

Assigning DOIs to publically-accessible IOTC documents and their publication on the open-access data repository Zenodo

Anne-Elise Nieblas¹, Fabio Fiorellato², Emmanuel Blondel³, Paul DeBruyn², Julien Barde⁴

¹ IRD, La Reunion, France.

² IOTC, Victoria, Seychelles.

³ FAO, Rome, Italy.

⁴ UMR MARBEC, IRD, Sète, France.

Abstract

DOIs are persistent, unique identifiers that are increasingly becoming a standard method to help identify research objects. They give a digital ‘footprint’ that follows the research object throughout its lifespan, regardless of where it is physically hosted. A workshop was held in April 2019 at the Indian Ocean Tuna Commission (IOTC) Secretariat in Victoria, Seychelles called by the Institut de Recherche pour le Développement (IRD) as part of the “Fisheries and aquaculture” use case of the European OpenAire-Connect H2020 project. The aim of this ‘use case’ is to publish the publicly-accessible IOTC documents that are of scientific interest to the “Fisheries and aquaculture” community on the open-access data repository, Zenodo (i.e. assign digital object identifiers, or DOIs to IOTC documents). [Zenodo](#) is an open-access data repository developed under the [OpenAIRE](#) project, operated by [CERN](#), and supported by the European Commission, which assigns DOIs to research objects free of charge. The product focus of Zenodo is on datasets, codes (e.g. linked via Github repositories), publications, presentations (i.e. slides), and other research objects. A workflow was established to 1) filtering IOTC documents according to an agreed upon selection criteria; 2) validate the selected documents by the IOTC Secretariat; 3) publish the documents on Zenodo/Zenodo Sandbox and obtain a DOI for each; and finally 4) publish the assigned DOIs on the IOTC website. Of the 6124 documents derived from the IOTC website and defined as an appropriate document type for the scientific community, 3314 documents were retained for subsequent publication on Zenodo following the selection criteria. Validation of the documents was limited due to time constraints of the Secretariat and the duration of the OpenAire project (ended the 30th of June 2019). Within the timeframe of the project, 601 documents, including meeting reports and executive summaries, were published on Zenodo (~18% of the total filtered documents). The short-term objective of this project remains the complete validation by the IOTC and publication on Zenodo of the documents identified by this project as appropriate for the scientific community. The long-term objective of this project is to publish IOTC documents each year, following a standard protocol. The protocol and codes for the filtering and publication steps are complete and published (R code available at [aenieblas/Zenodo_DOI_IOTC](#)). The objective of this paper is to introduce working party participants to the concept of DOI, the benefits of DOIs and the workflow of assigning DOIs to future IOTC documents of working parties and meetings. A next step is to discuss a strategy to continue publication of IOTC documents on Zenodo in the future.

Introduction

DOIs are persistent, unique identifiers that are increasingly becoming a standard method to help identify research objects. They give a digital ‘footprint’ that follows the research object throughout its lifespan, regardless of where it is physically hosted. Metadata are also increasingly being tied to DOIs, and can be used to export standardized bibliographic references. A DOI allows a research object to be

accurately identified, and reliably links the work to its authors. DOIs are also meant to increase the reach and impact of the research object as well as its citation.

DOIs are evolving with end-user requirements/desires toward reproducible and open science. DOIs are generally given to documents, such as publications and presentations, but they are also now given to datasets, code, parameters, queries (e.g. [GBIF](#) gives DOIs for database queries, a method that enables a completely reproducible scientific process).

Zenodo is an open-access data repository developed under the [OpenAIRE](#) project, operated by [CERN](#), and supported by the European Commission, which assigns DOIs to research objects free of charge. The product focus of Zenodo is on datasets, codes (e.g. linked via Github repositories), publications, presentations (i.e. slides), and other research objects. For each submission of a product, a DOI is created. Zenodo is a trustworthy publication and sharing mechanism for scholarly activities. It promotes open licencing, provides tracking statistics of publication views and downloads, and its contents are harvested by large search engines, including Google services like [Google Dataset Search](#).

Submissions to Zenodo will be retained for at least 20 years, which is the length of time that CERN is currently funded. DOIs are assigned to each published file and cannot be changed (some flexibility is given depending on a case-by-case basis). There are some mandatory metadata elements (based on [Datacite schema](#)) required prior to publication, but metadata can also be updated after publication.

