

## NOTE ON THE COLLABORATION OF THE SECRETARIAT WITH INTERNATIONAL PARTNERS ON DATA-RELATED MATTERS

PREPARED BY: IOTC SECRETARIAT, LAST UPDATED: 26 NOVEMBER 2022

### Purpose

To provide the participants at the 18th Working Party on Data Collection and Statistics (WPDCS18) with information on the potential collaborations of the Secretariat with institutes and non-governmental organisations developed to improve the management of metadata for describing the data sets managed by the Secretariat and enrich the sources of information available for scientific analyses.

### Background

The Secretariat of the Indian Ocean Tuna Commission (IOTC) manages several data sets related to the fisheries catching tuna and tuna-like fisheries in the Indian Ocean. The size, coverage, resolution, and origin of the data sets varies according to the type of data considered. Most data are collated from the Contracting Parties and Cooperating Non-Contracting Parties (CPCs) through data requirements included in IOTC Conservation and Management Measures (CMMs). Fisheries data sets include vessel registries (Res. [10/08](#) and [19/04](#)), quantitative information on effort exerted by the fisheries (e.g., numbers of hooks deployed, daily position of satellite-tracked buoys equipping drifting fish aggregating devices) (Res. [15/02](#) and [19/02](#)), magnitude and composition of the retained and discarded catches (Res. [15/02](#)) as well as detailed information on gear configuration (e.g., numbers of hooks between floats), vessels positions, and other fisheries activities (e.g., search) collected through scientific observers at sea (Res. [22/04](#)). Information on retained catches is also collected through observers deployed on carrier vessels who monitor transshipments at sea from large-scale longliners (Res. [22/02](#)) while port calls and landings are monitored by national authorities, mostly within the framework of Port States Measures (PSMs; Res. [16/11](#)).

In addition to fisheries data, the Secretariat maintains a database of morphometric data to derive relationships and conversion factors for harmonizing size-frequency data and computing individual weights from length measurements (IOTC 2005; Geehan and Pierre 2013). Furthermore, some morphometric, biological (e.g., age, sex), and tagging data on tuna and tuna-like species have been collected through national and regional projects conducted by CPCs under the umbrella of the IOTC (Murua et al. 2015; Farley et al. 2019, 2021). To complement this information, the Secretariat collaborates with some CPCs to collate additional morphometric data so as to increase sample size and improve the conversions used in data processing procedures (e.g., IOTC et al. 2022). Similarly, the Working Party on Billfish has recently noted the interest of sharing tagging data at regional level to increase sample size and requested the Secretariat to develop a regional database for tagging data (IOTC 2022). Finally, the IOTC has also the mandate to “keep under review the economic and social aspects” of the Indian Ocean tuna and tuna-like fisheries and the Secretariat has started collating price data from the Pacific Islands Forum Fisheries Agency (FFA) to circumvent the limited amount of socio-economic data available for scientific analysis (IOTC 2021).

### Data confidentiality aspects

Most data sets managed by the Secretariat are in the public domain and routinely updated and published on the [IOTC website](#) while non-public domain data can be accessible under certain conditions. The confidentiality policy and procedures linked to the storage and dissemination of catch, effort, and tagging data managed by the IOTC Secretariat are set in [Res. 12/02](#). Briefly, access to logbook-level information or detailed observer data is restricted to IOTC staff

and data can only be made available with the permission of the source or after aggregation following standard stratification, provided that the catch of no individual vessel can be identified within a time/area stratum. Furthermore, the FAO administrative circular 2022/06 sets out the FAO data protection principles which are aligned with international data protection standards and controls, including those of the United Nations system and are applicable to the entire lifecycle of data processing, from initial collection, storage to deletion ([IOTC-2022-WPDCS18-INF04](#)). The FAO data protection policy relies on five first principles (fairness, integrity, responsibility, security, and transparency), defines four confidentiality levels (public, internal, confidential, and strictly confidential), and states that all FAO personnel, including IOTC secretariat staff, are individually responsible with this policy and related instruments. Hence, the FAO administrative circular 22/06 ensures the full confidentiality and security of the data sets that would be shared with the Secretariat at levels of stratification finer than the standard stratification described in Res. 12/02.

## Augmenting data sources

Several data sets have been identified as potential additional sources to complement the data available from the Secretariat and support scientific analyses conducted as part of the IOTC work. These data sets can concern both the biology and ecology of tuna and tuna-like species as well as the socio-economic dimension of the coastal and high seas fisheries, and the vessels movements and their areas of operations (**Table 1**).

**Table 1.** Examples of data sets of potential interest to support scientific analyses conducted as part of the IOTC work

Data type	Methods	Applications	Sources
Biology	Morphometrics	Length-length and length-weight relationships, conversion factors	CPCs, NGOs
	Direct age estimation	Growth, maximum age	
	Histological analysis	Maturity at size/age, spawning season, reproductive capacity	
	Oocyte counts	Fecundity at size/age, maternal effects	
Ecology	Conventional tagging	Growth, movements and migrations, mortality	CPCs, NGOs
	Satellite tagging	3D movements, habitat preferences, post-release mortality	
	Stable isotopes of N and C	Trophic position and niche, feeding habits, habitat use, intra and inter-species competition	
	Stomach contents	Feeding preferences, trophic competition	
Economics	Time series of price	Economic value, fishery targeting	CPCs, NGOs, <a href="#">FAO</a> , <a href="#">FFA</a>
	International trade statistics	Trade flows, food security	
	GDP from fisheries	Fisheries dependence in coastal countries	
	Employment in fisheries	Socio-economic importance of tuna fisheries	
Vessel activities	VMS, AIS, VIIRS day/night band	Fishing grounds, fishing effort	<a href="#">GFW</a> , <a href="#">SeaVision</a>

Regarding biological and ecological data, recent contacts have been made with the International Game Fish Association (<https://igfa.org/>) who has been collecting morphometric data on billfish over several decades and running a satellite tagging program, the IGFA Great Marlin Race, that has been tagging billfish since 2011 in the world oceans. Other NGOs such as The Billfish Foundation (<https://billfish.org/>) and Marine Megafauna (<https://marinemegafauna.org/>) have also been involved in the tagging of billfish with both conventional and satellite tags in the Indian Ocean (Rohner et al. 2018, 2020, 2021). Regarding the spatial extent of the habitat of tuna and tuna-like species due to their highly migratory behaviour and the high costs associated with sampling and tagging operations in the high seas, the sharing and pooling of data appears to be essential to describe ecological features and patterns and study their drivers at

---

regional scales, e.g., to characterize shark migrations and vertical movements (Queiroz et al. 2019; Andrzejaczek et al. 2022).

