

Editors' Note: Special Issue on Canonical Workflow Frameworks for Research Postprint

Authors: Peter, Wittenburg, Alex, Hardisty, Amirpasha, Mozzafari, Limor, Peer, Nikolay, Skvortsov, Alessandro, Spinuso, Zhiming, Zhao, Peter, Wittenburg

Date: 2022-11-28T00:00:00+00:00

Abstract

This special issue is on Canonical Workflow Frameworks for Research (CWFR). A workflow refers to a

sequence of activities, which may be more or less computer-based, used with regularity in the research

process. CWFR aim to identify common patterns in such scientifically motivated workflows and to offer

libraries of components based on FAIR Digital Objects as the integrative standard. Such CWFR components

can be reusable independent of particular technologies, benefitting researchers in their daily work by

making recurring activities more efficient, using automated workflow methods that would immediately

create FAIR compliant data without adding burden.

Full Text

Preamble

Editors' Note: Special Issue on Canonical Workflow Frameworks for Research

Peter Wittenburg^{1†}, Alex Hardisty², Amirpasha Mozaafari³, Limor Peer⁴, Nikolay Skvortsov⁵, Alessandro Spinuso⁶, Zhiming Zhao⁷

¹Gemeindweg 55, 47533 Kleve, Germany

²Cardiff University, Cardiff, South Glamorgan, CF14 3UX, Wales, UK

³Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

⁴Institution for Social and Policy Studies, Yale University, New Haven, CT 06520, USA

⁵Vavilov 44/2, 121351 Moscow, Russia

⁶Utrechtseweg 297, 3731 GA De Bilt, the Netherlands

⁷University of Amsterdam, PO-Box 94323, 1090 GH Amsterdam, the Netherlands

Citation: Wittenburg, P. et al.: Editors' Note: Special Issue on Canonical Workflow Frameworks for Research. *Data Intelligence*. 4(2), 149-154 (2022). DOI: 10.1162/dint_e_{00122}

This special issue focuses on Canonical Workflow Frameworks for Research (CWFR). A workflow refers to a sequence of activities, which may be more or less computer-based, used with regularity in the research process. CWFR aims to identify common patterns in such scientifically motivated workflows and to offer libraries of components based on FAIR Digital Objects as the integrative standard. Such CWFR components can be reusable independent of particular technologies, benefitting researchers in their daily work by making recurring activities more efficient, using automated workflow methods that would immediately create FAIR-compliant data without adding burden.

The goal of this special issue is to provide readers with a deep exploration of CWFR and how it relates to research-driven workflows, existing workflow technologies, and the use of FAIR Digital Objects. The issue covers articles examining core research activities including experimentation, data processing and analysis, data management, reproducibility, and publication. The articles comment on CWFR and its relation to these workflows, either conceptually in view of the current research ecosystem and infrastructure or more practically, focusing on a specific implementation, design, tool, or context relating to CWFR.

The contributing authors are experts in their area. They include researchers, data professionals, data managers and curators, IT specialists, and others who are using, developing, or experimenting with the effective use of canonical workflows and workflow patterns for data-intensive research. As guest editors, it has been a privilege to work with such accomplished authors.

It is our hope that this issue will stimulate further exploration of this subject. The papers in this issue address timely questions such as: What are the recurring patterns of work within or across institutions and research communities? What are the core elements of workflow technologies and how can they relate to the core ideas of CWFR? How well do existing integration standards and best practices address this? What is the potential of FDOs to support the goals of CWFR? How can research be protected against the ever-changing technological fashions?

Finally, we are grateful to the journal for the opportunity to publish this special issue and to Dr. Fenghong Liu, Managing Editor-in-Chief, for her skilled

guidance and support.

Corresponding author: Peter Wittenburg (Email: peter.wittenburg@mpcdf.mpg.de; ORCID: 0000-0003-3538-0106)

Author Biographies

Peter Wittenburg has a background in electrical engineering and has worked as Technical Director at the Max Planck Institute for Psycholinguistics for many years, acting as a member of the IT Advisory Board of the President of the Max Planck Society. The Max Planck Institute focused from the beginning on digital technologies to understand the functioning of the brain with respect to language processing. The institute's need for accessing data from other institutes to feed the stochastic engines they applied rather early led him to become an expert in building data/research infrastructures. He was responsible for the technological aspects of three large international and European research infrastructures: DOBES, CLARIN, and EUDAT. In this function, he understood that data work across silos is highly inefficient and that harmonization and standardization are required to improve the situation. This was the reason he co-founded the Research Data Alliance in 2013 and the FAIR Digital Object Forum in 2019. ORCID: 0000-0003-3538-0106

