Craig A. Willis

CONTACT INFORMATION

School of Information Sciences
University of Illinois at Urbana-Champaign
Champaign, IL 61820 USA

willis8@illinois.edu
`https://craig-willis.github.io`

RESEARCH INTERESTS

Computational reproducibility; information storage and retrieval; research data and access; scientific metadata; scholarly communications

EDUCATION

University of Illinois at Urbana-Champaign, Ph.D, Library and Information Science, 2020
University of North Carolina at Chapel Hill, M.S. Library Science, 2012
University of Colorado at Boulder, B.A. Geography, 2007

PROFESSIONAL EXPERIENCE

School of Information Sciences, University of Illinois at Urbana-Champaign

Research Programmer	**2020 -**

National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

Senior Research Programmer	**2017 - 2020**
Research Programmer	**2015 - 2017**

ProQuest, LLC, Seattle, Washington

Lead Software Developer, Discovery Products Group	**2005 - 2010**

Sony Recording Media, Boulder, Colorado

Senior Software Engineer, eMedia Division	**2001 - 2005**

ACADEMIC EXPERIENCE

School of Information Sciences, University of Illinois at Urbana-Champaign

Co-Instructor Information Storage and Retrieval (LIS 456)	**Spring 2016**
Research Assistant	**2013 - 2015**

III: Improving Information Retrieval by Analysis of Temporal Evidence in a Unified Model (NSF 1217279, PI: M Efron).
IMIRSEL: International Music Information Retrieval Systems Evaluation Laboratory (PI: J. Stephen Downie)
HTRC: HathiTrust Research Center (PI: J. Stephen Downie)

Metadata Research Center, University of North Carolina at Chapel Hill

Research Assistant	**2010 - 2012**

HIVE: Helping Interdiscplinary Vocabulary Engineering (IMLS, PI: Jane Greenberg).

Funding

Publications

Dissertation

Willis, C. (2020). *Trust, but verify: An investigation of methods of verification and dissemination of computational research artifacts for transparency and reproducibility.* University of Illinois at Urbana-Champaign

Peer Reviewed Articles

Chard, K., Gaffney, N., Hategan, M., Kowalik, K., Ludaescher, B., McPhillips, T., Nabrzyski, J., Stodden, V., Taylor, I., Thelen, T., Turk, M. J., and **Willis, C** (2020). Toward enabling reproducibility for data-intensive research using the Whole Tale platform. *CoRR*

Yang, B., Kalyanam, R., **Willis, C.**, Lambert, M., and Kirkpatrick, C. (2019). CHEESE: Cyber Human Ecosystem of Engaged Security Education. In *Proceedings of the 20th Annual SIG Conference on Information Technology Education*, page 189–190

Chard, K., Gaffney, N., Jones, M. B., Kowalik, K., Ludäscher, B., McPhillips, T., Nabrzyski, J., Stodden, V., Taylor, I., Thelen, T., Turk, M. J., and **Willis, Craig** (2019a). Application of BagIt-Serialized Research Object Bundles for Packaging and Re-Execution of Computational Analyses. In *2019 15th International Conference on eScience (eScience)*, page 514–521. IEEE

Chard, K., Gaffney, N., Jones, M. B., Kowalik, K., Ludäscher, B., Nabrzyski, J., Stodden, V., Taylor, I., Turk, M. J., and **Willis, C.** (2019b). Implementing computational reproducibility in the Whole Tale environment. In *Proceedings of the 2nd International Workshop on Practical Reproducible Evaluation of Computer Systems*, page 17–22

LeBauer, D. and **Willis, C.** (2019). Vocabularies, APIs, and Formats for High Throughput Crop Phenotyping: The TERRA Ref Case Study. In *Plant and Animal Genome XXVII Conference (January 12-16, 2019)*. PAG

Mecum, B., Jones, M. B., Vieglais, D., and **Willis, C.** (2018a). Preserving reproducibility: Provenance and executable containers in dataone data packages. In *2018 IEEE 14th International Conference on e-Science (e-Science)*, page 45–49. IEEE

Mecum, B., Wyngaard, S., **Willis, C.**, Turk, M., Thelen, T., Taylor, I., Stodden, V., Perez, D., Nabrzyski, J., Ludaescher, B., and et al. (2018b). Science, containerized: Integrating provenance and compute environments with the Whole Tale. *AGUFM*, 2018:IN53A–02

Burnette, M., Kooper, R., Maloney, J., Rohde, G. S., Terstriep, J. A., **Willis, C.**, Fahlgren, N., Mockler, T., Newcomb, M., Sagan, V., and et al. (2018). TERRA-REF data processing infrastructure. In *Proceedings of the Practice and Experience on Advanced Research Computing*, page 1–7