For developers, Zenodo provides services via a REST API (<https://developers.zenodo.org/>) to enable programmatic access. Zenodo also provides a service called “Zenodo Sandbox”, which is a mirror of Zenodo that allows authors and developers to practice publishing their research object prior to final publication and assignment of a permanent DOI.

Though DOIs are permanent, Zenodo provides the option to have a ‘version’ DOIs, i.e. a permanent and original ‘concept’ DOI is assigned to a document, and any subsequent versions of that document will have a unique ‘version’ DOI that is linked to the original ‘concept’ DOI ([Figure 1](#) for an example). This allows for a living scholarly record that can be updated, curated, and exchanged. This versioning mechanism can ensure, for example, snapshots of a time series or identifying releases of a software and, by doing so, sustain reproducible research.

The screenshot shows the Zenodo interface for a research object. At the top, the Zenodo logo is on the left, a search bar in the center, and 'Upload' and 'Communities' links on the right. A user profile for 'anne.elise.nieblas@gmail.com' is visible in the top right. The main content area displays the title 'geometa: Tools for Reading and Writing ISO/OGC Geographic Metadata in R' by Emmanuel Blondel, dated December 17, 2018. It includes a 'Files' section with a table listing 'geometa_0.4-0.tar.gz' (315.3 kB) and a 'Citations' section showing 'No citations.' On the right side, there are statistics for 97 views and 16 downloads, an 'Indexed in' section for OpenAIRE, a 'Publication date' of December 17, 2018, a DOI of 10.5281/zenodo.2371217, and a list of keywords including 'R', 'ISO/TC211', 'OGC', 'metadata', 'feature catalog', 'geographic', 'geospatial', and 'spatial'. The license is listed as MIT License. At the bottom right, a 'Versions' table lists four previous versions: Version 0.4 (Dec 17, 2018), Version 0.3 (Aug 22, 2018), Version 0.2 (Nov 28, 2017), and Version 0.1 (Jun 1, 2017).

Figure 1. An example of a research object with several versions (see bottom right corner for listing of the different versions available).

Zenodo also allows for research objects to be organised by “Zenodo communities”, e.g., a collaborative space for all research output of a domain, journal, project, workshop, institute, etc. Communities can assign a range of licenses, including Creative Commons licences, and a DOI to each product. Communities can be curated by multiple users to group the research object created by members of that community, e.g., [Biodiversity Literature Repository](#), [The 19th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun](#).

In terms of Zenodo policies, all fields of research and types of research artifacts are accepted. The content of the product must not violate privacy or copyright, or breach confidentiality or non-disclosure agreements for data collected from human subjects. Research data are accepted at any stage of the research lifecycle. Anyone may register and deposit content for which they possess the appropriate rights. By uploading content, no change of ownership is implied and no property rights are transferred to CERN. Finally, all data file formats allowed and each deposit can be up to 50Gb

A workshop was held in April 2019 at the Indian Ocean Tuna Commission (IOTC) Secretariat in Victoria, Seychelles called by the Institut de Recherche pour le Développement (IRD) as part of the “Fisheries and aquaculture” use case of the European OpenAire-connect H2020 project ([Appendix 1](#) for agenda). The aim of this ‘use case’ is to publish the publicly-accessible IOTC documents that are of scientific

interest to the “Fisheries and aquaculture” community on the open-access data repository, Zenodo (i.e. assign digital object identifiers, or DOIs to IOTC documents). The workshop aimed to define the various steps of the workflow to achieve the publication of IOTC documents: identify selection criteria to filter the IOTC documents that are appropriate for publication, obtain the metadata and actual files of the documents from the IOTC, enrichment of metadata and to discuss a strategy to continue publication of IOTC documents on Zenodo in the future ([Appendix 2](#) for workshop presentation).

The specific objectives of this project were to :

- 1) Define criteria to filter appropriate documents
- 2) ‘Reserve’ DOIs (on Zenodo)
- 3) IOTC to validate and publish reserved DOIs on IOTC website
- 4) We can then publish the DOIs on Zenodo
- 5) Enrich with improved metadata

Methods

A workflow was established following these steps:

- 1) Filtering of IOTC documents by the IRD team according to the agreed upon selection criteria:
 - Identification of IOTC documents;
 - Enrichment of metadata;
- 2) Validation of the selected documents by the IOTC Secretariat,
 - Verification that each document proposed for a DOI by IRD is appropriate;
 - Not previously been assigned a DOI;
 - IOTC to send back the list of validated documents;
- 3) Publication on Zenodo (first try on Zenodo Sandbox);
- 4) IOTC publishes the final DOIs on the IOTC website.