In this context, the Secretariat could act as data custodian by collating, gathering, and describing the data sets made available from different sources to support the IOTC scientific process. The level of confidentiality of each data set would be determined and a specific data request form would be used to define the modalities of use of the data, including contribution to the analyses and co-authorship ([Appendix I](#))

## Improving metadata management

The description of the IOTC data sets could be improved to increase their visibility and facilitate the use of the data to support the IOTC Working Parties. The development of proper metadata compliant with the FAIR data management principles (see [IOTC-2022-WPDCS18-24](#)) could strongly benefit from a regional project funded by the EU and led by IRD, the “Grand Observatoire de l’Océan Indien” (G2OI), which aims to enhance the value of research data in the southwest Indian Ocean region by bringing together a network of partners who manage reference data in the IOC countries (Reunion, Comoros, Madagascar, Mauritius, and Seychelles).

The overarching objective of the project is that the partners will create and feed a regional data infrastructure that will allow end-users (scientists, managers, and general public) to easily exploit the regional data heritage through web services made available on a data portal dedicated to the infrastructure. The selected technical solutions, supported by a series of meetings with partners, will allow to reinforce the national data infrastructures already in place and to exchange information more easily (interoperability) with similar initiatives in other regions of the Indian Ocean and at the global level. The technical description of the project is provided in [Appendix II](#).

Some preliminary activity linked to the G2OI project have been presented at the 15<sup>th</sup> session of the IOTC Working Party on Data Collection and Statistics (WPDCS) held in 2019 which agreed on the interest of these activities for improving the description and management of the data and publications managed by the Secretariat, further noting that some potential funding might be needed to continue this activity. The activities were subsequently included in the program of work of the WPDCS and endorsed by the Scientific Committees in 2019 and 2020 but not performed in absence of funding. The G2OI project provides an opportunity to strengthen the management of the data sets hosted at the IOTC Secretariat through funding and collaboration with some national institutions and universities located in the southwest Indian Ocean.