McPhillips, T., **Willis, C.**, Gryk, M. R., Nuñez-Corrales, S., and Ludäscher, B. (2019). Reproducibility by Other Means: Transparent Research Objects. In *2019 15th International Conference on eScience (eScience)*, page 502–509. IEEE

LeBauer, D., Kooper, R., Burnette, M., and **Willis, C.** (2017). TERRA REF: Advancing phenomics with high resolution, open access sensor and genomics data. *AGUFM*, 2017:B42A–02

Willis, C., Lambert, M., McHenry, K., and Kirkpatrick, C. (2017). Container-based analysis environments for low-barrier access to research data. In *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*, page 1–4

Willis, C., Sherman, G., and Efron, M. (2016a). What makes a query temporally sensitive? *Proceedings of the Association for Information Science and Technology*, 53(1):1–9

Willis, C., Sherman, G., and Efron, M. (2016b). What Makes a Query Temporally Sensitive? In *9th International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 1065–1068. ACM

Choi, K., Lee, J. H., **Willis, C.**, and Downie, J. S. (2015). Topic Modeling Users' Interpretations of Songs to Inform Subject Access in Music Digital Libraries. In *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, page 183–186

Efron, M., **Willis, C.**, and Sherman, G. (2014). Learning sufficient queries for entity filtering. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in information retrieval*, page 1091–1094

White, H., **Willis, C.**, and Greenberg, J. (2014). HIVEing: the effect of a semantic web technology on inter-indexer consistency. *Journal of documentation*

Fenlon, K., Senseney, M., Green, H., Bhattacharyya, S., **Willis, C.**, and Downie, J. S. (2014b). Scholar-built collections: A study of user requirements for research in large-scale digital libraries. *Proceedings of the American Society for Information Science and Technology*, 51(1):1–10

Fenlon, K., Cole, T. W., Han, M.-J., **Willis, C.**, and Fallaw, C. (2014a). Rethinking HathiTrust Metadata to Support Workset Creation for Scholarly Analysis. In *DH*

Green, H. E., Fenlon, K. S., Senseney, M., Bhattacharyya, S., **Willis, C.**, Organisciak, P., Downie, J. S., Cole, T., and Plale, B. (2014). Using Collections and Worksets in Large-Scale Corpora: Preliminary Findings from the Workset Creation for Scholarly Analysis Project. *iConference 2014 Proceedings*

Willis, C. and Efron, M. (2013). Finding information in books: Characteristics of full-text searches in a collection of 10 million books. *Proceedings of the American Society for Information Science and Technology*, 50(1):1–10

Willis, C. and Losee, R. M. (2013). A random walk on an ontology: Using thesaurus structure for automatic subject indexing. *Journal of the American Society for Information Science and Technology*, 64(7):1330–1344

Willis, C., Greenberg, J., and White, H. (2012a). Analysis and synthesis of metadata goals for scientific data. *Journal of the American Society for Information Science and Technology*, 63(8):1505–1520

White, H., **Willis, C.**, and Greenberg, J. (2012). *The HIVE impact: contributing to consistency via automatic indexing*, page 582–584

Greenberg, J., Losee, R., Agüera, J. R. P., Scherle, R., White, H., and **Willis, C.** (2011). HIVE: Helping interdisciplinary vocabulary engineering. *Bulletin of the American Society for Information Science and Technology*, 37(4):23–26

OTHER PUBLICATIONS

Nüst, D., Eddelbuettel, D., Bennett, D., Cannoodt, R., Clark, D., Daroczi, G., Edmondson, M., Fay, C., Hughes, E., Lopp, S., and et al. (2020). The Rockerverse: Packages and Applications for Containerization with R. *arXiv preprint arXiv:2001.10641*

Kowalczyk, S. T., Sun, Y., Peng, Z., Plale, B., Todd, A., Auvil, L., **Willis, C.**, Zeng, J., Pathirage, M., Liyanage, S., and et al. (2016). *Big data at scale for digital humanities: An architecture for the HathiTrust Research Center*, page 345–369. IGI Global

Sherman, G., Efron, M., and **Willis, C.** (2014). *The University of Illinois' Graduate School of Library and Information Science at TREC 2014*

Willis, C., Medlin, R., and Arguello, J. (2012b). *Incorporating Temporal Information in Microblog Retrieval*

Projects

Merging Science and Cyberinfrastructure Pathways: The Whole Tale (NSF Award 1541450) (October 2017 - current)

CHEESE: Cyber Human Ecosystem of Engaged Security Education (NSF Award 1820608) (Sept 2018 - May 2020)

Transportation Energy Resources from Renewable Agriculture Phenotyping Reference Platform (TERRA-REF) (DOE) (Nov 2016 - June 2019)

Crops *in silico* (http://cropsinsilico.org/) (May 2018 - June 2019)

National Data Service (https://nationaldataservice.org) (January 2015 - August 2019)