The workflow uses R programming language. It adapts scripts of the R [geoflow](#) package ([Blondel et al. 2019](#)), and uses the [zen4R](#) ([Blondel and Barde 2019](#)) as the R package to interact with Zenodo from R (binding for the underlying Zenodo API). The final codes are open source and published online to ensure future updates. These codes are available via GitHub (https://github.com/aenieblas/Zenodo_DOI_IOTC).

1) Define selection criteria to filter appropriate documents

During the workshop, the IOTC Secretariat helped establish selection criteria to filter documents that should be assigned DOIs. These were first filtered by IOTC ‘Document type’ as documents are organised on their website ([Table 1](#)), then by IOTC ‘Meeting’ ([Table 2](#)). Further filtering removed duplicates, presentations, information papers, ad hoc papers and withdrawn papers. Finally, individual files within each document type and meeting were verified by both IRD and IOTC ([Table 1](#)). A full protocol of the selection criteria can be found in [Appendix 3](#).

IOTC Document type

Documents types that are of scientific interest were identified by the IOTC, and given a high priority. Several document types have little or no interest to the broader community, and include, for example, logistical or financial information (i.e. meeting information; [Table 1](#)). Reference documents and information papers were not included as they are documents that are not discussed during the working parties and they are often already published and thus already have a DOI. Guidelines and publications were given medium priority as they may contain information of interest to the scientific community; for example, species guides; but it should be ensured that these documents do not already have a DOI associated. Though in theory, stock assessment input and output are highly relevant to the community, they were given a “future” priority as these files are not adequately organised to ensure that a publication would be useful.

National reports are considered a high priority: in the past, national reports were classified as Information papers (given the ‘INF’ code in their reference name; see Section 6 for more information about reference naming conventions) under ‘Meeting documents’, and special care should be taken to extract these from the ‘Meeting documents’ document type. In addition, the IOTC Secretariat may be listed as an author for these reports, but in fact, the author should be the country. Similarly, executive summaries and expert consultations were previously classified as “Meeting documents”, and attention should be given to extract these files.

Table 1. Selection criteria of IOTC resources by ‘Document type’ and individual files. Selection criteria identified based on IOTC priorities, i.e. keep all documents types relating to ‘high’ priority rating, and apply selection criteria on file types as described in the Selection criteria column. Priorities are color-coded such that high priorities are green, medium priorities are orange and low priorities/removals are red. Future priorities are in purple and indicate that these document types are not currently at the stage where clear documents exist or selection criteria can be easily applied.

IOTC document type	Number of files	Selection criteria for files	OpenAIRE Priority	IOTC priority
Meeting Report	396	Keep latest document by language, Remove low priority meetings	medium/high	high
Meeting information	170		low	no
Meeting documents	4421	Remove: low priority meetings, some titles relevant only for Secretariat (e.g. agenda), INFO papers Verify/reclassify: National reports, executive summaries, expert consultations	high	high
Reference Documents	36		low	no
Compliance reports			low	no
Compliance questionnaires			low	no
Response to feedback letters			low	no

Implementation reports			low	no
Executive Summaries	306	Remove: low priority meetings	medium	high
Information papers	467		Low, depending	no
Datasets	457	Remove: Data catalogue, dataset availability document, SF REF / CE REF, stock assessment data from Secretariat, combined zip of separate files, CPUE from Secretariat, parameter files, code Keep: nominal catch, separate catch and effort files, size frequency Future: equations, data forms	high	high
Stock Assessment Input and Output files	6	Note: Mixed into datasets; Metadata not well organised at present	high	future
Project reports	0			high
National reports	165	Note: author=country		high
Guidelines	6	Keep: species guides. Check that no DOI previously assigned		low
Publications	33	Keep: species guides. Check that no DOI previously assigned	medium	medium
Inspection reports			low	no
General			low	no
Letters of credentials (all)			low	no
CMM proposals			low	no
CNCP applications			low	no
Reports from other meetings			low	no
NGO statements		Already in meeting reports		no
Consultant reports	7		medium	no
FAO Documents	0		low	no

