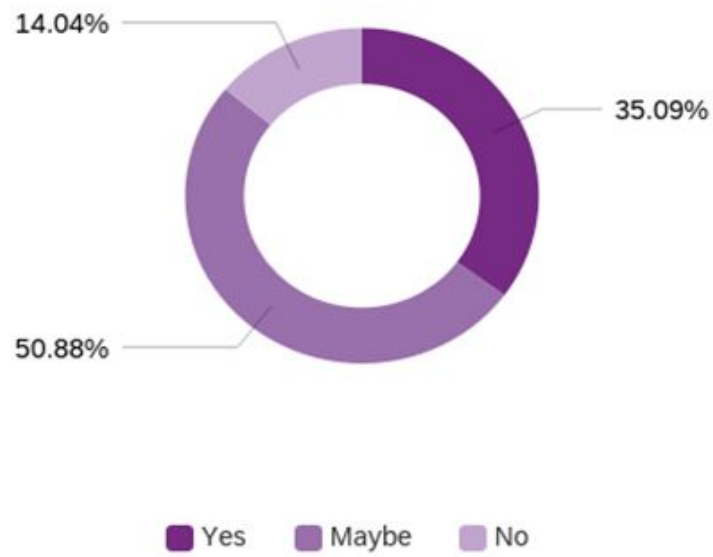
User-centered research on defining specific objectives and "stories" (in the agile sense) with concrete value; initiating concrete experiments and project work to give enough of a prototype that can then be iterated on and improved over time.

---

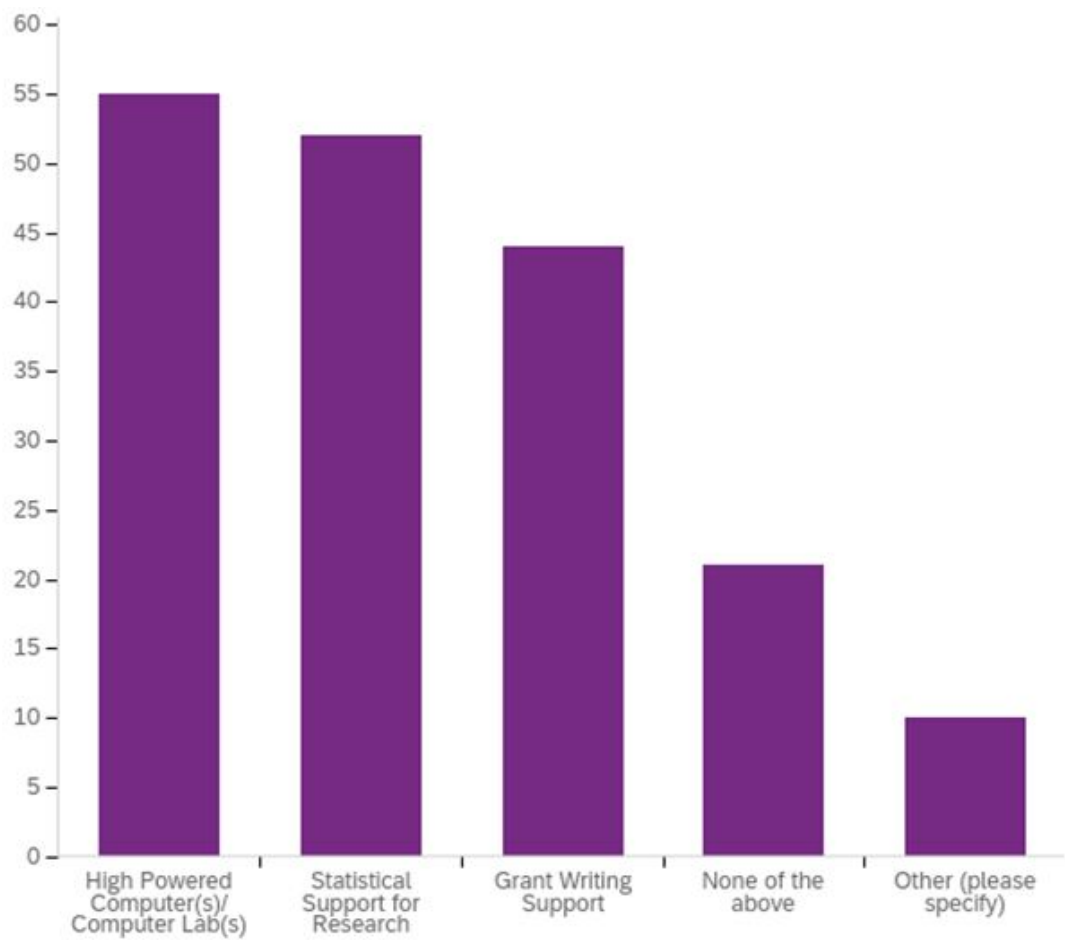
---

DIY learning

**Librarians - Does your library have resources to handle/address these obstacles?**



**Librarians - Which of the following resources do you have available to you for use in your research?**



Other (please specify)

Other (please specify) - Text

Data visualization and GIS support

Fellowships / Endowments

I am a librarian. I am not doing research but am producing projects.

n/a

Research Data Services

---

Metadata architect

---

All these options are available to faculty at my institution, however, as a non-faculty librarian they are not available to me