## References

- Andrzejczak, S., Lucas, T.C.D., Goodman, M.C., Hussey, N.E., Armstrong, A.J., Carlisle, A., Coffey, D.M., Gleiss, A.C., Huvneers, C., Jacoby, D.M.P., Meekan, M.G., Mourier, J., Peel, L.R., Abrantes, K., Afonso, A.S., Ajemian, M.J., Anderson, B.N., Anderson, S.D., Araujo, G., Armstrong, A.O., Bach, P., Barnett, A., Bennett, M.B., Bezerra, N.A., Bonfil, R., Boustany, A.M., Bowlby, H.D., Branco, I., Braun, C.D., Brooks, E.J., Brown, J., Burke, P.J., Butcher, P., Castleton, M., Chapple, T.K., Chateau, O., Clarke, M., Coelho, R., Cortes, E., Couturier, L.I.E., Cowley, P.D., Croll, D.A., Cuevas, J.M., Curtis, T.H., Dagorn, L., Dale, J.J., Daly, R., Dewar, H., Doherty, P.D., Domingo, A., Dove, A.D.M., Drew, M., Dudgeon, C.L., Duffy, C.A.J., Elliott, R.G., Ellis, J.R., Erdmann, M.V., Farrugia, T.J., Ferreira, L.C., Ferretti, F., Filmlalter, J.D., Finucci, B., Fischer, C., Fitzpatrick, R., Forget, F., Forsberg, K., Francis, M.P., Franks, B.R., Gallagher, A.J., Galvan-Magana, F., García, M.L., Gaston, T.F., Gillanders, B.M., Gollock, M.J., Green, J.R., Green, S., Griffiths, C.A., Hammerschlag, N., Hasan, A., Hawkes, L.A., Hazin, F., Heard, M., Hearn, A., Hedges, K.J., Henderson, S.M., Holdsworth, J., Holland, K.N., Howey, L.A., Hueter, R.E., Humphries, N.E., Hutchinson, M., Jaïne, F.R.A., Jorgensen, S.J., Kanive, P.E., Labaja, J., Lana, F.O., Lassauce, H., Lipscombe, R.S., Llewellyn, F., Macena, B.C.L., Mambrasar, R., McAllister, J.D., McCully Phillips, S.R., McGregor, F., McMillan, M.N., McNaughton, L.M., Mendonça, S.A., Meyer, C.G., Meyers, M., Mohan, J.A., Montgomery, J.C., Mucientes, G., Musyl, M.K., Nasby-Lucas, N., Natanson, L.J., O'Sullivan, J.B., Oliveira, P., Papastamtiou, Y.P., Patterson, T.A., Pierce, S.J., Queiroz, N., Radford, C.A., Richardson, A.J., Richardson, A.J., Righton, D., Rohner, C.A., Royer, M.A., Saunders, R.A., Schaber, M., Schallert, R.J., Scholl, M.C., Seitz, A.C., Semmens, J.M., Setyawan, E., Shea, B.D., Shidqi, R.A., Shillinger, G.L., Shipley, O.N., Shivji, M.S., Sianipar, A.B., Silva, J.F., Sims, D.W., Skomal, G.B., Sousa, L.L., Southall, E.J., Spaet, J.L.Y., Stehfest, K.M., Stevens, G., Stewart, J.D., Sulikowski, J.A., Syakurachman, I., Thorrold, S.R., Thums, M., Tickler, D., Tolloti, M.T., Townsend, K.A., Travassos, P., Tyminski, J.P., Vaudo, J.J., Veras, D., Wantiez, L., Weber, S.B., Wells, R.J.D., Weng, K.C., Wetherbee, B.M., Williamson, J.E., Witt, M.J., Wright, S., Zilliaccus, K., Block, B.A., and Curnick, D.J. 2022. Diving into the vertical dimension of elasmobranch movement ecology. *Sci. Adv.* **8**(33): eabo1754. American Association for the Advancement of Science. doi:10.1126/sciadv.abo1754.
- Farley, J., Eveson, P., Bonhommeau, S., Dhurmeea, Z., West, W., and Bodin, N. 2019. Growth of albacore tuna (*Thunnus alalunga*) in the western Indian Ocean using direct age estimates. IOTC, Kuala Lumpur, Malaysia, 14-17 January 2019. p. 19. Available from <https://www.iotc.org/documents/WPTmT/07/21>.
- Farley, J., Krusic-Golub, K., Eveson, P., Clear, N., Luque, P.L., Artetxe-Arrate, I., Fraile, I., Zudaire, I., Vidot, A., Govinden, R., Ebrahim, A., Romanov, E., Chassot, E., Bodin, N., Murua, H., Marsac, F., and Merino, G. 2021. Estimating the age and growth of bigeye tuna (*Thunnus obesus*) in the Indian Ocean from counts of daily and annual increments in otoliths. IOTC-2021-WPTT23-18.
- Geehan, J., and Pierre, L. 2013. Biological data on tuna and tuna-like species gathered at the IOTC Secretariat: Status report (update). IOTC, Ninth session of the IOTC Working Party on Data Collection and Statistics. p. 23. Available from <https://iotc.org/documents/biological-data-tuna-and-tuna-species-gathered-iotc-secretariat-status-report>.
- IOTC. 2005. Biological data on tuna and tuna-like species gathered at the IOTC Secretariat: Status Report. IOTC, Phuket, Thailand, 18-22 July 2005. p. 22. Available from <https://iotc.org/documents/biological-data-tuna-and-tuna-species-gathered-iotc-secretariat-secretariat>.
- IOTC. 2021. Fostering the work of the IOTC with socio-economic data sets sourced from FAO-GLOBEFISH and the Pacific Islands Forum Fisheries Agency (FFA). IOTC, Virtual meeting, 29 November - 3 December 2021. p. 5. Available from <https://www.iotc.org/documents/fostering-work-iotc-socio-economic-data-sets-sourced-fao-globefish-and-pacific-islands>.
- IOTC. 2022. Report of the 20th Session of the IOTC Working Party on Billfish. IOTC, Online Meeting, 12-15 September 2022. Available from <https://www.iotc.org/documents/WPB/20/RE>.
- IOTC, Wu, R.-F., Zhu, J., Nishida, T., Bonhommeau, S., Kitakado, T., and Hoyle, S. 2022. Preliminary analysis of the variability in the length-weight relationship of Indian Ocean albacore. IOTC, Virtual meeting, 25-29 July 2022. p. 21. Available from <https://www.iotc.org/documents/WPTmT/802/06>.
- Murua, H., Eveson, J.P., and Marsac, F. 2015. The Indian Ocean Tuna Tagging Programme: Building better science for more sustainability. *Fish. Res.* **163**: 1–6. doi:10.1016/j.fishres.2014.07.001.
- Queiroz, N., Humphries, N.E., Couto, A., Vedor, M., da Costa, I., Sequeira, A.M.M., Mucientes, G., Santos, A.M., Abascal, F.J., Abercrombie, D.L., Abrantes, K., Acuña-Marrero, D., Afonso, A.S., Afonso, P., Anders, D., Araujo, G., Arauz, R., Bach, P., Barnett, A., Bernal, D., Berumen, M.L., Bessudo Lion, S., Bezerra, N.P.A.,