IOTC File type

The “Selection criteria for files” column of [Table 1](#) gives details about filters used to select the individual files within the IOTC Document types. In addition to these specific rules, the Secretariat notes that the first eight documents of each meeting can often be ignored as they relate to documents of interest only to the Secretariat (e.g. draft agendas). Working party documents numbered 5 and 7 are sometimes useful, and cannot be deleted outright (i.e. they can be conservation management measures or data). It was determined that text mining of the file title may provide the most useful

filter for these documents as the meeting document number of the file name is not always a good indicator of their usefulness to the scientific community

Within each IOTC document type, individual files are filtered to remove low priority meetings (see below, [Table 2](#)). Filtering the individual files within the Datasets document type requires particular attention. Data catalogs are removed, as are the data reference file, whereby several different individual catch and effort files are zipped into one. The separate catch and effort files are conserved. At present, equations and data forms are removed as they will be updated in the near future. CPUE data and stock assessment inputs are removed as they are information summarised by the Secretariat. The raw catch and effort data are more informative. The stock assessment inputs are what are provided by the IOTC to the modeler, and are not the stock assessment outputs.

IOTC Meeting type

IOTC Documents on the IOTC website are organised by Meeting. Many of these meetings are of interest to the greater scientific community, and some are of interest only to the Secretariat (e.g., SCAF; [Table 2](#), line 4). The Secretariat identified the IOTC meetings with the highest priority for science. The low priority meetings were used as a filter to remove files that should not be assigned a DOI.

Table 2. Selection criteria of IOTC resources by focusing on each specific IOTC meeting. Selection criteria were identified based on IOTC priorities, i.e. to filter IOTC files to remove meetings of low scientific priority. Priorities are color-coded such that high priorities are green, medium priorities are orange and low priorities/removals are red.

IOTC Meeting	OpenAIRE priority	IOTC priority
Commission		low
Special session of the commission		no
Compliance committee (CoC)		no
Standing committee on administration and finance (SCAF)		no
Performance review of the IOTC (PRIOTC)		no
Technical committee on allocation criteria (TCAC)	medium	medium
Technical committee on performance review (TCPR)		no
Management procedures dialogue (MPD)		no
Technical committee on management procedures (TCMP)		medium
Scientific committee (SC)	high/medium	high
Working party on billfish (WPB)	high	high
Working party on tropical tunas (WPTT)	high	high
Working party on neritic tunas (WPNT)	high	high
Working party on temperate tunas (WPTmT)	high	high
IOTC Working party on the implementation of Conservation and Management Measures (WPICMM)	medium	medium

Working party on data collection and statistics (WPDCS)	high	high
Working party on ecosystem and bycatch (WPEB)	high	high
Working party on methods (WPM)	high	high
Working party on fishing capacity (WPFC)	high	low
Working party on tagging (WPT)	high	low
Working party on tagging data analysis (WPTDA)	high	low
IOTC ad hoc Working Group on FADs (WGFAD)	high/medium	Medium
Technical meeting on the IOTC Regional Observer Scheme		low
Management Strategy Evaluation workshop (MSE)	high/medium	High (available?)
Workshop on Connecting the IOTC science and management processes		low
Regional workshop to support compliance with IOTC requirements for the collection and reporting of fisheries data to the IOTC		low
Regional workshop to support the implementation of Resolutions of the IOTC		low
Consultation/validation workshop on the development of e-MARIS		low
Workshop on CPUE standardisation	high	Low (for now sensitive docs)
Science/Stock assessment workshop	high	no
Workshops		no
Indian Ocean shark year program (IO-ShYP) workshop	high	no
Weight of evidence workshop	high?	no
Workshop on depredation in tuna longline fisheries of the Indian Ocean (DeWS)	high	no
Indian Ocean Tuna Tagging symposium	high	high
Indian Ocean Coastal states meeting		no
High seas boarding and inspection scheme (HSBIS)		no
Regional training on the IOTC e-PSM application (E-PSM)		no

Languages

Languages are generally English and French. Of files with the same reference name, the most recent documents by language are kept. Each translation of the same document is given a different DOI. We decided that the intention of the Zenodo option to allow for different versions of the same document was more for updates of datasets or revisions of the same document, rather than for translations. In addition, we believe that giving different DOIs for the different translations ensures that both languages' metadata will be displayed; thus enhancing document and data discovery. Furthermore,

the IOTC website is currently organised such that each translation has its own page, so giving each translation its own DOI follows suit. Finally, DOIs can be linked on Zenodo via their metadata, which can list the reference name of all translations, thereby relating the documents and the search results.