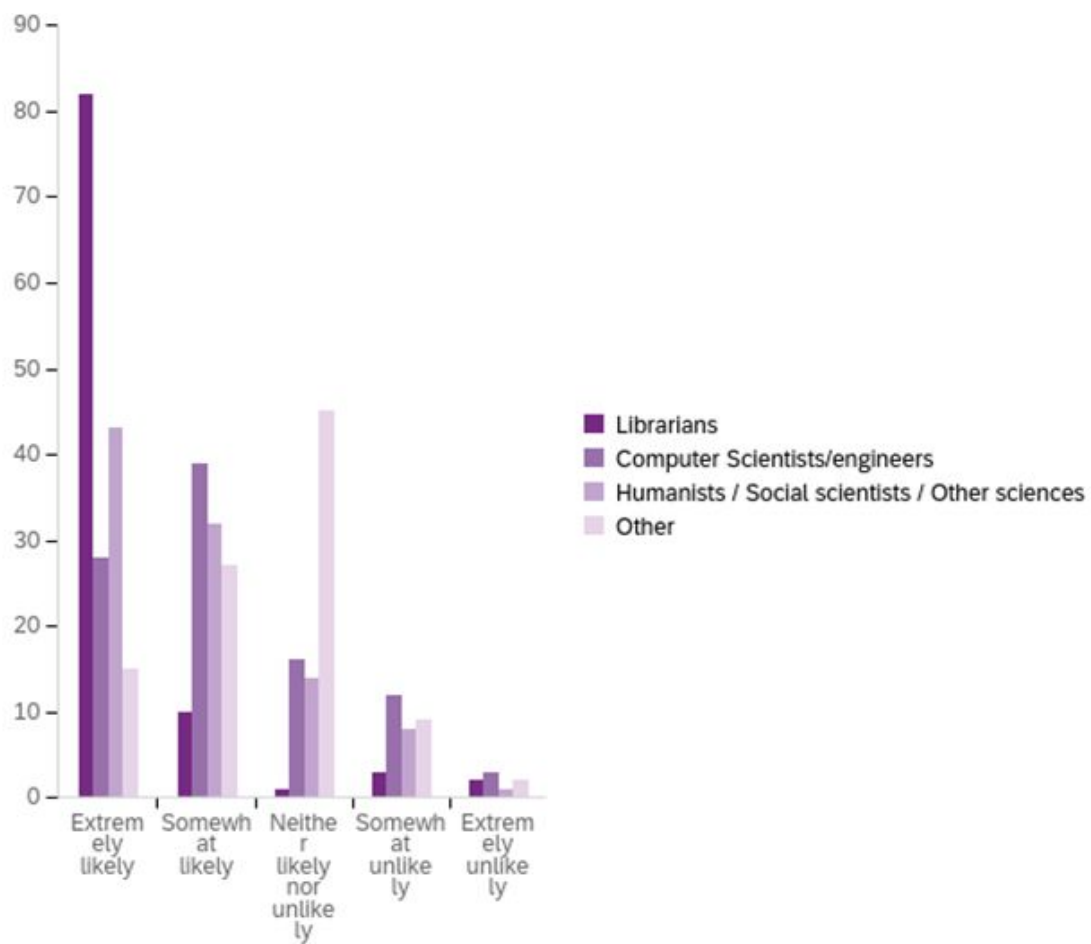
---

We have a community of over 2K instructors worldwide

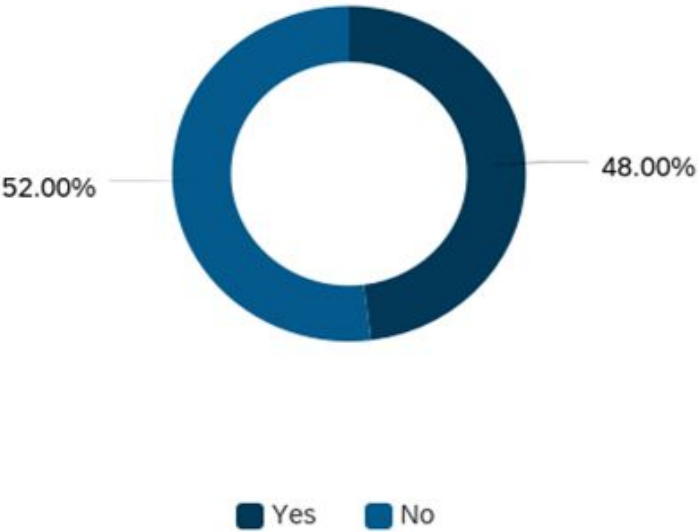
---

bank of virtual computers of various types

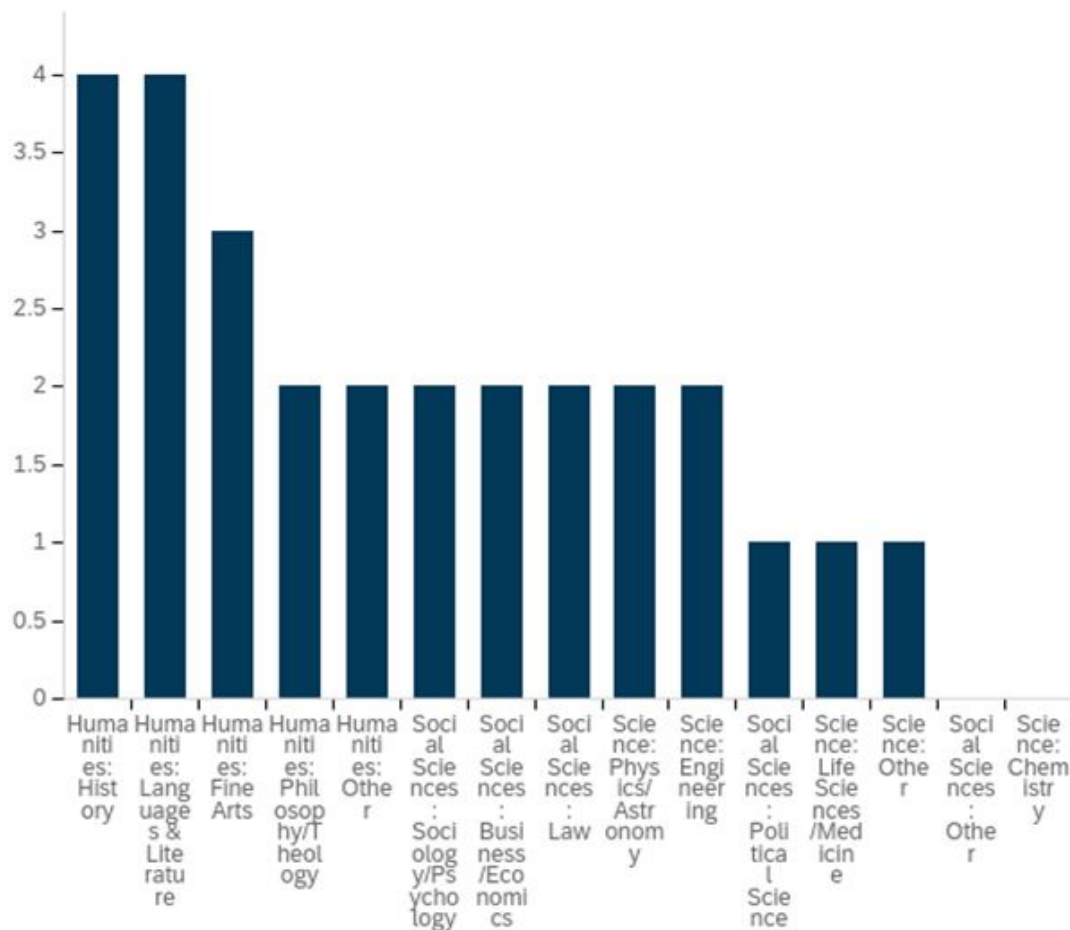
**Librarians - How likely are you to collaborate with people from each of the following areas:**



**Computer scientists - Do you collaborate with scholars from other disciplines in your research?**



Computer scientists - For what disciplines are you producing cross-disciplinary research?



**Computer scientists - In two or three sentences, please describe your cross-disciplinary research.**

In two or three sentences, please describe your cross-disciplinary research.

---

I may not have responded properly because I wouldn't say what I do myself is research. I attempt to be an enabler and enthusiastic cheerleader for these new possibilities. I work in the historical collection area of my institution and part of my job is to enable research of others.

---

We've used book metadata to demonstrate the spread of new concepts. This is important to economists

---

Data science

---

I support research activities in a national library

---

I am a computer programmer who works with researchers at [institution] from all disciplines.

---

Our research is to reveal the hidden patterns in student learning behavior, discover the strongest predictor for successful learning outcomes, and identify and boost students who are not thriving academically.

---

I have worked with linguists, since my research focuses on developing machine learning tools that pertain to endangered and low-resource languages, with the ultimate goal of helping in the language documentation process.

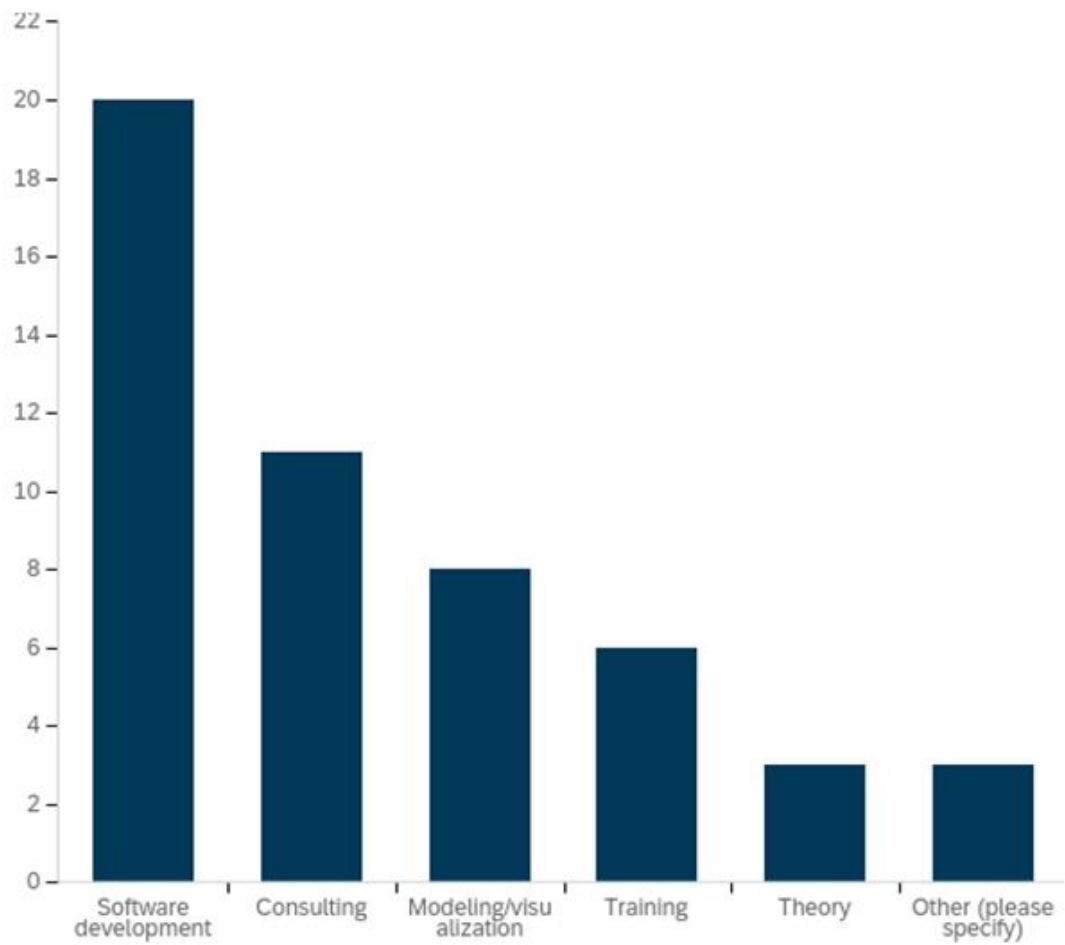
---

Just starting a project to train a model to identify the copyist for a corpus of musical scores.

---

I may have not answered this correctly as I am not a researcher myself, but am part of an informal team involving historians and others here at USHMM.

Computer scientists - What is your role in facilitating cross-disciplinary research/scholarship?



Other (please specify)

Other (please specify) - Text

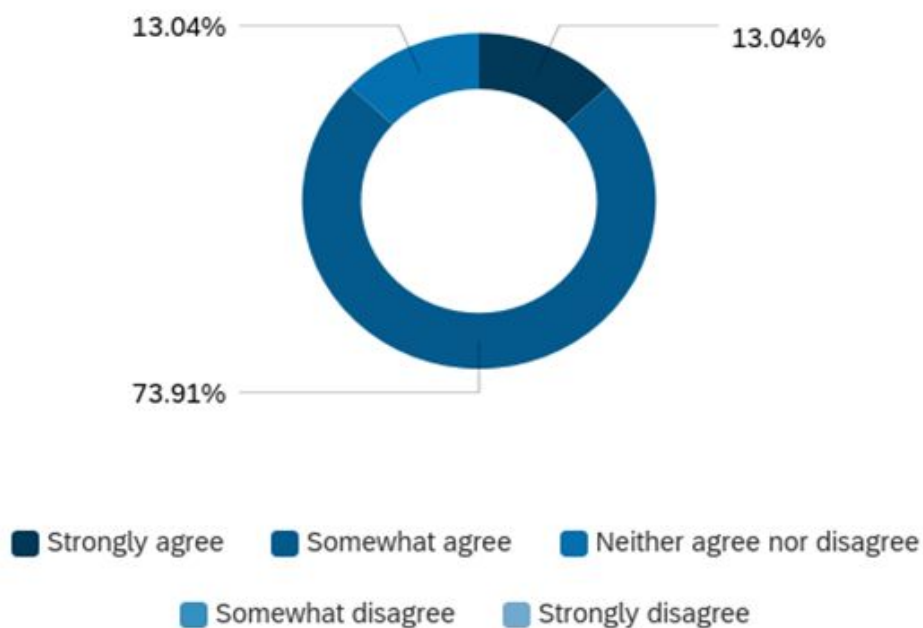
Attempting to expand management support for these activities

Server infrastructure for ML

Manager (overseeing other devs)



**Computer scientists - To what degree do you agree that terminology (jargon) is an obstacle to cross-disciplinary research?**



**Computer scientists - When doing cross-disciplinary research/scholarship, what obstacles do you usually encounter?**

When doing cross-disciplinary research/scholarship, what obstacles do you usually encounter?

