



## Implementing RDA Data Citation Recommendations: Case Study in South Africa

Wim Hugo

NRF, SAEON, PAARL, South Africa (wim@saeon.ac.za)

SAEON operates a shared research data infrastructure for its own data sets and for clients and end users in the Earth and Environmental Sciences domain. SAEON has a license to issue Digital Object Identifiers via DataCite on behalf of third parties, and have recently concluded development work to make a universal data deposit, description, and DOI minting facility available. This facility will be used to develop a number of end user gateways, including DataCite South Africa (in collaboration with National Research Foundation and addressing all grant-funded research in the country), DIRISA (Data-intensive Research Infrastructure for South Africa - in collaboration with Meraka Institute and Department of Science and Technology), and SASDI (South African Spatial Data Infrastructure).

The RDA recently published Data Citation Recommendations [1], and this was used as a basis for specification of Digital Object Identifier implementation, raising two significant challenges:

1. Synchronisation of frequently harvested meta-data sets where version management practice did not align with the RDA recommendations, and
2. Handling sub-sets of and queries on large, continuously updated data sets.

In the first case, we have developed a set of tests that determine the logical course of action when discrepancies are found during synchronization, and we have incorporated these into meta-data harvester configurations. Additionally, we have developed a state diagram and attendant workflow for meta-data that includes problem states emanating from DOI management, reporting services for data depositors, and feedback to end users in respect of synchronisation issues.

In the second case, in the absence of firm guidelines from DataCite, we are seeking community consensus and feedback on an approach that caches all queries performed and subsets derived from data, and provide these with anchor-style extensions linked to the dataset's original DOI. This allows extended DOIs to resolve to a meta-data page on which the cached data set is available as an anchored download link. All cached datasets are provided with checksum values to verify the contents against such copies as may exist.

The paper reviews recent service-driven portal interface developments, both services and graphical user interfaces, including wizard-style, configurable applications for meta-data management and DOI minting, discovery, download, visualization, and reporting. It showcases examples of the two permanent identifier problem areas and how these were addressed.

The paper concludes with contributions to open research questions, including (1) determining optimal meta-data granularity and (2) proposing an implementation guideline for extended DOIs.

[1] A. Rauber, D. van Uytvanck, A. Asmi, S. Pröll, "Data Citation Recommendations", November 2015, RDA. <https://rd-alliance.org/group/data-citation-wg/outcomes/data-citation-recommendation.htm>