- Blaison, A.V., Block, B.A., Bond, M.E., Bonfil, R., Bradford, R.W., Braun, C.D., Brooks, E.J., Brooks, A., Brown, J., Bruce, B.D., Byrne, M.E., Campana, S.E., Carlisle, A.B., Chapman, D.D., Chapple, T.K., Chisholm, J., Clarke, C.R., Clua, E.G., Cochran, J.E.M., Crochelet, E.C., Dagorn, L., Daly, R., Cortés, D.D., Doyle, T.K., Drew, M., Duffy, C.A.J., Erikson, T., Espinoza, E., Ferreira, L.C., Ferretti, F., Filmalter, J.D., Fischer, G.C., Fitzpatrick, R., Fontes, J., Forget, F., Fowler, M., Francis, M.P., Gallagher, A.J., Gennari, E., Goldsworthy, S.D., Gollock, M.J., Green, J.R., Gustafson, J.A., Guttridge, T.L., Guzman, H.M., Hammerschlag, N., Harman, L., Hazin, F.H.V., Heard, M., Hearn, A.R., Holdsworth, J.C., Holmes, B.J., Howey, L.A., Hoyos, M., Hueter, R.E., Hussey, N.E., Huveneers, C., Irion, D.T., Jacoby, D.M.P., Jewell, O.J.D., Johnson, R., Jordan, L.K.B., Jorgensen, S.J., Joyce, W., Keating Daly, C.A., Ketchum, J.T., Klimley, A.P., Kock, A.A., Koen, P., Ladino, F., Lana, F.O., Lea, J.S.E., Llewellyn, F., Lyon, W.S., MacDonnell, A., Macena, B.C.L., Marshall, H., McAllister, J.D., McAuley, R., Meyer, M.A., Morris, J.J., Nelson, E.R., Papastamatiou, Y.P., Patterson, T.A., Peñaherrera-Palma, C., Pepperell, J.G., Pierce, S.J., Poisson, F., Quintero, L.M., Richardson, A.J., Rogers, P.J., Rohner, C.A., Rowat, D.R.L., Samoily, M., Semmens, J.M., Sheaves, M., Shillinger, G., Shivji, M., Singh, S., Skomal, G.B., Smale, M.J., Snyders, L.B., Soler, G., Soria, M., Stehfest, K.M., Stevens, J.D., Thorrold, S.R., Tolotti, M.T., Towner, A., Travassos, P., Tyminski, J.P., Vandeperre, F., Vaudo, J.J., Watanabe, Y.Y., Weber, S.B., Wetherbee, B.M., White, T.D., Williams, S., Zárate, P.M., Harcourt, R., Hays, G.C., Meekan, M.G., Thums, M., Irigoien, X., Eguiluz, V.M., Duarte, C.M., Sousa, L.L., Simpson, S.J., Southall, E.J., and Sims, D.W. 2019. Global spatial risk assessment of sharks under the footprint of fisheries. *Nature* **572**(7770): 461–466. doi:10.1038/s41586-019-1444-4.
- Rohner, C.A., Bealey, R., Fulanda, B.M., Everett, J.D., Richardson, A.J., and Pierce, S.J. 2021. Movement ecology of black marlin *Istiompax indica* in the Western Indian Ocean. *J. Fish Biol.* **99**(3): 1044–1059. doi:10.1111/jfb.14809.
- Rohner, C.A., Bealey, R., Fulanda, B.M., and Pierce, S.J. 2020. Movement and habitat use of striped marlin *Kajikia audax* in the Western Indian Ocean. *J. Fish Biol.* **97**(5): 1415–1427. doi:10.1111/jfb.14508.
- Rohner, C.A., Richardson, A.J., Jaine, F.R.A., Bennett, M.B., Weeks, S.J., Cliff, G., Robinson, D.P., Reeve-Arnold, K.E., and Pierce, S.J. 2018. Satellite tagging highlights the importance of productive Mozambican coastal waters to the ecology and conservation of whale sharks. *PeerJ* **6**: e4161. PeerJ Inc. doi:10.7717/peerj.4161.

**Appendix I: Draft template of data request form**

Name of the institution(s) requesting the data	
Contact details for the head researcher	
Project outline	
Specifications of the data required (fields, formats)	
Name(s) and position(s) of the staff accessing the data	
Intentions with respect to publication of the results of the proposed work	
Usage agreement between data owner(s) and user(s)	<ul style="list-style-type: none"><li>• To not diffuse the data to a third person</li><li>• To include the data owner(s) in the analysis and/or interpretation</li><li>• To include data owner(s) as co-authors in publications</li><li>• To delete all data copies once analyses are complete</li></ul>

## Appendix II: Technical description of the G2OI project

The main objective of the INTERREG V project "Grand Observatoire de l'Océan Indien" (G2OI or swIOOS, "South West Indian Ocean Observing Systems") is to enhance the value of research data in the region by bringing together a network of partners who manage reference data in the IOC countries (Reunion, Comoros, Madagascar, Mauritius, Seychelles). The project partners will create and feed a regional data infrastructure that will allow end-users (scientists, managers, general public) to easily exploit the regional data heritage through web services made available on a data portal dedicated to the infrastructure. The selected technical solutions, supported by a series of meetings with partners, will allow to reinforce the national data infrastructures already in place and to exchange information more easily (interoperability) with similar initiatives in other regions of the Indian Ocean and at the global level.

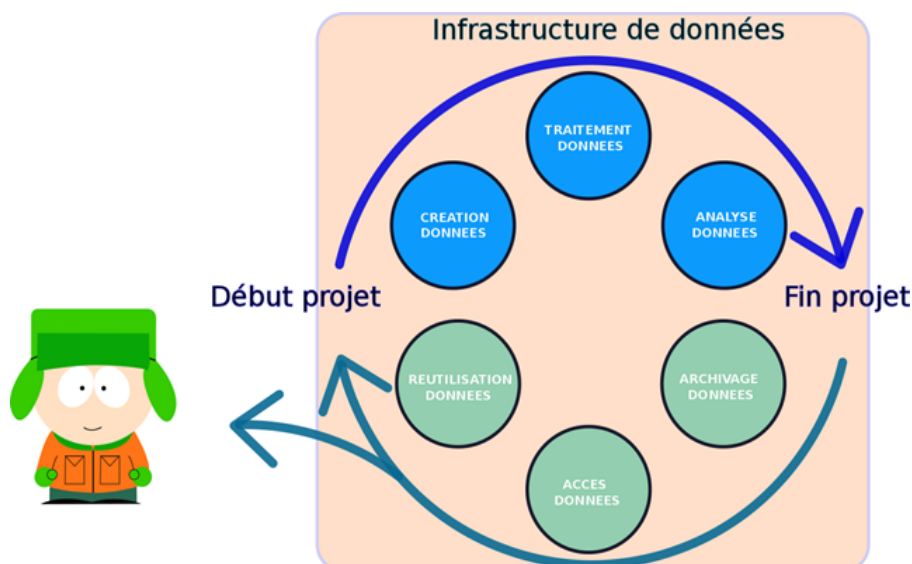
In the south-western Indian Ocean, ERDF and EDF funding has enabled IOC Member States to finance projects in various fields (fisheries, biodiversity, agronomy, health, education, energy, natural hazards, etc.) for over 30 years. In the past, these projects were generally not designed or coordinated to ensure the sustainability of the data and information they have accumulated. As a result, the resulting data heritage, which is the main legacy of this significant economic investment, is at risk. Indeed, much of the data is dispersed, remains difficult to access and has not been valorised once the projects have been completed. If better managed and made accessible to the general public as well as to development and research actors, this heritage can help improve the sustainable management of regional resources and support the development of new projects. In particular, good data management makes it possible to define reference states and study changes in many areas (e.g., exploitation of regional resources, health status of coral reefs, etc.) but it also makes it possible to answer new questions and create added value by combining different types of data that were previously managed separately. However, as it stands, potential users are unable to archive, discover, reuse and therefore add value to most of the data that exists in the region. These problems are historically linked to the heterogeneity of data collection and management methods in the projects and among the partners, and have been amplified by the absence of shared regional solutions to guarantee basic services on the region's data (archiving and referencing, search, access, processing). The G2OI project proposes a solution to these recurring problems by coordinating the implementation of a regional data infrastructure that will reinforce the national initiatives of the partners. Below, we detail the major challenges of data infrastructures at global and regional levels and then explain how the G2OI project will build a regional data infrastructure that will meet user needs by mobilising and strengthening a network of partners in the region.