---

Funding sources and managerial support move gradually along with interest in the field. I think it's a matter of time for the opportunities to become more widely understood.

---

a fundamental lack of understanding about how filthy data actually is

---

Workflows, data management, jargon.

---

different operating systems, lack of disk space on user's machine, lack of training, question that's difficult to operationalize...

---

---

As a computer scientist, I often struggle with the theoretical aspects of linguistics, as they often don't have an equivalent in computer science. On the other hand, many computer scientists are not familiar with core linguistics theories.

---

Balancing priorities of various individuals.

---

At this point finding the right kind of people who can do this type of work is itself the first obstacle. There are other barriers (e.g. Information security is concerned about allowing software running on machines that access our network, so we find ways to copy digital materials to a local machine for processing)

### **Computer scientists - How do you currently address these obstacles?**

How do you currently address these obstacles?

---

good examples

---

One-on-ones

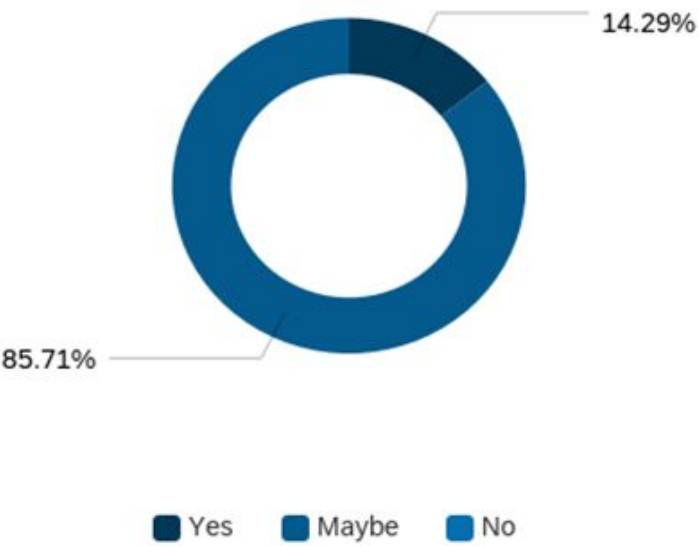
---

By using virtual machines, hosting training workshops, and transforming questions into those that can be more directly operationalized...

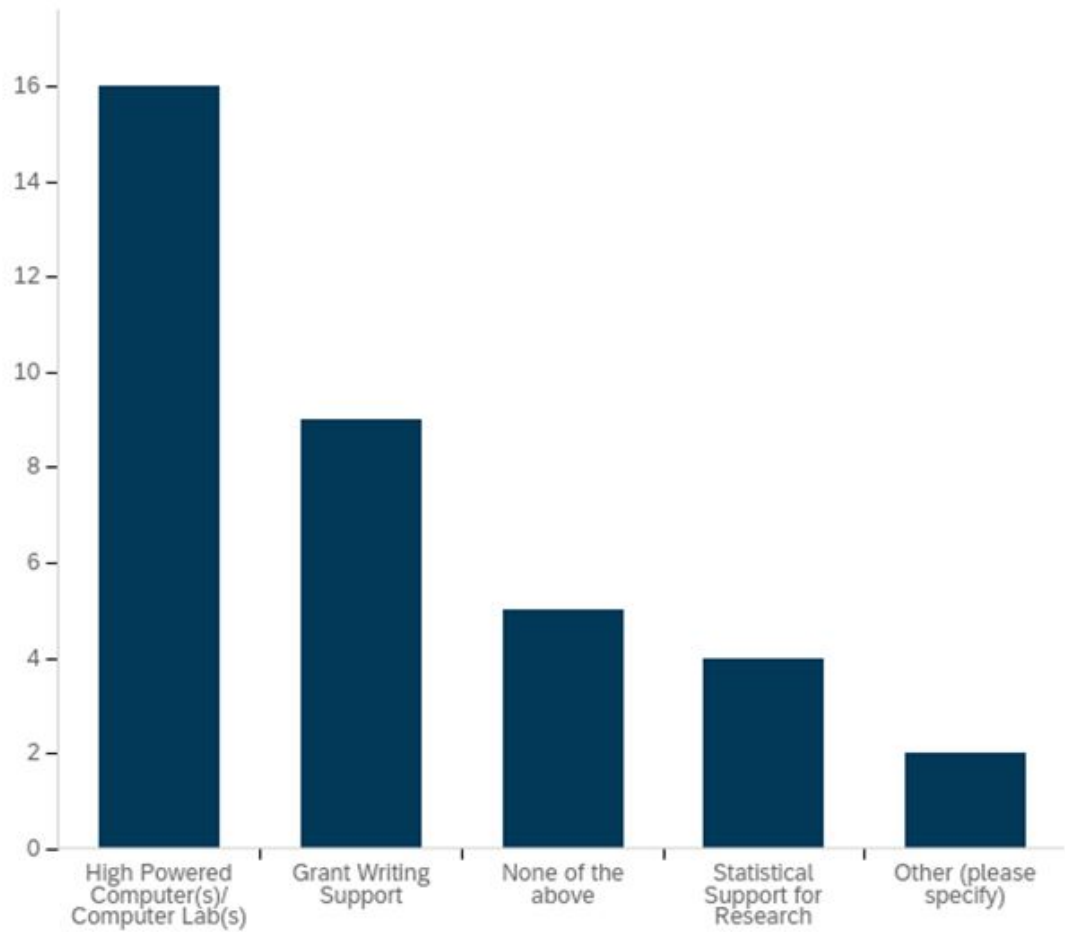
---

We are pretty new at this but once we find the right people, we've found ways to make progress.

**Computer scientists - Does your university/institution have resources to handle/address these issues?**



**Computer scientists - Which of the following resources do you have available to you for use in your research?**



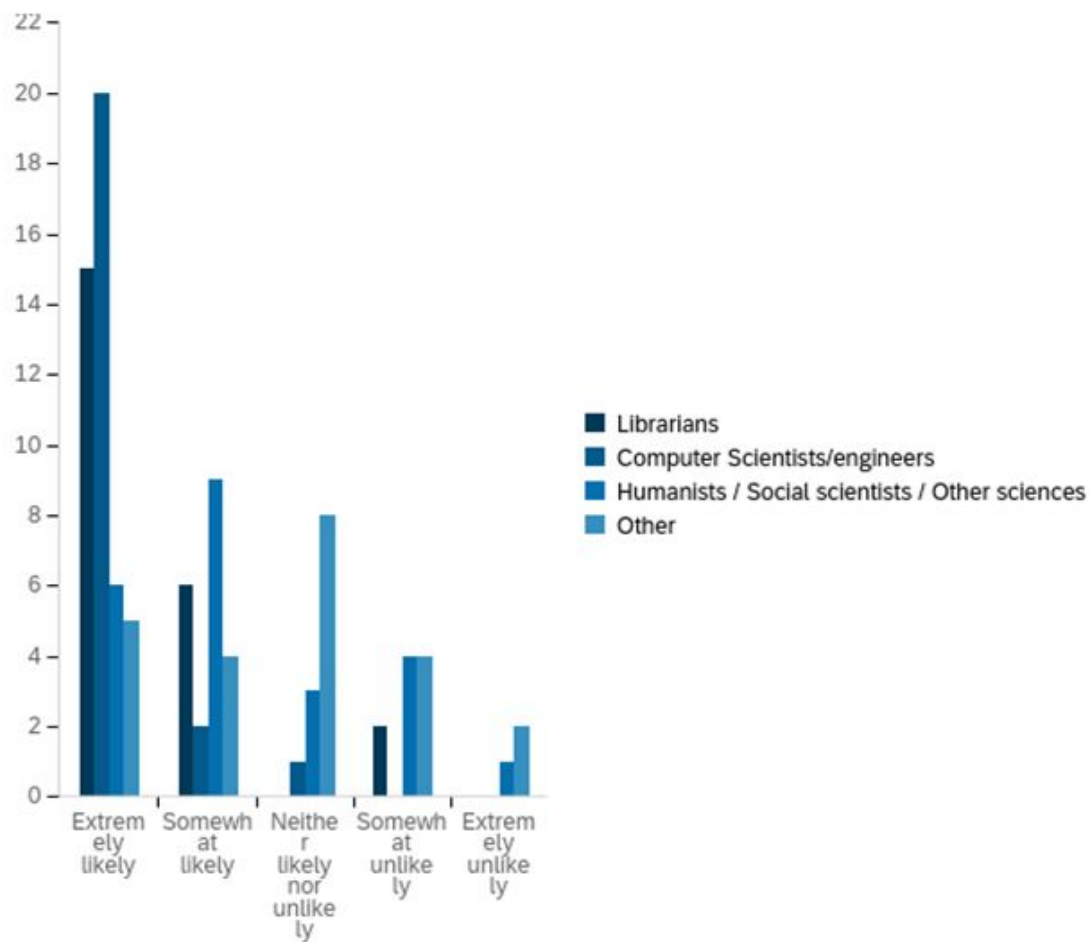
Other (please specify)

