Annual South African National Research Data Workshop 2024

Title: Crafting Open Data for Open Science: Technical Innovation and Data Management in Environmental Research- The SAEON Open Data Platform

Authors: Mark Jacobson (mj.jacobson@saeon.nrf.ac.za) and Daniëlle Seymour (da.seymour@saeon.nrf.ac.za), uLwazi Node, South African Environmental Observation Network

Abstract

The South African Environmental Observation Network (SAEON) is one of the National Research Foundation (NRF)'s Research Infrastructure Platforms and serves as a sustained, coordinated, responsive and comprehensive South African earth observation network. SAEON delivers long-term, reliable data for scientific research and informs decision-making to support a knowledge society and improve quality of life.

The Open Data Platform (ODP) is one of SAEON's research data infrastructure comprising an aggregation of databases, services and web applications that facilitate the preservation, publication, discovery, and dissemination of earth observation and environmental data in South Africa. The ODP was certified as a trustworthy data repository by CoreTrustSeal in 2023. In the evolving landscape of environmental data, which continuously changes in response to technological, societal, and scientific developments, effective curation and publication processes are crucial for ensuring dataset accessibility and usability. SAEON's commitment to the FAIR principles-making data Findable, Accessible, Interoperable, and Reusable-guides data management and dissemination practices. SAEON ensures that metadata is comprehensive and adheres to established standards, enabling users to understand the context and quality of the data.

While the assemblage of systems constituting the ODP has grown organically over many years, an information architecture has simultaneously evolved to enable the centralised metadata management system to accept data submissions from a variety of sources. It permits quality control and value-added annotations by SAEON's data curation team. Additionally, it allows for interoperability with and publication of metadata to a variety of data cataloguing systems both locally and globally. The development of this abstract information architecture has led to the emergence of useful high-level patterns including many-to-many connectivity between data producers (archives) and data consumers (catalogues),

differentiated access control supportive of multitenancy, and an extensible ontological framework.

In this presentation, we will cover:

- SAEON's approach to managing environmental data publication, including rigorous curation process, quality checks, and the assignment of DOIs;
- The role of comprehensive metadata in enhancing discoverability and user engagement;
- Characteristics of the ODP information architecture and its utility in real-world use cases.