Licensing

The documents that have already been published on the IOTC website are in the public domain, and as yet have no license attributed to them. A license is required for each document published on Zenodo. Thus, the Secretariat attributed the [“Creative Commons Attribution Non Commercial 4.0 International”](#) license to all IOTC documents in this project. This license is recommended by the FAO Fisheries and Research Monitoring System (FIRMS).

Reference naming conventions

The standard naming convention for the reference name given to IOTC documents should follow:

IOTC-YYYY-WPXX#-DOC#

Where “YYYY” is the 4-digit year, “WPXX#” indicates the working party (WP), the meeting acronym (XX) and the meeting number (#), “DOC#” indicates the document number assigned to each meeting document. The separator between elements should be the short dash “-”.

Meeting reports and scientific committee reports are potentially referred to with two different reference numbers, depending on the language. The current (not yet fully standardized) reference template for such documents is:

IOTC-YYYY-WPXX#-R[#LANG]

Where “LANG” indicates the language that the meeting report is written. Particular care should be taken to ensure that LANG is a standardized code to aid in programmatic identification of language versions (e.g., currently the LANG code for English is both “EN” and “E”).

National reports should be noted as ‘NR#’ instead of ‘WPXX#’; similarly, scientific committee reports should be noted as ‘SC#’, and expert consultations should be noted as “EC#”. When revisions are submitted for a document, this should be indicated with a “_Rev#” after the full reference name, e.g., IOTC-YYYY-WPXX#-R[#LANG]_Rev#. Currently, there is no distinction in the reference name to differentiate a presentation from a paper or report, but presentations should be noted in the reference name of the document as “Pres”, e.g. IOTC-YYYY-WPXX#-DOC#-LANG-Pres. A more detailed naming convention is outlined in [Appendix 4](#).

Enrich with improved metadata

Metadata elements displayed by Zenodo for each document can be improved at any time, including after publication of the DOI. The IOTC metadata already available for each document and that were published with the IOTC documents when available are the abstract, keywords, and the authors.

Abstracts

The IOTC began publishing abstracts for documents online in 2018, and some are available in the metadata output provided by the IOTC. A vast majority do not have a published abstract, and these are currently listed as “Abstract not available” for the documents that are currently published on Zenodo. Future updates could undertake text mining of the IOTC document file to update Zenodo metadata for abstracts. Users might also be asked to provide enriched metadata directly to ensure relevant updates.

Keywords

The IOTC metadata for each document provides a field entitled ‘all_taxonomy_terms’, which includes the year of the meeting, the author name (generally ‘IOTC Secretariat’), and the meeting name. In addition, for each document, we added ‘IOTC’, ‘Indian Ocean’, the meeting name (e.g. WPEB), the document type (e.g., Meeting report), and the scientific name, English, and French common names for any species specified in the title.

Author list

Currently, IOTC metadata includes the last name and first initial of authors. Where listed in the metadata, these author names were added to the author list. Otherwise, authors were listed as the IOTC Secretariat. Future efforts to improve metadata can include text mining of the document for author names, identifying an ORCID for each author where possible. In any case, authors are encouraged to get and share their own ORCID to foster data citation.

Zenodo communities

A [Zenodo IOTC community](#) was created and agreed upon for all documents related to the IOTC. Documents in the automated workflow developed in this project are automatically added to the IOTC community. Individual Zenodo users can request that their documents are uploaded to this community, which can be either accepted or rejected by the curators of the community (i.e., the IOTC Secretariat). If rejected, the document will still be uploaded on Zenodo, but will not be listed as a document of the IOTC community. Sub-communities relating to the IOTC meeting can also be created in the future (e.g. WPTT, WPDCS, etc).

R filtering workflow

The first step of the R workflow is the filtering step whereby IOTC files are selected according to criteria agreed upon at the IOTC-IRD workshop. A google sheet (hereafter, ‘gsheet’) of the IOTC document metadata is created for the selected documents, following the structure required by the geoflow package.

2) Validation of the selected documents

Once the filtering step has been completed, the resulting gsheet of document metadata is sent to the IOTC for approval. The Secretariat verify that each document proposed by IRD for publication is appropriate for the scientific community.