Other (please specify) - Text

Software application support for research

Good funding

**Computer scientists - How likely are you to collaborate with people from each of the following areas:**



**Everybody - What is your job title? (Removed for respondent privacy.)**

## Appendix C: Workshops

### Notre Dame Workshop

*Monday, March 11, 2019*

*Location: Hesburgh Library, room 246*

#### Agenda

- |       |   |                   |
|-------|---|-------------------|
| I.    | Lite Breakfast/Socialize  | 08:00am - 08:30am |
| II.   | Introduction: Why We're Here  | 08:30am - 09:00am |
| III.  | Lightning Talks and Demos - What We're Doing  | 09:00am - 10:15am |
|       | A. "Social Systems Artificial Intelligence" Tim Wenninger,<br>Professor of Engineering, University of Notre Dame  |                   |
|       | B. "A Personalized Account-based Recommender Within<br>the University Library Mobile App" Jim Hahn,<br>Orientation Services and Environments Librarian, University<br>of Illinois, Urbana-Champaign |                   |
|       | C. "Image Analysis for Archival Discovery" Elizabeth Lorang,<br>Associate Dean and Humanities Librarian, University<br>of Nebraska, Lincoln   |                   |
|       | D. "Expandable AI" Amy Guarino, Learning Analytics<br>Architect, Kyndi  |                   |
| IV.   | Break   | 10:15am - 10:30am |
| V.    | Discussion About Approaches   | 10:30am - 11:30am |
|       | A. What kinds of tools do you use?  |                   |
|       | B. How do you approach this type of research?   |                   |
|       | C. Clustering versus classification   |                   |
|       | D. Unsupervised versus supervised   |                   |
| VI.   | Lunch   | 11:30am - 12:30pm |
| VII.  | Live Survey About Interest/Experience in Classification<br>Text mining, etc.  | 12:30pm - 12:45pm |
| VIII. | Discuss Survey Results, Identify Breakout Categories  | 12:45pm - 01:30pm |
| IX.   | Deficits in Tools for Cross-disciplinary Research   | 01:30pm - 02:30pm |
| X.    | Break   | 02:30pm - 02:45pm |
| XI.   | Cross-Professional Breakout Discussions   | 02:45pm - 03:45pm |
| XII.  | Report Out  | 03:45pm - 04:30pm |
| XIII. | Wrap Up - Conclusions, Next Steps   | 04:30pm - 05:00pm |

## Workshop Participants

<b>Name</b>	<b>Institution</b>
Jennifer Liss	Indiana University
Amy Guarino	Kyndi
Arend Hintze	Michigan State University
Mark Philips	North Texas University
Mark Robison	Notre Dame
Anibal Perez-Linan	Notre Dame
Melissa Harden	Notre Dame
Madhav Joshi	Notre Dame
Meng Jiang	Notre Dame
Xiaoqing Duan	Notre Dame
Christina Leblang	Notre Dame
Tim Weninger	Notre Dame
Jeffrey Bergstrand	Notre Dame
Patrick Flynn	Notre Dame
Mark Dehmlow	Notre Dame
John Wang	Notre Dame
Alex Papson	Notre Dame
Helen Hockx-Yu	Notre Dame
Don Brower	Notre Dame
Dan Johnson	Notre Dame
Eric Morgan	Notre Dame
Laurie McGowan	Notre Dame
Nastia Guimaraes	Notre Dame
Matthew Hannah	Purdue University
Dean Lingley	Purdue University
Sue Wiegand	St. Mary's College
Jim Hahn	University of Illinois, Urbana-Champaign
Samuel Hansen	University of Michigan
Elizabeth Lorang	University of Nebraska- Lincoln
Douglas Duhaime	Yale University

## Afternoon Discussion

### Tools and Cross-Disciplinary Research Challenges (Will Become Brainstorming Activity in Subsequent Workshops)

- Tools for identifying and mitigating bias are needed.
  - Toolkit for thinking critically about algorithms
  - AI discriminates
    - Recommended reading:
      - “Algorithms of Oppression” - Safiya Umoja Noble
      - “Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.” - Virginia Eubanks
    - Datasets should not be built with common or easy data in order to facilitate variety and inclusion.
      - IE, Moby Dick as a sample
- Library of Congress Subject Headings
  - Not always helpful because they did not help facilitate cross disciplinary research (re Convocate)
    - Too broad
    - Accurate, but not granular
- Collaboration should be incentivised
  - Some our apprehensive about cross disciplinary research
    - Protecting one’s research
    - Hard to determine who’s research is who’s, who gets the credit?
- Jargon, Taxonomy, Vocabulary
  - Disciplines use different phrases meaning the same thing
    - Patterns with Features vs. Observations with Variables
    - Font faces mean different things in different places
    - Vocabulary within different areas of study within the same discipline isn’t always the same
- There is a challenge gathering data from different disciplines when you aren’t sure what you are looking for
  - Is there a way to increase synergy between data of different disciplines
- Recommenders
  - Can you intentionally bias recommenders? (yes!) (it already happens)
    - Is that ethical?
  - Do recommenders create bias inherently?



## Breakout Discussions

### *Topic: Library Collections as Data for use in ML-based Research*

#### Group 1

- Opportunities
  - Getting more value out of existing assets
    - A lot is buried (forgotten)
  - Eliminate silos
    - Solve problems faster
  - Understanding uniqueness in collections
  - Leveraging engineers and infrastructure
  - Finding opportunities for collaboration
  - Crowdsourcing
  - Fostering new ways of teaching/learning
  - Older data under cataloged, use ML for easier location
  - Understanding journal acceptance and rejection data
  - Improve collection analysis
- Challenges
  - Extracting information in a specific format
  - OCR everything
  - Licensing
  - Different platforms/discovery tools
  - Full inventory

### *Topic: Ethics, Ethical Considerations, and ML (including bias and algorithms as constructions of worldviews)*

#### Group 2

- IRB
  - Not keeping up with what we can do
  - Something was ok, and now it isn't
  - Internet data collection, observation vs. "figuring out people"
- Recommendations
  - Recommender cannot be an authority on what the students do/see/read
- Opportunity Cost
  - What are you missing by using the recommender?
- Define recommendations
  - What are we trying to recommend?
    - Interest?
    - Specific areas of study?
- Commercial
  - Data retention ethics
- What data is being kept?
  - Transparency in what is being held

- Example: Ring Doorbell
- Surveillance
- Safety vs. Functionality vs. Efficiency (self driving cars)
  - If self driving cars are safer do we have an obligation to use them?
- At what degree should or shouldn't we automate people out of a job?
  - What do we do with the people whose jobs are automated?

*Topic: Machine Learning for Automated Collection Metadata and Description*

#### Group 3

- Different ways we talk about data & what metadata we are measures
- Discussion on sizes of data (videos)
  - Not length, etc, but storage size required
- Newspaper collection level metadata vs. item level metadata
- Use ML for “aboutness” for topic modeling
  - Quality based on how specifically you can train models
- Use ML for record extraction
  - Contributes to the openness of ML

*Topic: How can we utilize technologies to aid in humanities research for those who are put off or extremely confused by algorithms?*

#### Group 4

- Hard sciences have leveraged open data and ML
- Combined methods
  - IE, Datasets and crowdsourcing
- Data science in the disciplines as a major supported effort
  - How can we benefit from this trend?
- ML for automating keyword descriptions
  - MSU assigns keywords to theses
    - Too many to do manually
- Bringing together digitized images
  - How can we open large amount of undescribed objects for research
  - When data that is not canonical
- 3D Objects
  - Repositories exist
    - ML collection sharing across institutions

*Topic: Using Machine Learning to Enhance Description or Automate Enhanced Description*

Group 4

- Notre Dame using text mining to find additional locations of where particular species of yeast may be present
  - Information about similar species may lead to more information about species of interest
  - Unique topic due to limited data points
  - Using open source linked data
- MSU looking at pieces for electron dissertation collection
  - A group of people are training model to start assigning subjects to dissertation uploads
    - Due to the volume of material workload is too great to do manually
  - Mostly about extraction at this point
  - Digital Scholarship lab group - recently opened and already busy
- Duke wants to add data science to every discipline (IE Humanities & Data Science, etc.)
  - Week long courses offered on machine learning
  - Data and visualizations department doing workshops throughout the semester to support courses and effort
  - Developing prerequisites for ML courses
  - Currently extra curricular
- Notre Dame (Will Martin) establishing visualization studio
  - A couple computers not easily accessible
  - Acquired 3D scanner
    - Pottery by Margaret Kelly Cable
    - Fossilized clams
  - Can we use ML for any of the 3D scan data?
    - Duke creating hyrax repository for 3D scans

