

# Approaches for Computing Specimen Image Research Data



Presenters from the College of Computing and Informatics, Drexel University, Philadelphia PA, U.S.A.

- David Breen, Professor of Computer Science
- Joel Pepper, PhD student, Computer Science
- Jane Greenberg, Alice B. Kroeger Professor, Information Science

Over the last several decades advances in computing, imaging, and cyberinfrastructure have had a major impact on scientific research and discovery. One area of considerable activity is the digitization of the biological specimens that have been collected by museums and researchers all over the planet. The scanning of these specimen collections and the placement of the resulting images into easily accessible repositories on the Internet is enabling new scientific studies based on the previously unavailable data. Unfortunately, potential scientific advances are hindered by the lack of high-quality and pertinent metadata associated with the image collections. Metadata is required to search the repositories for the imaged specimens needed for a particular study. Since the collections may each contain tens of thousands of images, producing metadata for each image via a manual process is prohibitively labor-intensive and infeasible. Methods for automatically computing metadata from images are therefore needed to fully exploit biological image repositories for scientific discovery. In the first part of this talk we will present some approaches used in image informatics, the extraction of quantitative information from images, within the context of a fruit fly study. The second part will describe our current efforts to automatically analyze fish images as a step toward improving metadata in specimen research repositories. These fish specimens are being studied for a larger project (Biology Guided Neural Networks), which is developing a novel class of artificial neural networks that can exploit the machine readable and predictive knowledge about biology that is available in the form of specimen images, phylogenies and anatomy ontologies. We will close by highlighting the challenges that must be overcome when developing automatic image metadata generation techniques.

## Presenters Biographical Notes

David E. Breen is a Professor of Computer Science at Drexel University. He has held research positions at the Max Planck Institute for the Physics of Complex Systems, the California Institute of Technology, the European Computer-Industry Research Centre, the Fraunhofer Institute for Computer Graphics, and the Rensselaer Design Research Center. His research interests include textile modeling and design, biomedical image informatics, geometric modeling, and self-organization algorithms. Breen received a BA in Physics from Colgate University and MS and PhD degrees in Computer and Systems Engineering from Rensselaer Polytechnic Institute. He is a recipient of the NSF CAREER Award.

Jane Greenberg, Alice B. Kroeger Professor and Director of the Metadata Research Center, is an Information Science Professor at Drexel University. Her research activities focus on big metadata and machine-driven approaches, knowledge organization/linked data, data science, and information economics. She is a Principal investigator (PI) on the NSF Spoke initiative, "A Licensing Model and Ecosystem for Data Sharing" and the NSF Harnessing the Data Revolution (HDR) collaborative initiative, "Biology-Guided Neural Networks for Discovering Phenotypic Traits"). She is the lead PI for the IMLS LIS Education and Data Science Integrated Network Group (LEADING). Greenberg holds a BA in Political Science from the University of Rochester, a MLS from Columbia University, and a PhD degree in information science from the University of Pittsburgh.

Joel Pepper completed his undergraduate studies at The Ohio State University in 2019 in computer science and biomedical engineering. His undergraduate research was in laboratory-based cell and tissue engineering. He is currently pursuing a PhD in computer science at Drexel University with a research focus primarily in computer graphics. He is a research assistant working on the metadata aspects of the National Science Foundation sponsored "Biology Guided Neural Networks" project, working with Drexel Professors Jane Greenberg and David Breen as well as other collaborators.