IRD identifies whether the document has been previously assigned a DOI. This step assumes that most documents have not been published elsewhere. The documents with the highest likelihood of

publication are the Information papers, which are removed at the filtering step, and Publications. The Publications are few in number and an internet search is performed by the IRD team for each one to verify that there are no other publication records or a DOI assigned.

The document titles are checked against the titles in the physical document file, using a text mining script ([check_files.R](#)) to identify that the correct file is associated with the correct metadata.

Finally, the Secretariat verifies the metadata associated with each document and suggests corrections and improvements for the IRD team to implement. The IRD team implements these changes, and the IOTC gives its final approval of the documents and the metadata.

3) Publication on Zenodo/Zenodo Sandbox

Source data

This workflow uses the IOTC document metadata that have been provided by IOTC and downloaded locally as csv files from <https://iotc.org/test/biblio/docs>.

The csv files that were used here were downloaded on 11/4/2019. They have been made available in a gdrive folder ([IOTC Metadata SOURCE](#)). Note, that 'Expert_consultations' is not a document type searchable/downloadable from the IOTC website, and the [expert_consultations.csv](#) file that is in the IOTC_Metadata_SOURCE folder was created in the filtering step.

Set up for Zenodo and Zenodo Sandbox

Prior to publication on zenodo.org, the upload/publishing of the documents are tested on sandbox.zenodo.org. This site mirrors that of Zenodo, and allows the full workflow to be tested prior to real publication. This point is very important, as once a document is published on Zenodo, it is given a DOI that is assigned to the file for at least 20 years (the planned lifetime of the Zenodo host project CERN). In the case of automated publications such as in this project, the Zenodo team has agreed to remove mistakes in publications within a 1 month time frame, or on the rare occasion when removal from publication is required (e.g. for political reasons).

In order to publish, [sign up](#) for a Sandbox account and get a token. The token can be created by first clicking on the Account tab on the top right (username listed here), and navigating to 'Applications'. From there, a user can create a new token (check all options). Record the token somewhere safe before exiting the window as the token will not be viewable again once the site is closed or navigated away from.

Set up for Gdrive

This workflow uses google sheets for final upload to Zenodo. Manual set up is required to enable the sharing function of the gsheets.

1. Create a gdrive folder for the first instance of following this workflow
2. Make a folder in the home gdrive directory, i.e.,


```
#drive_mkdir("iotc_csv_share")
#drive_mkdir("published")
```

3. Navigate to the new folder in the gdrive and share the FOLDER. The share function must be manually activated by right clicking on the gdrive folder and clicking 'Get shareable link'.

Workflow to assign Zenodo DOIs to IOTC documents

All scripts for this work are available on the [aenieblas/Zenodo DOI IOTC](#) github page. The entire workflow for this project can be called from the wrapper script [ZENODO DOI IOTC.R](#).

This workflow follows three major steps :

1. The filtering process : applies the selection criteria for each different IOTC document types (as described [above](#)). This step filters the files by IOTC document type and creates gsheets of the IOTC document metadata, following the structure required by geoflow.
2. Step 1 of the geoflow workflow : deposits metadata on Zenodo/Sandbox and reserves DOIs.
3. Step 2 of the geoflow workflow : uploads and publishes files on Zenodo/Sandbox, then creates a final gsheet with the Version DOIs and Concept DOIs for publication on IOTC website.

1. Filtering process

The csv source files are called in the [iotc 2 gsheet filters.R](#) script. This script applies the selection criteria as decided upon at the Zenodo/OpenAire workshop held in Victoria, Seychelles on 28/4/2019 and as described in the [above](#) and in the filtering protocol ([Appendix 3](#)).

The selection criteria are applied, the files that have been filtered out are removed and put in a different gsheet <iotc_doc_type>_removed, with a column "Filter used" added to describe why the file was removed.

The files that remain are put into the <iotc_doc_type>_filtered gsheet. Several columns are then populated with the syntax required for geoflow (i.e. modifications are made posteriorly to their Identifier, Relation, Rights¹, Creator, Description, and Subject fields.

The gsheet information is returned from this function and can be used to derive the share link of the gsheet is then used for the next step.

2. Publishing on Zenodo Sandbox, step 1 : deposit metadata, reserve DOIs

The first step of the geoflow workflow is to deposit the metadata of the documents and reserve a DOI. This step is called from the function [geoflow_step1.R](#).