## Palo Alto Workshop

*Monday, April 1, 2019*

*Location: 345 Hamilton Ave, Palo Alto, CA (AT&T building)*

### Agenda

- |       |   |                   |
|-------|---|-------------------|
| I.    | Lite Breakfast/Socialize  | 08:00am - 08:30am |
|       | A. Introduce team and logistics, agenda   |                   |
| II.   | Introduction: Objectives, Why We're Here  | 08:30am - 09:15am |
| III.  | Presentations   | 09:15am - 10:45am |
|       | A. "Appraising, Processing, and Providing Access to<br>Historical Email Archives with ePADD" Josh Schneider,<br>Assistant University Archivist, Stanford University<br>and Peter Chan, Digital Archivist, Stanford University |                   |
|       | B. "Untitled" Erin McCabe, Digital Scholarship Librarian,<br>University of Cincinnati   |                   |
|       | C. "Machine Learning + Digital Scholarship" Andrew<br>Janco, Digital Scholarship Librarian, Haverford College   |                   |
|       | D. "Cross-Disciplinary Research" Avery Hua, Product<br>Manager, Kyndi   |                   |
|       | E. (Should we include Abdul-Mageed?)  |                   |
| IV.   | Break   | 10:45am - 11:05am |
| V.    | Brainstorming Activity  | 11:05am - 12:00pm |
| VI.   | Lunch   | 12:00pm - 01:00pm |
| VII.  | Live Survey About Interest/Experience in Classification<br>Text mining, etc.  | 01:00pm - 01:15pm |
| VIII. | Cross Professional Conversations  | 01:15pm - 02:15pm |
|       | A. Groups organized for diversity mix (researchers,<br>computer scientists, librarians)   |                   |
|       | B. Survey Topics - #1   |                   |
|       | 1. How CS approaches  |                   |
|       | 2. How Librarian approaches   |                   |
|       | 3. How Researcher approaches  |                   |
| IX.   | Break   | 02:15pm - 02:35pm |
| X.    | Cross Professional Conversations  | 02:35pm - 03:35pm |
|       | A. Groups organized for diversity mix (researchers,<br>computer scientists, librarians)   |                   |
|       | B. Survey Topics - #2   |                   |
|       | 1. How CS approaches  |                   |
|       | 2. How Librarian approaches   |                   |
|       | 3. How Researcher approaches  |                   |
| XI.   | Report Out  | 03:35pm - 04:35pm |
| XII.  | Wrap Up - Conclusions, Next Steps   | 04:35pm - 05:00pm |

## Workshop Participants

<b>Name</b>	<b>Institution</b>
Andrew Janco	Haverford College
Chris Markman	City of Palo Alto
Muhammad Abdul-Mageed	University of British Columbia
Svein Arne Brygfeld	National Library of Norway
Dan Lou	City of Palo Alto
Heather Richards-Rissetto	University of Nebraska-Lincoln
Mike Taylor	Northern Arizona University
Erin E. McCabe	University of Cincinnati
Melissa Gill	The Getty
Josh Schneider	Stanford University
Joshua Hussey	University of Georgia
Peter Chan	Stanford University
Patrice-Andre Prud'homme	Oklahoma State University
Avery Hua	Kyndi
Marisa Plumb	San Jose State University
Catherine N Coleman	Stanford University
Tom Cramer	Stanford University
Eric Morgan	Notre Dame
John Wang	Notre Dame
Mark Dehmlow	Notre Dame
Nastia Guimaraes	Notre Dame

## Brainstorming Activity

COMMON TOOLS USED TO OVERCOME CROSS-DISCIPLINARY RESEARCH OBSTACLES	CROSS-DISCIPLINARY RESEARCH CHALLENGES	SUCCESSFUL STRATEGIES FOR CROSS-DISCIPLINARY RESEARCH	WHICH CROSS-DISCIPLINARY PROBLEMS CAN BE SOLVED? WHICH CANNOT?
<b>General Tools:</b>	<b>General Challenges:</b>	<b>General Strategies:</b>	<b>General Problem Solving:</b>
Complementary research agenda	High F&A rate	Problem setting	Can - knowledge wasted
	Black Box loyalists		Can - insufficient cross over
	Software sustainability		Cannot - Mode of scholarship
	Sustainability		
	Translation of RQ to something mappable to ML outputs		
	Mode of scholarship		
	Different scholarly outputs - single vs multiple author scholarship		
<b>Organizational Tools:</b>	<b>Organizational Challenges:</b>	<b>Organizational Strategies:</b>	<b>Organizational Problem Solving:</b>
Creation of suitable corpuses for training	Operation costs	Bringing faculty to the table - tools must make work easier, more productive time, UX design	Can - value in pursuing this effort
	Funding	Partnering outside the library	Can - high price tags
	Time	Ensuring robust inclusive project team	Can - justification of work, time, funding, personell
	Money	Commitment	
	Publishing expectations		
	Publish results		
	Reserach publication norms and expectations		
	Measures of success		

	differ		
	High price tag for CS but humanities grants low		
<b>Resource Tools:</b>	<b>Resource Challenges:</b>	<b>Resource Strategies:</b>	<b>Resource Problem Solving:</b>
		Incentives for starting and maintaining required collaboration (financial?, ?)	Can - humanists don't like failure
		Regular check-ins, passionate people	Can - jargon
<b>Collaboration Tools:</b>	<b>Collaboration Challenges:</b>	<b>Collaboration Strategies:</b>	<b>Collaboration Problem Solving:</b>
Uncontrolled vocabularies	Communication	Good team understanding + capacity for research question, data, tools/methods	Can - better communication across fields "common terms"
How might crowd-sourcing be used to assist with topic creation and/or controlled vocabulary construction?	Finding appropriate collaborators	Exchanging not only technical skills but histories of cultural analysis	Can - range of discipline = different opportunities standards for ACAD/RIGOR
Coffee	Speaking different Jargon	Successful story sharing	Can - range of discipline = different opportunities standards for ACAD/RIGOR
	Habits of mind	In person meetings	
	Common terms	Dialogue	
	AI and ML models are usually too narrow and very hard to be shared across domains/disciplines	Common goals	
	Domain specific terminology	Open lines of communication	
	Different objectives when building tools for research vs for scalable applications	Librarian/Tech lead "guided review" of results session	
	People with right skills		
	People - collaboration minds + contextualization		

Existing Data Tools:	Existing Data Challenges:	Existing Data Strategies:	Existing Data Problem Solving:
GitHub issues, projects	Use cases/research questions appropriate for tools	Open source tools	Can - search assistant
Wikidata	Provenance + context	ML model sharing	Can - find relevant text/images/objects based on a starting sample
Wikis	Meaning-making varies	(semi)Standardized data profiles as model exploring jumping-off point	Can - privacy is fairly standardized
Stanford NLP		Data sharing	Can - classify, sort, organize, parse, describe a corpus
Stanford NLP NLTK			Can - multimedia cross analysis
Google Vision			Cannot - Language
Slack, QGIS, Google Docs			Cannot - google etc. having useful but secret proprietary analysis tools
PY Torch (deep learning)			
Spark (big data)			
Langid (language identification)			
Jupyter Notebook, Binder			
Library search - DB aggregators			
Web-based tools - promote wider data use			
Where to find about new tools: NLP courses, Python for data science/applied ML course			
<b>Planning and Implementation Tools:</b>	<b>Planning and Implementation Strategies:</b>	Planning and Implementation Strategies:	<b>Planning and Implementation Problem Solving:</b>
	Determining whether a problem/area is best addressed through ML	Planning	



		Common understanding of why a tool was built and what it cannot do	
		User research	
		Identify research outcomes that have value to peers in both Arts and Sciences	

## Breakout Discussions

*Topic: How important is it for faculty to understand the math and statistical principles built into algorithms?*

### Group 2

- Ethics of algorithms
  - Need to understand the implications
- A cross-disciplinary group can strengthen research from a thoroughness standpoint, noticing deficiencies from different perspectives
- It is important to understand the limitations of the algorithms and working back from there

*Topic: How important is it for faculty to understand the math and statistical principles built into algorithms?*

### Group 3

- It's not critical to know the math, but have to have some responsibility for making sure it works properly
  - Some scholars are expected to know the math and the computer science as well

*Topic: How important is it for faculty to understand the math and statistical principles built into algorithms?*

### Group 1

- There is a threshold where researchers need to verify validity

*Topic: When is it better to train from scratch versus transfer learning?*

### Group 1

- If there are questions, ask about them in advance instead of just plugging in a model and hoping it works
- Would be nice to have environmental scan for models that are available to support one's research question
- Depends on your question, time, resources
- Awareness of some of the ethical questions
- Specificity of question is important too
- Transfer learning - training on a model that already exists, you can begin from an existing model and tune it to your needs, this will save a lot of work and allows for taking advantage of others' labor

*Topic: What different approaches/methods exist and how do we find out more about the potential utility of approaches for different research goals and data types?*

Group 3

- Don't be afraid to experiment
- Critical analysis
- Being aware for machine learning can do and what the results mean
- Libraries may be a bit intimidated by taking on the efforts - need to develop confidence
- Talked about cultural shift in libraries

*Topic: What will a library centered AI look like? Do we need a library centered AI? If so, how to build it?*

Group 1

- Focused on different cultural centers
- Services that libraries can offer such as project management
- Library means neutral territory
- Metadata generation and discoverability
- For public libraries, there are digital literacy concerns on how AI works

*Topic : What infrastructure should academic libraries provide to support machine learning?*

Group 2

- Consider what campus IT looks like and resources, that will inform what can do
- Keep in mind the curation of data
- Developing best practices and guidelines
- Building a network and direct people to the best resources on campus to help

