

International Journal of Population Data Science

Journal Website: www.ijpds.org



CR8TOR: Research Object Crates for provisioning reproducible infrastructure in Kubernetes-native federated Trusted Research Environments

Mike Harding¹, Alwin Thomas², Saurav Nair Sudhar², and Vishnu Vardhan Chandrabalan³

¹Lancaster University, Lancaster, United Kingdom

²Lancashire Teaching Hospitals NHS Foundation Trust, Preston, United Kingdom

³Lancashire Teaching Hospitals NHS Foundation Trust, Preston, United Kingdom; Lancaster University, Lancaster, United Kingdom

Orchestration of infrastructure and data resources for research on sensitive data in Trusted Research Environments (TREs) is complex and time-consuming, with fragmented processes for data governance, identity management, disclosure checks, and resource provisioning. Research Object Crates (RO-Crates) aim to make metadata and analytical outputs FAIR (findable, accessible, interoperable, reusable), but current approaches neither record infrastructure requirements in RO-Crates nor use them for automated provisioning. CR8TOR addresses this challenge by extending standardised metadata models (Five-Safes RO-Crate) to take an active role in creating project-dependent infrastructure. Built on the Kubernetes Operator pattern within K8TRE (a Kubernetes-native, cloud-agnostic TRE implementation), CR8TOR demonstrates how contextual project information modelled in RO-Crate-like schemas can provide machine-readable representations of research resource requirements. By extending the Kubernetes API with custom resource definitions (CRDs), these resource

descriptions sit alongside traditional research object metadata while maintaining FAIR principles, acting as computable units that can be recreated when required. CR8TOR supports automated data ingress from external sources, project governance controls, data provenance tracking, and on-demand reconciliation of project resources including analytics workspaces, user accounts, and access/network policies. Deployed across cloud and on-premises infrastructure supporting multi-institution collaborations, the same declarative models can be shared across deployments, enabling reproducible and portable research settings. This work demonstrates how provisionable metadata may strengthen methodological foundations of federated research by allowing TRE infrastructure specifications—including workspace configuration, controlled access rules, and project-specific identities—to be described, shared, and reproduced with clarity, consistency, and automation, ultimately lowering operational costs while improving governance and reproducibility.