### 1/ Global and regional contexts: setting up data infrastructures

From a global point of view, the economy and science of data are growing rapidly. On the one hand, the largest private companies (such as Google, IBM, Amazon, Apple, Microsoft, Facebook) are adding value to the data they collect or reference through various services (search engine, archiving, access, processing). On the other hand, governments are integrating Artificial Intelligence and data management into their development strategy (e.g., Open Data and the European Directive INSPIRE, Infrastructure for Spatial Information in Europe). Ultimately, user communities (public or private) expect better data, better decisions in all fields (health, ecology, natural hazards, etc.) and at different scales (local, regional or global). This technological revolution is based on the exponential increase in data acquisition capacities (sensor networks and connected objects, Internet of Things, etc.) combined with the emergence of new techniques for data management and analysis (Big Data) from the world of research (e.g., neural networks for deep learning). In addition to specific human skills, the development of these activities relies entirely on the implementation of data infrastructures that bring together the hardware components (machines stored in data centres) on which the various software components (servers and applications) that provide the operational services needed to manage the data (archiving, discovery, access and data processing services) are hosted. In addition, the value of the data managed in these infrastructures depends on the following key points:

- the availability of long time series of data: indeed, observations, repeated over long periods of time, are essential to provide reference states on systems and to understand their evolution in many fields (fisheries, coral reef management, effects of climate change, health, food security, etc.);
- the good management of the spatial dimension of data is particularly important as geographic data federate different user communities (spatial data are essential in multidisciplinary contexts) and improve the handling of data services (in particular search and data visualisation);
- the standardisation of data formats and protocols for accessing them conditions the interaction capacity (interoperability) of an infrastructure with the information systems of other partners. For example, in the European context, the INSPIRE directive obliges its member states to follow international standards in order to coordinate the implementation of a Spatial Data Infrastructure (SDI) at the European level. Thus, INSPIRE-compatible infrastructures can communicate with each other and can also feed other systems (e.g., governmental Open Data portals);
- Public and private organisations, aware of these strategic issues, are now streamlining the management of their data by adopting data management plans (**Figure 1**) which clarify the life cycle of the data they are funding and identify which infrastructures will be used to manage them once the projects are completed.

This INTERREG project proposal will provide a regional response to these global challenges by promoting the establishment of a data infrastructure for the South West Indian Ocean.



**Figure 1:** Main services of an infrastructure to implement a data management plan

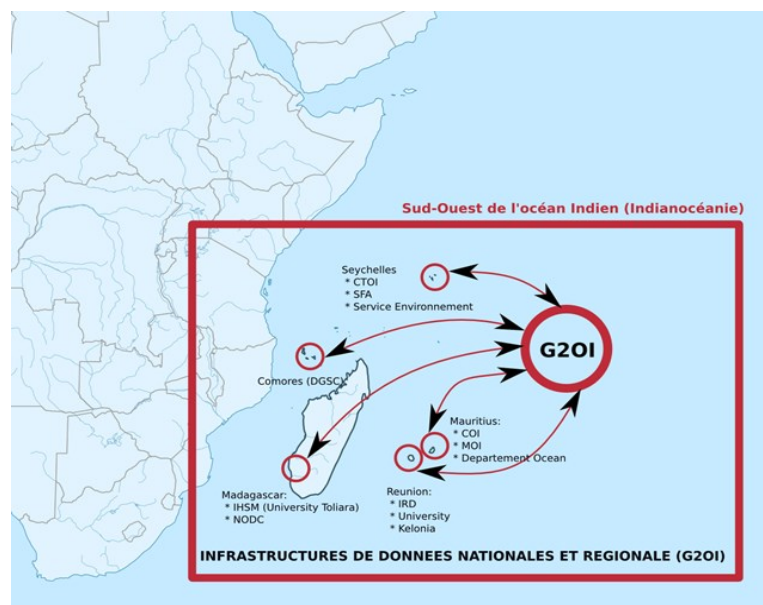
In the south-west Indian Ocean, ERDF or EDF projects have thus made it possible, for more than 30 years, to accumulate data and knowledge in various fields. Despite considerable investments, the data heritage resulting from these projects remains poorly managed. It is currently difficult to know what data exists in the region and even more complicated to mobilise it outside the context of its acquisition. The identification and description of this regional information heritage as well as the archiving of the identified data are necessary actions to secure and valorise these investments in other contexts (by other projects or other user communities). To this end, it is necessary to mobilise partners in the region to coordinate the establishment of a regional spatial data infrastructure. Such an infrastructure will allow, on the one hand, to better reference the data heritage from past projects and to implement data management plans for future projects in IOC countries (see Figure 1). On the other hand, it will allow the coordination and consolidation of spatial data infrastructures managed by different research actors in Reunion or by their partners in the other IOC member states. Indeed, a regional infrastructure will bring additional expertise and visibility to partners producing data in the South West Indian Ocean and will also create links with international partners working on similar initiatives in neighbouring regions. For example, an infrastructure in the Southwest Indian Ocean will be a significant contribution



to the Global Ocean Observing System (GOOS) by complementing similar infrastructures in neighbouring regions of the Indian Ocean (to the East with IMOS in Australia and to the South with SOOS in the Antarctic, IORA, etc.) or similar initiatives in other regions (PacIOOS which brings together small island states in the Pacific).

