

Curriculum Vitae

Kathryn Knight

April 2024

1 Publications

Akidau, T., Begoli, E., Chernyak, S., Hueske, F., **Knight, K.**, Knowles, K., Mills, D., and Sotolongo, D. (2021). Watermarks in stream processing systems: semantics and comparative analysis of Apache Flink and Google cloud dataflow. *Proceedings of VLDB Endowment*. 14(12), 3135–3147. <https://doi.org/10.14778/3476311.3476389>

Begoli, E., Akidau, T., Hueske, F., Hyde, J., **Knight, K.**, and Knowles, K. (2019). One sql to rule them all-an efficient and syntactically idiomatic approach to management of streams and tables. *Proceedings of the 2019 International Conference on Management of Data (SIGMOD '19)*. Association for Computing Machinery, New York, NY, USA, 1757–1772. <https://doi.org/10.1145/3299869.3314040>

Begoli, E., Bridges, R. A., Oesch, S., and **Knight, K.** (2021). What clinical trials can teach us about the development of more resilient ai for cybersecurity. *European Cybersecurity Journal*. 7(1), 30-42.

Begoli, E., Goethert, I., and **Knight, K.** (2021). A lakehouse architecture for the management and analysis of heterogeneous data for biomedical research and mega-biobanks. *2021 IEEE International Conference on Big Data (Big Data)*. Orlando, FL, USA, (pp. 4643-4651). <https://doi.org/10.1109/BigData52589.2021.9671534>

Begoli, E., **Knight, K.**, Redmon, D., Rice, A., and Varma, L. (2021). Hospice landscape report. (publication number ORNL/SPR-2021/2271), Oak Ridge National Lab, Oak Ridge, TN (United States). <https://doi.org/10.2172/1830119>

Ferreira da Silva, R., Wilkinson, S., and **Knight, K.** (2022). The jot and tittle of workflows interoperability: Towards fair computational workflows via metadata apis. *Proceedings of the ASCR Workshop on the Management and Storage of Scientific Data*. <https://doi.org/10.2172/1843500>

Goethert, I., **Knight, K.**, and Linares, F. (2023). An approach to data management planning for protected data projects. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States).

Knight, K. and Boehm, S. (2022). Towards an international portal of ontologies and metadata standards for science. *Proceedings of the ASCR Workshop on the Management and Storage of Scientific Data*. <https://doi.org/10.2172/1843500>

Knight, K. and Wilkinson, S. R. (2022). Understanding the ai in fair data management support for ai and complex workflows. *Proceedings of the ASCR Workshop on the Management and Storage of Scientific Data*. <https://doi.org/10.2172/1843500>

Knight, K., Kurte, K., and Anantharaj, V. (2022). Toward a machine-actionable future for the doe science mission. *Proceedings of the ASCR Workshop on the Management and Storage of Scientific Data*. <https://doi.org/10.2172/1843500>

Knight, K., Honerlaw, J., Danciu, I., Linares, F., Ho, Y.-L., Gagnon, D. R., Rush, E., Gaziano, J. M., Begoli, E., Cho, K., et al. (2020). Standardized architecture for a mega-biobank phenomic library: the million veteran program (mvp). *AMIA Summits on Translational Science Proceedings*, 2020, 326–334.

Mahbub, M., Dams, G. M., Srinivasan, S., Rizy, C., Danciu, I., Trafton, J., and **Knight, K.** (2024). Leveraging large language models to extract information on substance use disorder severity from clinical notes: A zero-shot learning approach. arXiv preprint arXiv:2403.12297.

