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

Teaching and Research Interests

Computational reproducibility and transparency; peer review; 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

Experience

School of Information Sciences, University of Illinois at Urbana-Champaign
Teaching Assistant Professor **2023 -**

School of Information Sciences, University of Illinois at Urbana-Champaign
Research Programmer **2020 - 2023**

National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign
Senior Research Programmer **2017 - 2020**
Research Programmer **2015 - 2017**

School of Information Sciences, University of Illinois at Urbana-Champaign
Research Assistant **2013 - 2015**

Metadata Research Center, University of North Carolina at Chapel Hill
Research Assistant **2010 - 2012**

ProQuest, LLC, Seattle, Washington
Lead Software Developer, Discovery Products Group **2005 - 2010**

Training and Outreach

Publishing Transparent and Reproducible Computational Research with Whole Tale. CSDMS Annual Meeting, May 19th, 2022. Co-Presenter.

Developing, Packaging, and Sharing Reproducible Research Objects: The Whole Tale Approach. DataONE Webinar, October 8th, 2019. Co-Presenter.

The Whole Tale: Merging Science and Cyberinfrastructure Pathways. University of Washington eScience Institute, June 10th, 2019. Presenter.

Workshop on Education and Training for Reproducible Research (2019). National Center for Supercomputing Applications, March 18th, 2019. Co-Organizer.

Container applications in research computing and research data access. Practice and Experience in Advanced Research Computing (2018). Panel moderator.

USDA/NIFA DataDrivenAg Workshop and Hackathon. January, 2018. Infrastructure and support.

Program for Interdisciplinary and Industrial Internships at Illinois (PI4) Bootcamp. June, 2017. Infrastructure and support.

ThinkChicago Civic Tech Challenge. November, 2017. Infrastructure and support.

Einstein Toolkit School. 2017. Infrastructure and support.

National Data Service Workshop on Containerized Analysis Environments (2017). Co-Organizer.

Grants and Awards

Collaborative Research: Elements: TRAnsparency CErtified (TRACE): Trusting Computational Research Without Repeating It. NSF Award 2209628. June 2022 – July 2025. Co-Principle Investigator ($349,999)

Collaborative Research: CHEESE: Cyber Human Ecosystem of Engaged Security Education. NSF Award 1820608. July 2018 – June 2021. Principle Investigator ($149,917)

Feedback-based Expansion Models for Data Search. Subcontract award from NIH BioCADDIE project. May – July 2017. Subawardee ($79,300)

Projects

Collaborative Research: SI2-SSI: Inquiry-Focused Volumetric Data Analysis Across Scientific Domains: Sustaining and Expanding the yt Community (NSF 1663914, Jan 2023 - current)

High-throughput Materials Discovery for Extreme Conditions (DOD, Jan 2023 - current)

Merging Science and Cyberinfrastructure Pathways: The Whole Tale (NSF 1541459, Oct 2017 - current)

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

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

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

National Data Service (https://nationaldataservice.org) (Jan 2015 - Aug 2019)

III: Improving Information Retrieval by Analysis of Temporal Evidence in a Unified Model (NSF 1217279, 2014 - 2015)

IMIRSEL: International Music Information Retrieval Systems Evaluation Laboratory (2012 - 2014)

HIVE: Helping Interdiscplinary Vocabulary Engineering (IMLS, 2010 - 2012)

Software

Whole Tale (https://github.com/whole-tale/): Platform for publishing transparent and reproducible computational research.

NDS Labs Workbench (https://github.com/nds-org/ndslabs): Platform to provide low-barrier computational access to research data.

CHEESE Hub (https://github.com/cheese-hub/): Learning platform for network security, secure programming, and cryptography.

TERRA-REF (https://docs.terraref.org/): Data products, computing and analysis pipeline for large-scale agricultural phenotyping project.

ICASA Ontology (https://github.com/craig-willis/icasa): Prototype ontology for agricultural research.

Publications

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

McPhillips, T. M., Thelen, T., **Willis, C.**, Kowalik, K., Jones, M. B., and Ludäscher, B. (2021). CPR-A Comprehensible Provenance Record for Verification Workflows in Whole Tale. In Glavic, B., Braganholo, V., and Koop, D., editors, *Provenance and Annotation of Data and Processes*, Lecture Notes in Computer Science, page 263–269. Springer International Publishing

Willis, C. and Stodden, V. (2020). Trust but Verify: How to Leverage Policies, Workflows, and Infrastructure to Ensure Computational Reproducibility in Publication. *Harvard Data Science Review*, 2(4)

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, C.** (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. (2016). 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. (2014). 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

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