## New York Workshop

*Thursday, April 25, 2019*

*Location: Columbia University, Butler Library, room 203*

### Agenda

- |       |  |                   |
|-------|--|-------------------|
| I.    | Lite Breakfast/Socialize   | 08:00am - 08:30am |
|       | A. Introduce team and logistics, agenda  |                   |
| II.   | Introduction: Objectives, Why We're Here   | 08:30am - 09:00am |
| III.  | Presentations  | 09:00am - 10:45am |
|       | A. "Facilitating Cross-Disciplinary Research" Paul Edlblut,<br>Vice President, Cognitive Computing, Vantage Solutions  |                   |
|       | B. "Studying Undergraduate Student Writing" Charlie Harper,<br>Digital Learning and Scholarship Librarian, Case<br>Western Reserve University                                  |                   |
|       | C. "DPLA & Data Engineering" Audrey Altman, Developer,<br>Digital Public Library of America  |                   |
|       | D. "Counting Giraffes" Michael Lesk, Professor of Library<br>And Information Science, Rutgers University   |                   |
|       | E. "Machine Learning and Analytics" Corey Harper,<br>Technology Research Directory, Elsevier   |                   |
|       | F. "The Freedom of Information Archive" Matthew Connelly,<br>Professor of History, Columbia University and Raymond<br>Hicks, Project Manager, History Lab, Columbia University |                   |
| IV.   | Break  | 10:45am - 11:05am |
| V.    | Brainstorming Activity   | 11:05am - 12:00pm |
| VI.   | Lunch  | 12:00pm - 01:00pm |
| VII.  | Live Survey About Interest/Experience in Classification<br>Text mining, etc.   | 01:00pm - 01:15pm |
| VIII. | Cross Professional Conversations   | 01:15pm - 02:15pm |
|       | A. Groups organized for diversity mix (researchers,<br>computer scientists, librarians)  |                   |
|       | B. Survey Topics - #1  |                   |
|       | 1. How CS approaches   |                   |
|       | 2. How Librarian approaches  |                   |
|       | 3. How Researcher approaches   |                   |
| IX.   | Break  | 02:15pm - 02:35pm |
| X.    | Cross Professional Conversations   | 02:35pm - 03:35pm |
|       | A. Groups organized for diversity mix (researchers,<br>computer scientists, librarians)  |                   |
|       | B. Survey Topics - #2  |                   |
|       | 1. How CS approaches   |                   |
|       | 2. How Librarian approaches  |                   |
|       | 3. How Researcher approaches   |                   |

- XI. Report Out  
XII. Wrap Up - Conclusions, Next Steps

03:35pm - 04:35pm  
04:35pm - 05:00pm

Workshop Participants

Name	Institution
Sarah Melton	Boston College
Melanie Wacker	Columbia University
Monica Maceli	Pratt Institute
Pamela Graham	Columbia University
Brian Luna Lucero	Columbia University
Violeta Ilik	Columbia University
Paul Edelblut	Vantage Solutions
Serenity Sutherland	SUNY Oswego
Achim Koh	Pratt Institute
Charlie Harper	Case Western Reserve University
Patrick Smyth	Columbia University
Timothy Ryan Mendenhall	Columbia University
Amanda Rust	Northeastern University
Suzanne Wones	Harvard University
Elisabeth Joyce	Edinboro University
Audrey Altman	Digital Public Library of America
Michael Lesk	Association for Computing Machinery
Corey A Harper	Elsevier
Janice Kim	MIT
Timothy Thompson	Yale University
Charlotte Nunes	Lafayette College
Alex Whelan	Columbia University
Raymond Hicks	Columbia University
Matthew Connelly	Columbia University
Christopher Gass	CUNY
John Wang	Notre Dame
Mark Dehmlow	Notre Dame
Eric Morgan	Notre Dame
Nastia Guimaraes	Notre Dame

## Brainstorming Activity

<b>COMMON TOOLS USED TO OVERCOME CROSS-DISCIPLINARY RESEARCH OBSTACLES</b>	<b>CROSS-DISCIPLINARY RESEARCH CHALLENGES</b>	<b>SUCCESSFUL STRATEGIES FOR CROSS-DISCIPLINARY RESEARCH</b>	<b>WHICH CROSS-DISCIPLINARY PROBLEMS CAN BE SOLVED? WHICH CANNOT?</b>
<b>General Tools:</b>	<b>General Challenges:</b>	<b>General Strategies:</b>	<b>General Problem Solving:</b>
Wishlist: Multidimensional tree-view browser of the "nodes" of concepts and their many dimensions in order to exp/extract and drill-down.	The curse of knowledge is why we communicate badly	Train documents of people who work together powerful if also negative ??? items	Physical Create (?) CAN be solved
	Information overload	(?) can motivate	Cross-disciplinary discovery may be easier to solve than cross-disciplinary research
			How to measure cross-disciplinary success - or - implementation rates? Should we even try?
			Promotion - as progress in decades (?)
<b>Organizational Tools:</b>	<b>Organizational Challenges:</b>	<b>Organizational Strategies:</b>	<b>Organizational Problem Solving:</b>
Granting structures in place to fund cross-disciplinary research	Competing priorities	Leadership support	Can - Funding CAN be solved
Multidisciplinary funding consortia	Different/competing systems of reward/recognition/credit	Organizational push/support, e.g., targeted internal grants	Cannot - vast differences in technical skills across disciplines (unless we change support and standards for tenure, promotion, etc.)
	University "silos"	Library as neutral partner for bringing different disciplines together	
	Funding & project leadership	Advertising library & research services and	

		improving our visibility	
	Equal credit for participants (one area may be perceived as support).	Demonstrating value through examples of successful ML, i.e., getting research buy-in	
	Culture/bias	Having people be physically co-located	
		Creating culture of stakeholders gathering, sharing disciplinary values & methods and hearing each other about why these roles & methods are in place	
<b>Resource Tools:</b>	<b>Resource Challenges:</b>	<b>Resource Strategies:</b>	<b>Resource Problem Solving:</b>
	Easy to learn one tool deeply, hard to get breadth of knowledge in a new field	Similar interests can motivate	
	trust - experts have to relinquish control - respect	Respecting each other's expertise	
	Lack of respect for expertise from other domains - confidential	Practical problems as motivation	
	ability to work, collaborate as teams	Skills at translating knowledge between domains (people skills, not AI)	
	Social Constructs - Incentives to collaborate even across Labs/Inst	Concept of "play", gamification for finding new connections	
<b>Collaboration Tools:</b>	<b>Collaboration Challenges:</b>	<b>Collaboration Strategies:</b>	<b>Collaboration Problem Solving:</b>
Controlled vocabularies with cross-references	Different subject knowledge; different skill sets	Forming teams with complementary skillsets	Can - Specific meanings in content-specific terms/jargon CAN be solved
In-person workshops - carve out time for sustained thinking	Different definitions of "research"	Diverse project team	
	Diverse specialist vocabularies	Find and use the different channels	

		through which different disciplines communicate	
	Discipline-specific vocabularies	Creating interdisciplinary ontologies	
	Communicating domain knowledge & problems (e.g. vocabulary)	Ontology mapping	
	Terminology	Centering on common desired outcome of project	
	Sense/Acronym Ambiguity	Form teams, advisory groups that intentionally mix people from different backgrounds	
	Different lingo	Working groups, interest groups, reading groups	
	Domain specific language	Cohesive project teams	
	Mismatched disciplinary values & expectations	Bring other domain experts into multi-disciplinary research	
	Lack of understanding of debates within other domains (e.g., why choose one .ai package over another?)	Sharing skillsets across domains	
	How to communicate & share data?		
	Dispersed information different practices in how information is generated		
	Different standards for precision/rigor in data collection & analysis		
	Different motivations & expected outputs		
	different values - quantitative vs qualitative		
	Different research outputs		
	Shared assumptions/perspectives within communities		



	Lack of participant diversity in most domains (some worse than others)		
Existing Data Tools:	<b>Existing Data Challenges:</b>	Existing Data Strategies:	Existing Data Problem Solving:
Wolfram Alpha.com for advanced searching	How to organize data? CVS, xml, linked data	Generalizable Tooling - Configurable	Can - Discovery CAN be solved
Bespoke taxonomies	Bias in ML training sets	Compliance requirements	Can - for more advanced search WolframAlpha.com is very relevant. Research & functions made available by the above tools has been brought into this context?
Browsing stacks as building content	"Text mining" visual language/data at scale		Cannot - Irony and context for text analysis
Aggregations, curated collections e.g. DPLA!	Generating/interpreting meaningful topics in fragmented/non-linear texts (i.e. Twitter hashtag conversations)		Cannot - Different data structures? It's solvable only to an extent
Tools w/prebuilt GUIs like Voyant, (anything easing dissemination?)	Semantics - lack of or imperfect overlap		Cannot - Issues of interpretation may be hard to solve
Linked Data tools	Contextualization at scale		Cannot - Research outputs CANNOT be solved
Shared datasets for benchmarking	Ontology matching		
Data dictionaries	Reconciling differing domain models		
Context sensitive vectors (BERT & ECM!)	Anthropomorphism (which can hinder understanding of technology)		
Google as a tool better than local search engine			
Google & ibm.com for business			
Textual analysis improved search, discovery, etc			
Pre-trained models &			

transfer learning			
Visualizations (Networks, Maps)			
Text & Data Analyses			
discovery			
large-scale databases			
Ubiquitous cloud storage/documents			
Planning and Implementation Tools:	<b>Planning and Implementation Challenges:</b>	Planning and Implementation Strategies:	Planning and Implementation Problem Solving:
	Quality metrics -- how are they define? What is a "successful" outcome?	Scope projects in a way that serves multiple audiences and needs. Everybody gets something.	
	Scoping the research project - Scale	Atomizing problems to lowest common domain	
	Finding all the best networks of experts for collaboration	Establishing clear project goals and labor equity	
	Communication sharing expertise	Good project planning where clear benefit to ALL partners	

## Breakout Discussions

[First topic group missing from recording]

*Topic: Library roles in data management for machine learning: data ethnography, bias detection, training data generation, and dataset description.*

### Group 1

- Key tensions.
  - Public vs Private. Related to ethical management and use of data.
  - Public/Private tensions causes difficulty for academic library: for some areas there is a huge skew between this, e.g. an expert in data science considering whether to work in for-profit corporation vs being a data science librarian. How can libraries adapt to handle (and recruit) this kind of expertise.
  - Speed vs Control. In a field that is developing and and seemingly infinite possibility, how to manage this things that require time to develop (e.g. taxonomies, governance, standards, review).
  - Context vs Lack-of-context. When working with aggregation (DPLA), how much do you treat as a body of data unto itself. Dumping data into a big pool without the context of where it came from can prevent us from really understanding what is there. What dangers are inherent in making data available without context?
  - Libraries are in a place to raise these kinds of questions (esp with information literacy) and in building bridges.