Alex Hardisty was, before his recent retirement, Director of Informatics Projects in the School of Computer Science and Informatics at Cardiff University, UK. He is interested in bio/geodiversity informatics, the engineering of large-scale distributed information systems for data management and processing, virtual research environments, and socio-technical issues of new technology adoption. Alex is a technical architect. Before his retirement, he was leading DiSSCo technical work on open Digital Specimens (openDS), Minimum Information about Digital Specimens/Collections (MIDS/MICS), and exploiting machine-actionable FAIR Digital Objects. Alex co-chaired the CWFR Working Group of the FDO Forum and was a member of the FDO Forum's Technical Specification and Implementation (TSIG) Working Group and the FDO Forum Steering Committee. ORCID: 0000-0002-0767-4310

Amirpasha Mozaffari is a postdoctoral researcher in the group on Earth System Data Exploration (ESDE) at the Jülich Supercomputing Centre (JSC). He is trained as a geoscientist and recently defended his PhD in computational geohydrophysics from RWTH Aachen. He is active in the field of data management, workflow design, and FAIR data practices. He is co-chair of the canonical workflow framework for research in the FAIR Digital Object forum. ORCID: 0000-0001-6719-0425

Limor Peer, PhD, is Associate Director for Research and Strategic Initiatives at the Institution for Social and Policy Studies (ISPS), Yale University. Limor works on research transparency and reproducibility and is especially interested in the connection between generating and preserving scientific knowledge. She created the ISPS Data Archive, a digital repository for research produced by

scholars affiliated with ISPS with a focus on experimental design and methods. She led the project to develop YARD, the Yale Application for Research Data, a workflow tool for reviewing and enhancing research outputs. Limor is co-founder of the CURE (Curation for Reproducibility) Consortium of social science data archives. She co-chairs the CURE-FAIR working group at the Research Data Alliance and the Practices working group of the ACM's Emerging Interest Group on Reproducibility and Replicability. She sits on the board of the Roper Center for Public Opinion Research and serves on a number of advisory and task force groups working on data curation and research transparency. Prior to joining ISPS, Limor was Research Director at Northwestern University's Media Management Center and Readership Institute and Associate Professor (clinical) at the Medill School of Journalism. ORCID: 0000-0002-3234-1593

Nikolay Skvortsov's general research interests are ontological and conceptual modeling of research domains and data semantic interoperability issues. He has been affiliated with the Institute of Informatics Problems, Federal Research Center "Computer Science and Control," Russian Academy of Sciences, Moscow, Russia. In recent years, Nikolay Skvortsov has investigated requirements for the reuse of data, research methods, and processes in research communities, primarily using examples of problem development and solving in astronomical research domains. ORCID: 0000-0003-3207-4955

Alessandro Spinuso is a researcher at the R&D Observations and Data Technology division of the KNMI (Royal Netherlands Meteorological Institute). He earned his PhD in Computer Science at the University of Edinburgh (UK) in 2017. At KNMI, he covers the roles of Researcher and Product Owner within an Agile R&D team developing Provenance-aware Data Analysis services. His main research interest is the management and exploitation of provenance information in the context of user-controlled computational environments, providing notebooks and workflow systems for data-intensive analysis. He is involved in several international initiatives focusing on the development of e-science infrastructures for Earth Science research in Europe (EPOS, ENVRIFair, IS-ENES3, DARE, C3S). More recently, he is an invited expert to the IPCC TG-Data, a working group dedicated to the FAIR management of the data and methods that will be published in the next IPCC reports. ORCID: 0000-0002-0077-8491

Zhiming Zhao received his Ph.D. in computer science in 2004 from the University of Amsterdam (UvA). He is an assistant professor in the Multiscale Network Systems (MNS) at UvA and the technical manager of the Virtual Lab and Innovation Center (VLIC) of LifeWatch ERIC, a European research infrastructure in ecology and biodiversity science. His research focuses on innovative programming and control models for quality-critical systems on programmable infrastructures such as Clouds, Edges, and Software-Defined Networking using optimization, semantic linking, blockchain, and artificial intelligence technologies. He leads the UvA effort in several EU projects including ARTICONF, CLARIFY, ENVRI-FAIR, and SWITCH. ORCID: 0000-0002-6717-9418

Yann Le Franc, PhD is the CEO and Scientific Director of e-Science Data Fac-

tory S.A.S.U., a French R&D company aiming at proposing innovative solutions for FAIR data management to accelerate growth and progress. Yann Le Franc has a PhD in Neurosciences and Pharmacology (2004). After a postdoctoral experience in the US, he worked on data management projects in Neurosciences at the University of Antwerp (Belgium) and in the context of the International Neuroinformatics Coordinating Facility (INCF), where he developed a strong expertise in ontology design and semantic web technologies. He then contributed to several Horizon 2020 Research Infrastructure projects (EUDAT, EOSC-Hub, etc.) as an expert on Semantic Web and ontology design. He is co-chairing the Research Data Alliance Vocabulary and Semantic Service Interest Group and the FDO Semantic Group. He is also a member of the EOSC Semantic Interoperability Task Force. He is actively involved in the FAIRification and standardization of semantic artefacts in the context of FAIRsFAIR and OntoCommons projects. In parallel, he is the technical manager of the EOSC-Pillar project for the French National Computing Center for Higher Education (CINES). ORCID: 0000-0003-4631-418X

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.