2/ The data infrastructure of the G2OI project will meet the needs of users by relying on a network of partners in the region

The G2OI project will strengthen collaboration in Reunion and between the IOC countries through the creation and sharing of a spatial data infrastructure in the south-west Indian Ocean (**Figure 2**). This infrastructure will make it possible to coordinate the management and use of data and knowledge produced by various partners and research projects in the region (past, current or future projects). Beyond the data and servers, the stakes are human: the organisation of events (workshops, seminars, conferences, etc.) will strengthen the regional network of actors involved in data management. It is a question of the partners taking ownership of the project's objectives and acquiring the necessary skills to achieve them (learning new methods and reference tools for spatial data infrastructures). Initially, the project team will work in conjunction with the partners' data managers. Meetings organised by the project will allow the partners to share their experiences and to acquire new technical skills to better administer and feed the spatial data infrastructures (mastery of software components and different data sources). The network of partners will thus be able to interact more effectively with the project team and guide it in identifying, integrating and sustaining the reference data for the region in the infrastructure. Secondly, end-users will be able to easily exploit the data managed in the infrastructure by using the operational services that will be made available on the project's data portal (archiving, search, access and data processing services). Finally, the work will be regularly presented and discussed in workshops and conferences in order to allow new partners and future projects in the region to exploit, feed and sustain the infrastructure beyond the G2OI project (e.g., projects in Reunion funded by ERDF, ANR, Region and IOC projects funded by EDF or AFD).



**Figure 2:** Building infrastructure by building on and strengthening partners' national infrastructure

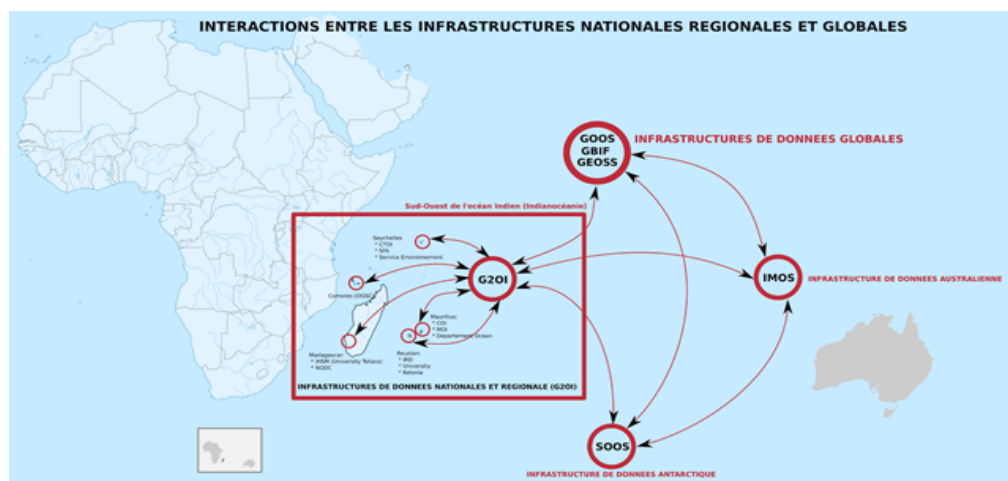
In Réunion, various actors (University, OSU-R, UMR ESPACE-DEV and SEAS-OI, ENTROPIE, MARBEC, Kelonia, CIRAD, IFREMER, etc.) have already set up applications that could feed this regional data infrastructure. Indeed, these initiatives generally comply with the recommendations of the European INSPIRE (Infrastructure for spatial information in Europe) directive, which guarantees their technical interoperability with other applications. From an institutional point of view, past projects have also made it possible to bring together and coordinate these separately initiated but complementary initiatives by federating actors around common themes (e.g., the ERDF project on Terrestrial Biodiversity and Marine Megafauna). For complex fields such as the management of marine ecosystems or Marine

Spatial Planning, it is essential to continue this work in order to federate the actors in charge of reference data in different fields (fisheries, aquaculture, land use, natural risks, energy, health, etc.). The data infrastructure set up by the G2OI project will improve the coordination of Reunion's stakeholders and can be shared in different contexts (e.g., future research projects).

In the other IOC countries, the G2OI project will benefit from the experience of reference partners in key areas (risk management, land use planning, coastal and marine ecosystem management, energy, oceanography, health, etc.). Some partners already manage their data in national data infrastructures and the coordination of their activities to set up a regional infrastructure will allow them to strengthen their skills and their own data management systems. The G2OI project will also enable the implementation of strategies (data management plans) for planning data management with regional organisations (e.g., with IOC or IOTC). In this way, the regional infrastructure can then be used and shared by future IOC projects (in particular under the 11th EDF) with new partners in the region who wish to streamline their data management in broader areas.

In conclusion, the G2OI project represents an opportunity to meet the needs of different user communities:

- the needs of partners who manage data in the region and who wish to enrich their skills and their own systems. The G2OI project will address these needs by organising regular meetings that will allow partners and the project team to collaborate and share methods for creating and enriching the regional data infrastructure;
- the needs of regional organisations and donors who fund a significant part of the data produced in the region and now wish to ensure the sustainability of investments by integrating management plans into future projects in the region (e.g., to coordinate the management of data from EDF and ERDF projects). These data management plans can be implemented using the G2OI regional data infrastructure;
- the needs of users who wish to access the regional data heritage but who do not yet have the necessary tools to exploit it. The data portal set up by the G2OI project will provide the operational services to easily exploit the data and knowledge of the infrastructure (archiving, searching, accessing and processing data);
- the needs of broader initiatives that aim to connect regional infrastructures to each other to make them visible on a global scale. These initiatives will be able to connect to the infrastructure set up by the G2OI project to access and disseminate more widely the information made available by the various partners in the South-West Indian Ocean (**Figure 3**).