### Group 2

- Libraries are in a good position to bridge groups, but might not have the expertise to detect problems with data, e.g. bias.
- Cannot remove bias from data, so people will always need to be aware of its presence, and this can use training.

### Group 4

- Thinking of ways people in libraries could make it easier for others to do machine learning. one idea: training sets of data
- What if they made the culture of machine learning to be more open and sharing?
  - Question led to questions on infrastructure: how to organize and share such experiences? What processes could they put in place to make sharing and communication sustainable? How to make the technology more accessible?

*Topic: the importance of unstructured, playful serendipitous access to cross-discipline content to discover unexpected connections*

Remote Group

- whenever one brings things together, there are not just two things coming together, there is also the needs of all the individual people.
- learning a second language or way of understanding the world is much different than learning the first one. Need a place where one can practice and learn.

*Topic: The ethical considerations of using machine learning in libraries*

Remote Group

- in IT people are hyperconscious of ethics issues.
- Europe has "right-to-be-forgotten" law. But is it really feasible to be forgotten on the internet.
- We all provide information to commercial concerns, for free, without thinking about it
  - Is it appropriate to educate people in the commons about what they are doing and the risks they are doing with their data by engaging with companies

*Topic: To what degree can machine learning be used in libraries?*

Group 1

- using machine learning with library metadata and collection materials:
  - as a gap analysis
  - applying to archival materials to improve discovery
  - applying to data to look at reuse and what people are doing with our collections
- what is needed/how do we scale up as a profession
  - often exploratory work is not supported well, because of needs to hit certain production targets, and the payoff is not always clear to administration---a lot of exploratory work fails or suggests other work.
  - could there be a website that provides sample projects, datasets, or example steps that people could follow to try out and learn these skills?
- discussed roles of "data librarianship" and "data collections"? These new positions often have ambiguous/unclear responsibilities that vary widely between institutions
- How can we interface with vendors (particularly single-search interface vendors) to "go deeper" to adjust the ranking and fine-tuning of things. The bottleneck is often just with getting a good set of data to work with.

*Topic: What role do libraries and educators have in educating the broader public in ML and its societal implications?*

#### Group 2

- often on search engines, have no idea where something came from nor why it was shown to you. what can libraries do to help?
- how can it be easier to become a part of the google trusted publisher program.

*Topic: How can I apply ML to metadata creation?*

#### Group 4

- what kind of tasks would ML be good for for finding metadata in video?
  - could transcribe text/speech in videos
  - difficulties, people don't always speak standard dialects
  - smaller tasks: only language detection
- questions of labor:
  - who might be edged out by these developments?
  - will it contribute to a greater % of workforce non-participation?
  -

#### Group 3

- thought about using ML for handwriting analysis, and applied to linked data
- use ML for assigning metadata to AV materials
- linked data: library fixation on linked data, whether we are doing it right or not, and the skepticism those in library leadership have on linked data solving problems.
  - observed that there are conferences in other disciplines about using linked data, e.g. financial services and credit card companies on knowledge graphs. why difference between the conception and success commercial
    - (opinion) library was too stuck on RDF, ontologies and inference, and not observing that there is a history of using graphs and computing over graphs.
    - (opinion) the "open world assumption" in linked open data is REALLY hard, and makes reasoning difficult. commercial approaches do not make this assumption and do not need to deal with it.

*Topic: How should machine learning be taught in higher education?*

#### Group 5

- how can non-specialists participate in machine learning?
- ethics and teaching the consumer is an important topic. Where are these ethics being taught? Appear like ethics is limited to a single course.
- topics in a potential ML course: python, presentation layer, solr, open vs non-open source, statistics, how to identify gaps in data
- process of doing ML is really iterative.
- need to teach people about PII and GDPR, how to anonymize data, how long to keep things
- important to bring in actual participants, not just theory, when teaching ML. e.g. people in hospitals practice ML.

## Washington D.C. Workshop

Friday, May 31, 2019

Location: Notre Dame Keough School D.C. Office, 1400 16th St NW, Washington, D.C.

### Agenda

- |       |  |                   |
|-------|--|-------------------|
| I.    | Lite Breakfast/Socialize   | 8:00am - 8:30am   |
|       | A. Introduce team and logistics, agenda  |                   |
| II.   | Introduction: Objectives, Why We're Here   | 8:30am - 9:00am   |
| III.  | Presentations  | 9:00am - 10:45am  |
|       | A. "AMP: The Audiovisual Metadata Platform" Jon Dunn,<br>Assistant Dean for Library Technologies, Indiana<br>University, Bloomington                               |                   |
|       | B. "Project Management in Yale's Digital Humanities<br>Lab" Douglas Duhaime, Digital Humanities Software<br>Developer, Yale University                             |                   |
|       | C. "Policy Analytics" Patrick McLaughlin, Director of<br>Policy Analytics and Senior Research Fellow, George<br>Mason University                                   |                   |
|       | D. "Reading Chicago Reading Project" Ana Lučić,<br>Digital Scholarship Librarian, DePaul University  |                   |
|       | E. "Machine Learning-Based Metadata Generation for<br>Library Archives" Harish Maringanti, Associate Dean<br>for IT & Digital Library Services, University of Utah |                   |
|       | F. "Machine intelligence and Moral Decision Making"<br>Bohyun Kim, CTO and Associate Professor, University<br>of Rhode Island                                      |                   |
| IV.   | Break  | 10:45am - 11:05am |
| V.    | Brainstorming Activity   | 11:05am - Noon    |
| VI.   | Lunch  | Noon - 1:00pm     |
| VII.  | Live Survey About Interest/Experience in Classification<br>Text mining, etc.   | 1:00pm - 1:15pm   |
| VIII. | Cross Professional Conversations   | 1:15pm - 2:15pm   |
|       | A. Groups organized for diversity mix (researchers,<br>computer scientists, librarians)  |                   |
|       | B. Survey Topics - #1  |                   |
|       | 1. How CS approaches   |                   |
|       | 2. How Librarian approaches  |                   |
|       | 3. How Researcher approaches   |                   |
| IX.   | Break  | 2:15pm - 2:35pm   |
| X.    | Cross Professional Conversations   | 2:35pm - 3:35pm   |
|       | A. Groups organized for diversity mix (researchers,<br>computer scientists, librarians)  |                   |

B. Survey Topics - #2

1. How CS approaches
2. How Librarian approaches
3. How Researcher approaches

XI.	Report Out	3:35pm - 4:35pm
XII.	Wrap Up - Conclusions, Next Steps	4:35pm - 5:00pm

Workshop Participants

Name	Institution
Abigail Potter	Library of Congress
Ana Lucic	DePaul University
Athina Livanos-Propst	PBS
Bohyun Kim	University of Rhode Island
Chris Iweha	Morgan State University
Clifford Lynch	Coalition for Networked Information
Doug Duhaime	Yale University
Erin Plettenberg	Federal Reserve Board
Harish Maringanti	University of Utah
Jason Cohen	Berea College
Jennifer Roper	University of Virginia
John Rees	National Library of Medicine
Jon Dunn	Indiana University
Kevin Gunn	Catholic University of America
Megan Potterbusch	George Washington University
Meghan Ferriter	Library of Congress
Michael Levy	United States Holocaust Memorial Museum
Michael Poston	Folger Shakespeare Library
Patrick McLaughlin	George Mason University
Ryan Wolfslayer	Federal Reserve Board
Tim McGeary	Duke University
Eric Morgan	Notre Dame
John Wang	Notre Dame
Laurie McGowan	Notre Dame
Mark Dehmlow	Notre Dame

