

Advancing Biomedical Research Reuse through a Machine-FAIR Approach



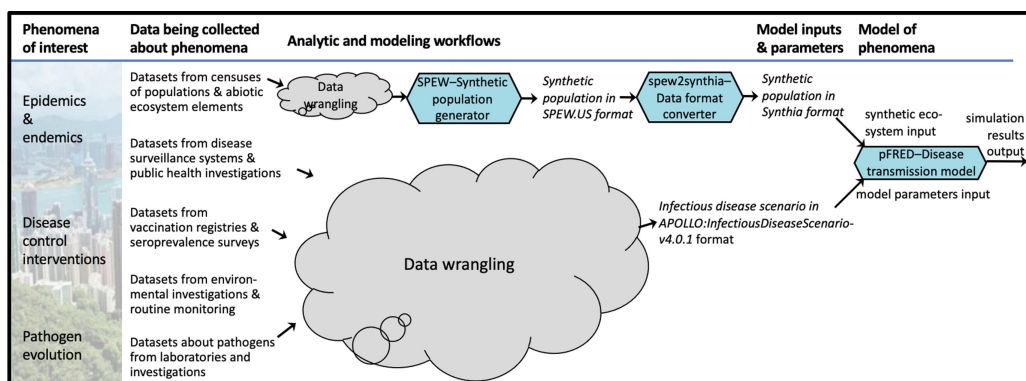
Efforts to promote reuse in biomedicine have a long history, dating back to 1879 with the creation of Index Medicus by the US National Library of Medicine (NLM). This initiative evolved into digital platforms like MEDLARS, MEDLINE, and PubMed, facilitating access to publications, datasets, and models. NLM's current strategic plan emphasises organising digital research objects to accelerate scientific discovery. The FAIR Guiding Principles, introduced in 2016, stress the importance of proper representation (metadata) to make digital objects findable, accessible, interoperable, and reusable by both humans and machines. Algorithmic search of digital repositories for valid workflow compositions has potential to accelerate scientific discovery but requires a scalable solution to the problem of knowledge acquisition about semantic constraints on software inputs. Additionally, practical limitations on the logical complexity of semantic constraints must be respected, which has implications for the design of software. [In a recent study published in the Journal of Biomedical Informatics](#), authors' objective was to develop representations and automated methods to satisfy the following novel machine use case (M1): Whenever a new digital object of research becomes available, a machine searches for existing objects that can be validly composed with the new object.

Addressing the Challenge of Automated Workflow Composition

The research aims to address the challenge of automatically searching through a collection of datasets and software to find valid compositions of workflows. This is crucial for facilitating the reuse of digital research objects in scientific endeavours. The primary focus is on developing a method, referred to as M1, which essentially involves automatically identifying valid compositions. This process requires acquiring knowledge about semantic constraints, which are essential for ensuring the validity of compositions. A key aspect of the proposed method is the utilisation of data-format matching as the foundation for composing workflows. This approach builds upon practices already employed by software developers to ensure that inputs to their software are valid. However, the research identifies significant discrepancies between current documentation practices and the requirements of M1 composition. This underscores the importance of incorporating M1-FAIR properties into software schemas and incentivizing developers to adhere to them.

Innovative Evaluation Methodology for Automated Workflow Composition

The evaluation methodology employed in the research is innovative, utilising error analysis to identify missing semantic constraints. The analysis also reveals conditions under which the recall of valid compositions can reach 100%. However, the results indicate that there is room for improvement, particularly in handling data services and representing semantic constraints. Despite achieving a precision rate of 61.7%, the research suggests that better handling of data services and semantic constraints could further enhance the accuracy of the method. Additionally, a significant portion of errors require the propagation of semantic constraints from inputs to outputs, highlighting the importance of comprehensive documentation.



Extending the Discussion: FAIR Interoperability and Future Directions

The discussion extends to the concept of FAIR Interoperability, conceptualising it as the composability of machines with the M1 approach. Different approaches, such as the Automated Workflow Generation (AWG) and Automated Service Composition (ASC), are compared in terms of their ability to address the challenge of semantic constraints. Looking towards the future, the research identifies the acquisition of semantic constraints as a critical area for further exploration. Scalable approaches, such as mining constraints from computational workflows or eliciting knowledge from software developers, are proposed. A roadmap is outlined, starting with making the developed method available and progressing towards versions incorporating semantic checking by procedural or declarative validators. The research aims to advance towards the creation of Machine-FAIR objects, which have the potential to significantly accelerate scientific discovery.

Source & Image Credit: [Journal of Biomedical Informatics](#)

Title Image Credit: [iStock](#)

Published on : Thu, 16 May 2024