**Figure 3:** Interchange (interoperability) with other regional and global infrastructures

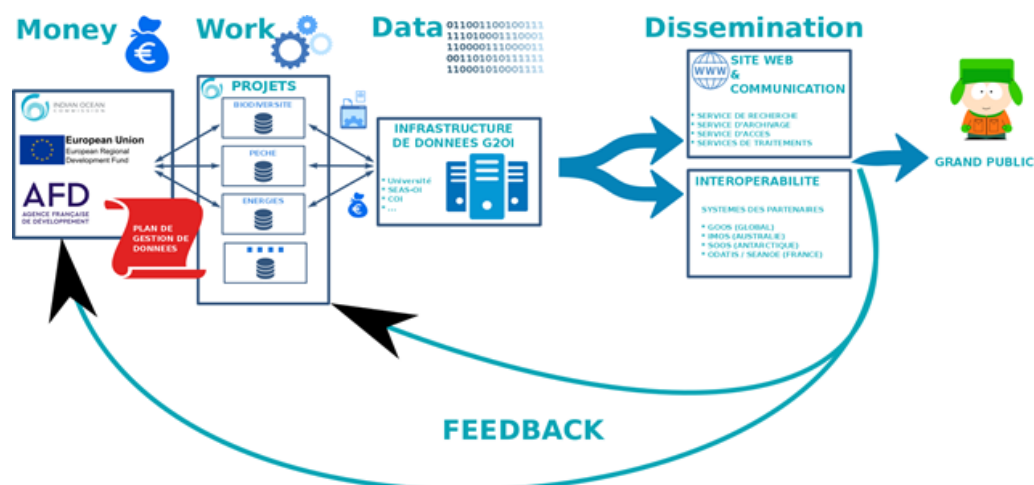
### Objectives

The data infrastructure created by the G2OI project should strengthen the regional dynamics for the conservation and valorisation of research data.

The first step will be to simplify and improve the methods used by the partners to administer the servers and applications with which they manage their data (and which will be connected to the regional infrastructure). This will be done both by optimising the technical solutions proposed and by acquiring the associated skills in workshops organised by the project. On this optimised hardware and software basis, the next step will be to coordinate data management. In particular, an important objective for the partners will be to better describe and archive the regional data heritage in the infrastructure. A large part of this heritage is the result of past projects in which the region's partners have been regularly involved (past or current projects: EDF, ERDF, AFD, etc.). In addition, the use of infrastructure services will be encouraged to manage the data that will be produced by future projects (in particular ERDF and IOC EDF projects) and thus guarantee their valorisation and durability. For future projects, this will concretely consist in adopting, before they are officially launched, data management plans (**Figure 1**) which will detail how their data will be managed and integrated into the regional infrastructure (referenced and archived). The challenge here is to show that the initial investment will pay off as the data can be reused once the projects are completed.

In a second phase, a data portal will be set up. It will be the public entry point that will direct external users to the operational services that allow the data contained in the regional infrastructure to be explored and exploited (archiving, discovery, access and data processing services). The data portal will show how the use of the regional infrastructure allows for better coordination of the national data management systems that comprise it. The data portal will integrate a classical project website to communicate also classical information (project activities, partners involved, etc.).

Finally, links will be established with similar international initiatives in neighbouring regions through events (seminars or conferences) organised by the project or the scientific community.



**Figure 4:** Data flow and valorisation through the G2OI data infrastructure and the associated web portal

These general objectives are illustrated in Figure 4 (above) and detailed in the points below:

- Objective 1.** To create an infrastructure based on optimised technical solutions: the regional infrastructure should not force partners to abandon their systems by imposing a centralised (monolithic) solution. On the contrary, it should allow partners to connect their applications easily and to enrich them. Following the example of the European approach (INSPIRE directive), the objective is to rely on a distributed architecture composed of national infrastructures located in the region and which will be strengthened by this collaboration (from a hardware, software and human point of view). The key point is to ensure the interoperability of the systems in place in the partners of the region by adopting the global standards and IT tools of reference. However, getting to grips with these methods and tools can be complex. Depending on the needs of the partners, workshops are planned to enrich the knowledge and technical skills needed to use the standards and

services provided by the infrastructure. Different options will be proposed according to the profile of the partners (not exclusive between them):

