Sharing is caring for the WIO: data governance and management practices of the region

Louis Celliers*, Nicolene Fourie, Julius Francis
*Presenting author

Western Indian Ocean Marine Science Association, United Republic of Tanzania
Louis Celliers (louis.celliers@gmail.com) Nicolene Fourie (nicolene.4e@gmail.com), Julius Francis (julius@wiomsa.org)

Background
Data are fundamentally part of scientific and social research. It is a versatile good that exists as both the raw material and the product of knowledge itself. ‘Open science’, or freely accessible research and data is only possible if data are managed. A lack of, or limited sharing of data results in “slow” science. The practical aspects of data management remain a struggle for many research communities. The Western Indian Ocean Marine Science Association (WIOMSA) is an important contributor of data, information and knowledge to the benefit of regional and national coastal and marine spatial planning. Currently, the location, state, availability and accessibility of a substantial regional investment in spatial data (and other) is not apparent. While there may be a repository of reports and other written outputs, the data underlying the MASMA-funded research is not discoverable and therefore not available or accessible for re-use. This paper reports on a survey of data governance and management practices in the WIOMSA community. It concludes with potential strategies and policies for WIOMSA in order to improve the return on investments made in data generation through, e.g., MASMA, and others.

Methods
The project team used three methods to gather data and information to address the objectives of this project. The first was a desktop review of literature relating to data management and governance practice that is relevant to WIOMSA and its institutional objectives. Secondly, Principal Investigators (and their institutions) of MASMA projects were questioned about their project data management and governance practices. This was done using a structured questionnaire in three parts, viz. S1 Project details; S2 Project data outputs; and S3 Data management. The objective of this questionnaire was to evaluate the extent and manner in which MASMA funded projects are compliant with international standards of data management. Finally, the project team analysed four data management scenarios for possible implementation by WIOMSA.

Results
The MASMA Programme has funded 48 projects up to the end of 2017. Eight of these projects were ongoing at the time that this project was undertaken (see Appendix 10.5). Questionnaires were sent to all Primary Investigators and 18 completed or partially-completed responses were received. This represents a 37.5% return of the total number of projects funded. The questionnaire requested Project Investigators to provide the detail of at least three datasets that were produced during the life-cycle of the MASMA project (normally 3-4 years). This resulted in 62 datasets being listed by Project Investigators.

Key findings are the paucity in the generation of metadata (only 29%) and the lack of adherence to the use of standards for data or metadata. Most of the data are quantitative with very little spatial attribution for any of the data generated. As expected, the reported data formats favoured the use of spreadsheets, with rare examples of data being stored in
databases. The large category of MASMA investigators reporting data stored in “other formats” require better resolution and possible action, depending on the actual formats used.

As a result of the lack of metadata, the data can only be found indirectly, or “stumbled upon”. A worrying tendency is the number of MASMA investigators (of completed projects) that list their data as conditionally available or unavailable due to use-restrictions. Although there may be legitimate reasons why data are not available, the combination of a lack of metadata and the imposed use-restrictions implies that a very small proportion of data are in the public domain and available for re-use.

Most projects suggested the existence of data management plans but the request for details of such plans indicated that such plans were not formally developed. Also, more than half of the dataset could not be found other than through word of mouth and contact with a project team member. In general, the responses to the questionnaire indicated a very low effort in maintaining discoverability of data.

Conclusions
From the findings of the questionnaire on current MASMA data governance, it is clear that there is substantial scope for the improvement of data management and the contribution of MASMA data and products to open data for the WIO region. It is feasible the MASMA Programme to make small changes in how is recognises the importance of data, which, if converted into action, will result in a substantial improvement in the return on investment in data generation.

This assessment describes four scenarios for the improvement of data management of the MASMA Programme data. The scenarios were defined according to actions relating to the creation of metadata, the use of identifiers (both data and people), adherence to the FAIR (findable, accessible, interoperable, usable) principles, the creation of DMPs, and the expenditure of resources for data management. The scenarios were also assessed according to the contribution to achieving the objectives of this project, which were to: 1) monitor and evaluate the use of MASMA data for regional impact; 2) communicate the value of MASMA project outputs, outcomes and impacts, including the long-term value of regionally important data; 3) reduce duplication of investment in data; and, 4) ensure that MASMA data can be known (discovered) and contribute to a regional pool of knowledge. The scenarios were:

1. MASMA “Business-as-Usual”: where WIOMSA undertakes no actions to improve data management of its funding programme
2. MASMA Metadata: where WIOMSA only requires the creation of metadata according to the ISO 19115 and Dublin Core standards; and the metadata is curated and published by WIOMSA
3. MASMA Open Data: where WIOMSA requires metadata data creation according to the mentioned standards, but also that project teams submit metadata and data to stable international, regional or national data platforms or repositories.
4. MASMA Data System: where WIOMSA invests in the development of its own data and metadata curation, storage and publication systems.

It is conceivable that WIOMSA could play a meaningful regional role to provide guidance and leadership with regards to good data management. This role could only be defined if the Association firstly improves its own data management processes and procedures, and negotiates this role with other regional agencies.
The authors gratefully acknowledge the discussion with, comments and input from Lucy Scott regarding the regional outlook for data management.