WHITE PAPER

New Database Engineering and Archive Construction Technology to Accelerate Bio-Imaging, Biomedical Engineering, and Covid-19 Research

LTS (Linguistic Technology Systems) is founded by Amy Neustein, Ph.D., Series Editor of **Speech Technology and Text Mining in Medicine and Health Care** (de Gruyter); Editor of **Advances in Ubiquitous Computing: Cyber-Physical Systems, Smart Cities, and Ecological Monitoring** (Elsevier, 2020); co-author (with Nathaniel Christen) of **Cross-Disciplinary Data Integration and Conceptual Space Models for Covid-19** (Elsevier, 2021); and co-editor of **Medical Image Processing and Machine Learning** (Institution of Engineering and Technology, forthcoming).

Team

Principal Investigator: Dr. James A. Rodger, Professor of Management Information Systems and Decision Sciences at Indiana University of Pennsylvania

Administrative Officer: Dr. Amy Neustein, founder of Linguistic Technology Systems (LTS)

Contributors

- Nathaniel Christen, Lead Software Architect, LTS
- Professor Amita Nandal, Department of Department of Computer and Communication Engineering at Manipal University, Jaipur
- Professor Arvind Dhaka, Department of Department of Computer and Communication Engineering at Manipal University, Jaipur; recently visiting scholar at University of Varna, Bulgaria
- Professor Todor Ganchev, Vice Rector of Research at University of Varna, Bulgaria

---

The "MOSAIC Data-Set Explorer" (MdsX) and "MOSAIC Structured Reporting" (MOSAIC-SR) are tools to help authors develop interactive presentations supplementing academic documents (MOSAIC is an acronym for "Multi-Paradigm Ontologies for Scientific and Technical Publications"). With MdsX, interactive presentations take the form of software applications that provide access to data sets, analytic techniques, or other digitally representable artifacts to document or encapsulate research work. With MOSAIC-SR, authors can implement or reuse code libraries that report on research/experiment methods, workflows, and protocols. Conceptually, MOSAIC-SR is functionally similar to the various domain-speciic recommendations collectively gathered into the "Minimum Information for Biological and Biomedical Investigations" (MIBBI) specifications, and indeed one use-case for MOSAIC-SR is that of implementing object models instantiating MIBBI policies. In some contexts, MOSAIC-SR and MIBBI overlap, because elements of scientific workflows are sometimes algorithms implemented within a code package concretizing authors' research.

MOSAIC-SR can express both computational workflows that are fully encapsulated by published code as well as real-world protocols concerning laboratory equipment and physical materials or samples under investigation. In the latter guise, MOSAIC-SR code can employ or instantiate standardized terminologies and data structures for describing experiments — such as MIBBI policies or BIOCODER functions. In this case, the role of MOSAIC-SR code is to serve as a serialization/deserialization endpoint for sharing research metadata. Conversely, when workflows are fully implemented within software developed as part of a body of research, MOSAIC-SR can provide a functional interface allowing this code to be embedded in scientific software. For these cases, MOSAIC-SR

provides a framework for modeling how a software component specific to a given research project exposes its functionality to host and/or networked peer applications. There are also scenarios where both scenarios are relevant — the MOSAIC-SR code would simultaneously document real-world experimental protocols and construct a digital interface as part of a workflow which is part digital and part "real-world."

This paper will focus on one specific application of MOSAIC-SR in the context of image analysis and bioimaging — specifically, a "Data Structure Protocol for Image-Analysis Networking" (**D-SPIN**), which both extends and adds a narrower focus to the overall MOSAIC-SR framework.