- *Option 1:* Partners can create or enhance their own systems so that they become distributed components that contribute to the regional infrastructure. This option does not pose any technical problems as long as the actors coordinate to allow a certain level of interoperability between their systems. To achieve this goal, the project will propose that partners administer standardised applications that share the same configurations and can be easily replicated on different servers in the region. Workshops will be organised to learn how to use these technical solutions (sharing virtualised applications) and to deploy them on servers.
  - *Option 2:* Partners use the common (shared) services and tools offered by the regional infrastructure to manage their new data sources. This option is relevant for actors who do not have the means (economic, human) or reasons to administer their own system in the long term (e.g., IOC projects that need to ensure long-term archiving and preservation of their data once they end). Depending on the needs, this option may eventually require the reinforcement of the physical infrastructure in place in the region (storage capacity, racks or machines to deploy new applications, data or processing servers) and the occasional involvement of project staff.
- **Objective 2.** To feed the infrastructure with data heritage from past ("Data Rescue") or ongoing projects: during the G2OI project, the partners will identify data (historical data or data being acquired in other projects) that can be preserved on the regional infrastructure servers (archived and described by the G2OI project team using harmonised metadata) In particular, IOC projects (10th EDF: SMARTFISH, BIODIVERSITY, ENERGIES) have funded activities that will feed the infrastructure with data from these projects. Depending on the constraints associated with these data, they can then be transformed and made accessible through standardised formats and access protocols. The types of data to be described and integrated into the infrastructure concern as a priority
    - spatial remote sensing data: earth observations (satellite, airborne, UAV data) and spatial oceanography (temperature, chlorophyll, wind, currents, salinity, etc.),
    - *in situ* sensor data (from different observation networks, platforms or stations): sensor networks (e.g., weather stations, Maïdo observatories, Piton de la Fournaise), observation resources at sea (buoys, oceanographic campaigns by scientific vessels: e.g., ANTEA or Marion Dufresne, etc.)
    - model data: oceanographic models (ROMS, PISCES, Drakkar, MARS-3D, etc.), atmospheric models, digital terrain models,
    - Field data: marine resource management (coral reefs, coastal or deep-sea fishing, aquaculture, marine biodiversity: cetaceans, turtles, etc.), other themes related to IOC's fields of interest (climate change, natural hazards, health, energy, food security, etc.), reference cartographic data (land use, urban mapping, etc.).
  - **Objective 3.** To feed and sustain the infrastructure by pooling it to manage data from future IOC ERDF and EDF projects: before the end of the INTERREG G2OI project, an important issue is to ensure that the data infrastructure can be fed with data from future ERDF and EDF projects that regularly involve IOC partners (in particular under the 11th EDF). For example, particular attention will be paid to EDF IOC projects that will be launched during the G2OI project (future projects on fisheries, biodiversity, health monitoring, risks, climate change, food security...) and that will be able to enrich the infrastructure in real time. To achieve this objective, future project proposals will integrate a Data Management Plan that can be implemented using the infrastructure services. Each project will thus be able to make a minimal financial contribution to benefit from these shared services (by providing a budget that covers part of the costs linked to the use of the infrastructure: participation in the manpower and maintenance of the material resources necessary to offer the services of

referencing, archiving, accessing or processing data on the servers of the regional infrastructure, cf. option 2 of objective 1).

- **Objective 4.** To set up a regional data portal that provides users with operational services to exploit the data managed in the infrastructure: the infrastructure must meet the needs of the Reunionese partners and the IOC countries to facilitate the management and valorisation of the region's data assets through a set of basic services. These services will be set up by the engineers recruited for the project (fixed-term contracts or services) and will all be accessible on the project's data portal (they may also be available on the data portals of partners who manage the required level of interoperability). The implementation of new services will be systematically accompanied by workshops that will allow the project partners to acquire the necessary skills to master these services and make them evolve. In particular, the following services will be offered to end users (scientists, managers, general public):
  - Permanent archiving: users will be able to deposit data on the portal with the guarantee that their data will not be lost (even if they are not accessible to everyone) and that they can be cited by the works that re-use them (allocation of unique identifiers: DOIs). To this end, an archiving service will be created at the University of Réunion,
  - Research: discover and locate data from the infrastructure that are relevant to a given theme or use. To this end, the regional data portal will include a search engine that will federate (enrich and interconnect) the various metadata catalogues of the partners,
  - Access: simplify the handling and interoperability of data by using standardised data formats and access protocols that guarantee reuse by the greatest number of people,
  - Data processing: automate the provision of processing to deliver routine products of interest to different domains (supporting sustainable fisheries, habitat detection, land use, etc.).
  
- **Objective 5.** Communication strategy to strengthen links with other actors, regional and international projects, inside and outside the IOC region: beyond the national partners and IOC projects that will be formally involved in the G2OI project, the aim is to promote the use of the data infrastructure to connect complementary data sources, created by other partners, projects or observation networks. To this end, the project will organise or participate in events (workshops, conferences) during which the infrastructure and the project results will be presented and discussed. For example, in Reunion, contacts could be made with initiatives that produce or consume data (citizen science projects, the Blue Institute, the marine herbarium network, the Life+ petrels project) and other types of projects (ANR, Region) or national initiatives (Odatis, the future information system on the marine environment, etc.) Beyond the IOC member countries in the south-west Indian Ocean, the aim will be to encourage the use and coordination of this initiative by donors and international organisations involved in the region (WIOMSA, FAO/IOTC/SWIOFC, UNEP, IUCN, AFD, FFEM, etc.). More broadly, the infrastructure will be designed so that it can also connect and exchange some of its content with similar initiatives carried out by neighbouring regions (ioGOOS in the Indian Ocean, IMOS in Australia, SOOS in Antarctica, PacIOOS for the Pacific island states, etc.) and other international infrastructures (GOOS, GBIF, GEOSS, Copernicus, etc.). In particular, a workshop dedicated to data infrastructures will be organised during the next WIOMSA symposium in 2019 to bring together partners interested in data management in the Indian Ocean.

### Expected results

The main expected result is to enhance the value of research data in the South West Indian Ocean region. To achieve this, it is necessary to create efficient services to facilitate the archiving, access and processing of data from partners

in the region. Indeed, at present, it is complicated for an ordinary user to know what data exists in the region and how to access it. To meet these needs, the project will build on the following results:

- Coordination of a regional data infrastructure that uses and enhances existing partner systems (existing partner applications and data, integration of new applications and data)
- Creation of a data portal to exploit the services offered by the regional data infrastructure:
  - Regional data discovery service to direct users to partner systems
  - Archiving service to allow partners who wish to do so to retain data and obtain DOIs to cite it when it is reused
  - Access (depending on rights) and processing service
  - Promotion of the portal to partners in the region, in particular by regional organisations (e.g., IOC, IOTC, FAO, World Bank, etc.)
- Provision of data management plans to facilitate the use of the data infrastructure in future projects:
  - Adoption of the data management plan in future ERDF or EDF (IOC) projects to maintain and evolve the infrastructure,
  - Use of the infrastructure in new regional projects (e.g., citizen science projects).

The G2OI project will be able to partly quantify these results by measuring:

- the number of hits and downloads on the project data portal,
- the number of new partners or projects that have contributed to the infrastructure
- the number of projects that use the regional infrastructure to implement their data management plans
- the number of citations of the infrastructure data (through DOIs) in scientific or expert papers.