## Brainstorming Activity

COMMON TOOLS USED TO OVERCOME CROSS-DISCIPLINARY RESEARCH OBSTACLES	CROSS-DISCIPLINARY RESEARCH CHALLENGES	SUCCESSFUL STRATEGIES FOR CROSS-DISCIPLINARY RESEARCH	WHICH CROSS-DISCIPLINARY PROBLEMS CAN BE SOLVED? WHICH CANNOT?
<b>General Tools:</b>	<b>General Challenges:</b>	<b>General Strategies:</b>	<b>General Problem Solving:</b>
coursera and Andrew Ng	traditionally the narrower the domain the better the model	will AI succeed because humans become more like computers	certain problems can be solved with more data made available
can focus on search too, not just metadata creation	understanding data does not solve or improve all problems	Run! Faster!	automate process requiring some human judgement
need ML driven services that can be easily used or "fine-tuned" with custom training data then used (examples, facial recognition, handwriting ocr, speech to text)	working with legacy systems and workflows	look at the research done in area outside of one's own specialty	Tricky to solve: monetization, incentive alignment, naming things, cross browser compatibility
	lack of OTS software	cold call an expert	Discipline specific access points allow more depth, but may decrease access from other disciplines.
	lack of understanding about how ML/AI works at a system level and how it may apply to one's field		
<b>Organizational Tools:</b>	<b>Organizational Challenges:</b>	<b>Organizational Strategies:</b>	<b>Organizational Problem Solving:</b>
library tradition of facilitating access to help developers find training data, code	balancing the research/teaching tool vs. admin priorities, funding, etc.	we need better understanding of what its important to know about other disciplines	Can - prioritizing staff/human resource commitment
need more good training sets of all kinds	lack of human resource capacity/time to experiment vs. library/university	outsourcing, hiring talent	Cannot - bias in funding distribution



	priorities		
pre-packaged/easily trainable ml models make ml accessible beyond computer scientist	Availability of funding opportunities that support all disciplines involved in a project, rather than focusing on one	design spaces for spontaneous interactions	
	funding, seed/pilot, project, sustained practice	recruit a dean to the cause	
	funding new positions	institutional support is generally more successful	
	expertise on campus vs. out of the box/ industry tool functions	create incentives such as grants (ex collections as data)	
	expert knowledge no cross disciplinary subject		
	resources for involving diversity of institutions and disciplines for a broad set of data		
	getting buy in from leadership		
<b>Resource Tools:</b>	<b>Resource Challenges:</b>	<b>Resource Strategies:</b>	<b>Resource Problem Solving:</b>
demonstrate appreciation acknowledge of the perspectives of the groups brought in	reeducating a library workforce	people need to be willing to invest time learning about other disciplines.	
	making library staff unafraid of the future	workshops	
	strong library history of hands-on, item-by-item approach	demonstrations	
	anxiety about being displaced	share examples of cross disciplinary research to raise awareness	
	one group thinks their idea/ perspective is the most important	active listening training	
	not seeing what is possible from collaborative research	training communities of practice	

		provide incentives for cross disciplinary research	
<b>Collaboration Tools:</b>	<b>Collaboration Challenges:</b>	<b>Collaboration Strategies:</b>	<b>Collaboration Problem Solving:</b>
team agreements	jargon	work together to set team expectations/working agreement	Can - cross domain tooling reuse
real time/near real time communications platform eg slack	different vocabularies	writing down assumptions	Cannot - The vocabulary choices will never be perfect but they can be enough.
development of a glossary for a project - help all understand terms and language	terminology	actively solicit projects that involve multiple groups	
	communication with different vocabulary	actively seek partners that have not been involved to date	
	Language barriers, jargon, notation	adequate time for project planning/understanding each other's approaches and goals	
	domain specific terms and jargon	transferring techniques across disciplines/data types	
	various departments do not regularly know each other's jargon		
	compatibility of data from different corpus		
	which disciplines can collaborate		
	success looks different in different disciplines (eg journal articles vs. conference proceedings vs. book vs. code)		
	different end goals		
	different domain languages and assumptions		
	different methodologies		

	different standards for precision/rigor in data collection and analysis		
	dispersed information practices in how information is generated		
<b>Existing Data Tools:</b>	<b>Existing Data Challenges:</b>	<b>Existing Data Strategies:</b>	<b>Existing Data Problem Solving:</b>
Annotated data	contextualization at scale	data integration	Can - Adopting fair data principles/ defining them
kera5, tensorflow, webgl, math math math, serverless	algorithmic indexing for discovery will have false positives but will also miss things - how to measure the amount of "miss" error and inform users	open data	Cannot - bias toward history because of copyright
cloud based storage	existing bias in data	domain specific share training data (ie historic newspapers)	Cannot - Finding/redacting PII in unstructured data
python libraries	Training data hard to find imbalances, needs formatting	developing good and FAIR research data management practices for new research	Cannot - privacy for the context of of indexing discovery is a very hard problem
highly curated info gives AI a big leg up	availability of machine readable data		
	lack of metadata standard requirement		
<b>Planning and Implementation Tools:</b>	<b>Planning and Implementation Challenges:</b>	<b>Planning and Implementation Strategies:</b>	<b>Planning and Implementation Problem Solving:</b>
project request, review, charter, workflow	identifying a problem	benchmark successful projects that address similar challenges outside of libraries	
humans articulate the "end user" of a corpus and how a feature related to end user	choosing appropriate tools	balanced teams (min 2 shared perspectives)	
	who are our users and what do they want	regular project updates for transparency and scope checking	
	project scope and duration	clear chains of custody of data and projects	

## Breakout Discussions

### *Topic: Personnel question:*

#### Group 2

- Who should be on the team? Team should be subject expert, statistician, computer programmer.
  - Could you remove need for statistician? Compare early cars. You'd need to understand how the car works, but now there's a better infrastructure. So could a ML type group be similar?
  - Another analogy: a browser. GUI mitigates the need to use a command line.
  - So, how much of stats do you need?
- Lost opportunities if we don't move in a cross-disciplinary direction.
  - Every time we (manually) do something that can be automated, we lose the opportunity to do something that can't.
  - Freeing up backlogs.
  - User expectations. As users become more familiar with tools outside of the library, they're going to expect the same from the library.

### *Topic: Automatic Tagging question:*

- Topic modeling versus discrete class assignment such as classification.
- 
- Tools: Sci-kit Learn, TM in R, KARES (wrapper for Tensorflow)

### *Topic: Personnel question:*

#### Group 3

- Why some things seem to work or not.
- Who should be on the team?
  - SMEs are going to be vital. As are project managers. Also need a cheerleader, not just for the team, but for the larger organization, to keep it excited.
  - 
  - Also [like Group 1] debated need for statistician/data analyst. What's really needed is something who can communicate out the results very well.
  -
- Fighting the desire to hold a project close to the chest. Success of the project needs to be inherently owned by the team, not the individual.
- Work of data cleaning/formatting is vital, not something you can just highlight. These are the incredibly important jobs, not the "little guys."

#### Group1

- Who should be on the team?
  - People who drive the process (such as leaders/admins), people who've been around for a long time (even skeptics/resisters), people who create the analysis, domain experts, people who are ok with uncertainty.

#### *Topic: Automatic Tagging and Named entity questions combined:*

- What kind of collections might get the highest benefit from this kind of work? Web archives, oral histories, newspapers, music, anything else without item-level metadata.
- Novel uses perhaps in images and machine learning.

#### *Topic: Named entity question:*

#### Group 4

- Should we build our own model or use someone else?
  - Spacy, Apache NLP, Stanford NLP, Natural Language Toolkit
- Create a pipeline approach – many small separate processes
- Can use NER for aboutness
- Using NER to create a semantic network and network graph

#### *Topic: Automatic Tagging question:*

- Using topic modeling as a model to tag new documents.
- Couldn't think how widespread auto tagging processes are in libraries

#### *Topic: Projects (?) Question:*

- Machine learning projects get harder to work with, with more specific datasets.
- Not about making the machine integral to the process, but finding what's integral to being a librarian, and letting the machine do all the rest. Librarian then freed up to do other things.

#### *Other Comments:*

- Developing internal staff skills versus contracting out.
- Library as hub, not spoke in a wheel, of machine learning activity.

## Writers' Workshop

October 25, 2019

Location: University of Notre Dame, South Bend, IN, Hesburgh Library, Rooms 246 and 247

### Agenda

***NOTE: Please bring your laptop with you! You will be sharing ideas about your proposed chapter and providing feedback to other authors selected for the collection of essays on Machine Learning.***

8:40 AM	Depart Inn at Saint Mary's hotel for Hesburgh Library
8:50 AM - 9:00 AM	Check-in for the workshop
9:00 AM - 10:00 AM	Welcome and 2-3 min introductions of proposed topics
10:00 AM - 10:15 AM	Morning break*
10:15 AM - 11:30 AM	Sharing ideas in small groups (pre-selected groups)
11:30 AM - 12:00 PM	Setting personal goals for the afternoon & accountability partners
12:00 PM - 1:30 PM	Lunch and optional tour of ND campus
1:30 PM - 3:30 PM	Individual and small group work time on essays
3:30 PM - 4:00 PM	Check in with accountability partners
4:00 PM - 4:30 PM	Wrap-up and Next Steps
4:45 PM	Depart Hesburgh Library for the hotel
6:30 PM	Dinner at a nearby restaurant (optional, but all are invited!)

## Workshop Participants

<b>Name</b>	<b>Institution</b>
Andrew Janco	Haverford College
Samuel Hansen	University of Michigan
Jason Cohen	Berea College
Mario Nakazawa	Berea College
Charlie Harper	Case Western Reserve University
Meng Jiang	University of Notre Dame
Bohyun Kim	University of Rhode Island
Ana Lucic	University of Illinois at Urbana-Champaign
Marisa Plumb	San Jose State University
Arend Hintze	Michigan State University
Jorden Schossau	Michigan State University
Audrey Altman	Digital Public Library of America
Sue Wiegand	St. Mary's College
Eric Morgan	University of Notre Dame
John Wang	University of Notre Dame
Daniel Johnson	University of Notre Dame
Mark Dehmlow	University of Notre Dame
Alex Papson	University of Notre Dame
Melissa Harden	University of Notre Dame
Don Brower	University of Notre Dame
Rebecca Leneway	University of Notre Dame
Nastia Guimaraes	University of Notre Dame
<b>VIRTUAL PARTICIPANTS</b>	
Michael Lesk	Association for Computing Machinery
Patrice-Andre Prud'homme	Oklahoma State University