Mahbub, M., Goethert, I., Danciu, I., **Knight, K.**, Srinivasan, S., Tamang, S., Rozenberg-Ben-Dror, K., Solares, H., Martins, S., Begoli, E., et al. (2024). Question-answering system extracts information on injection drug use from clinical progress notes. *Nature Communications Medicine*, 4(61). <https://doi.org/10.1038/s43856-024-00470-6>

Rodgers, D. and **Knight, K.** (2022). The challenge of capturing and converting primary to secondary and summary datasets. *Proceedings of the ASCR Workshop on the Management and Storage of Scientific Data*. <https://doi.org/10.2172/1843500>

Rush, E., Ozmen, O., **Knight, K.**, Park, B., Baker, C., Jones, M., Ward, M., and Nebeker, J. (2020). Characterizing sub-cohorts via data normalization and representation learning. *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, USA, 173–176. <https://doi.org/10.1109/CBMS49503.2020.00040>

Srinivasan, S., Begoli, E., Mahbub, M., and **Knight, K.** (2021). Nomen est omen-the role of signatures

in ascribing email author identity with transformer neural networks. *2021 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, USA, 291-297, <https://doi.org/10.1109/SPW53761.2021.00049>

Wilkinson, S., **Knight, K.**, Kuchar, O., Mehta, K., Shankar, M. A., and Wolf, M. (2022). Official report on the 2021 computational and autonomous workflows workshop (caw 2021). (publication number ORNL/TM-2022/2402). Oak Ridge National Lab, Oak Ridge, TN (US). <https://doi.org/10.2172/1862119>

Wilkinson, S. R., Eisenhauer, G., Kapadia, A. J., **Knight, K.**, Logan, J., Widener, P., and Wolf, M. (2022). F*** workflows: when parts of fair are missing. *2022 IEEE 18th International Conference on e-Science (e-Science)*, Salt Lake City, UT, USA, 507-512. <https://doi.org/10.1109/eScience55777.2022.00090>

2 Presentations

Knight, K. (2024, March 11-14). Reproducibility and other R words [conference presentation]. SOS 26, Cocoa Beach, FL, United States. <https://sos26.ornl.gov/>

Knight, K. and Goethert, Ian. (2023, June 27-30). Secure Radiology Image Transfer and Management: Lessons Learned [conference presentation]. NLIT Summit 2023, Milwaukee, WI, United States. <https://www.govevents.com/details/61976/nlit-summit-2023>

May, A., Kuchar, O., **Knight, K.**, and Srivastava, R. (2022, June 1-3). Towards a DOE Data Catalog: Ensuring Access, Sharing, and Protection [conference presentation]. DOE Data Days, virtual.

May, A., Kuchar, O., **Knight, K.**, and Srivastava, R. (2022, June 1-3). Towards a Big-Data Toolkit: Ensuring Data Governance & Ethical Considerations Are Applied to Large Datasets. [conference presentation]. DOE Data Days, virtual.

May, A., Kuchar, O., and **Knight, K.**, and Srivastava, R. (2022, June 6-9). Who is Afraid of a Petabyte Dataset? Rethinking Repository Infrastructures and Curation Workflows for the Scale and Type of Next Generation Data [conference presentation]. Open Repositories, Denver, CO, United States. <https://doi.org/10.5281/zenodo.6833481>

Stall, S., Erdmann, C., Kirkpatrick, C., Goble, C., da Silva, R. F., Sun, Z., and **Knight, K.** (2021, December 13-17). Data fair working session: Fair workflows- best practices, services, and tools for helping researchers increase the portability and reproducibility of their work. AGU Fall Meeting 2021, virtual. <https://www.agu.org/fall-meeting-2021/pages/schedule>

3 Posters

Knight, K. and Kuchar, O. (2022). FAIR Data Stewardship and Leadership Computing: What Works, What's Needed, What's Next [poster]. ORNL Summer 2022 Poster Symposium. Oak Ridge Leadership Computing Facility, Oak Ridge, TN, United States.

4 Other

Program Committee member. 36th International Conference on Scientific and Statistical Database Management, July 10-12, 2024. Rennes, France. <https://ssdbm.org/2024/index>

Judge. 28th annual EURēCA (Exhibition of Undergraduate Research and Creative Achievement). University of Tennessee, April 23, 2024. <https://eureca.utk.edu>