This script first makes the configuration file that is required to launch geoflow, using [make_json_step1.R](#). A configuration file is required to define the actions that geoflow must take during the workflow. The configuration file defines the url of the <iotc_doc_type>_filtered gsheet (the share link from the filtering process above), whether to delete any previously deposited drafts (clean=TRUE/FALSE) and whether or not to publish (publish=TRUE/FALSE). For this step, clean = TRUE and publish = FALSE. A Sandbox *token* is also required here, i.e. :

¹ The Rights column is not used in the present case and can be removed; however, the [geoflow_handler_entity](#) script in the geoflow package must be updated for this change.

```
$make_json_step1(sharelink_of_<iotc_doc_type>_filtered,token)
```

The geoflow workflow is launched via the `executeWorkflow()` function of the geoflow package, which reads the config.json file, e.g.,

```
$executeWorkflow('iotc_config_step1.json')
```

We have adapted some aspects of the [eblondel/geoflow](#) R package, to specifically deal with the metadata required for IOTC documents and we use the [zen4R](#) package to interact with Zenodo from R. These adaptations were made to the following scripts of the geoflow package:

the `geoflow_action_zen4R_deposit_record.R` file :

original : [eblondel/geoflow/R/geoflow_action_zen4R_deposit_record.R](#);

adapted : [aenieblas/Zenodo DOI IOTC/geoflow_action_zen4R_deposit_record.R](#)).

and the `geoflow_handler_entity.R` file :

original : [eblondel/geoflow/R/geoflow_handler_entity.R](#);

adapted : [aenieblas/Zenodo DOI IOTC/geoflow_handler_entity.R](#)).

The function `executeWorkflow('iotc_config_step1.json')` deposits the metadata derived from the gsheets of the filtered files onto Zenodo/Sandbox, and returns a csv into the local jobs/metadata folder of the DOIs and their respective reference IDs. The DOIs are added to the `<iotc_doc_type>_filtered` gsheets and the gsheets are copied to create a separate gsheets for the second step of the geoflow process, `<iotc_doc_type>_uploaded`. The Identifier column of the 'uploaded' gsheets is then modified to add the doi. The Data column is also modified with the url of the source file.

The output of the `geoflow_step1()` function is the gsheets information of the `<iotc_doc_type>_uploaded`, which can be used to derive its sharelink. This gsheets information is then used in the next step.

3. Publishing on Zenodo/Zenodo Sandbox, step 2 : upload files and publish

The next step of the workflow is to upload the files onto Zenodo/Zenodo Sandbox and publish them. This step is launched using the `geoflow_step2()` function. This function requires the gsheets information from the `<iotc_doc_type>_uploaded` gsheets.

This script first makes the configuration file that is required to launch the second step of for publishing on Zenodo/Zenodo Sandbox with geoflow, using [make_json_step2.R](#). This configuration file defines the url of the `<iotc_doc_type>_uploaded` gsheets (the share link from `geoflow_step1()`), sets `clean=FALSE` and `publish=TRUE`. A Zenodo Sandbox `token` is also required here, i.e. :

```
$make_json_step2(sharelink_of_<iotc_doc_type>_uploaded,token)
```

The second step of the geoflow workflow is then launched, again via the `executeWorkflow()` function of the geoflow package, which reads the config.json file, e.g.,

`$executeWorkflow('iotc_config_step2.json')`

The function `executeWorkflow('iotc_config_step2.json')` uploads the file found in the Data column of the `<iotc_doc_type>_uploaded` and publishes the document to Zenodo/Zenodo Sandbox. The `<iotc_doc_type>_uploaded` is then copied locally and appended for the 'Concept DOI', which is the 'top level' DOI. This is the DOI that is recommended for the IOTC to publish on their website. A new gsheet, `<iotc_doc_type>_published_<timestamp>` is then added to the gdrive. The timestamp indicates the date/approximate time of publication of the files on Zenodo/Zenodo Sandbox.

The output of the `geoflow_step2()` function is the gsheet information of the `<iotc_doc_type>_published_<timestamp>`, which can be used to derive its sharelink. This gsheet information is what is passed on to IOTC for subsequent publishing of the Concept DOI on their website.

IOTC created a Zenodo account during the workshop and shared the Zenodo API token with IRD to allow IRD to formally publish documents on Zenodo on behalf of the IOTC Secretariat. Files published to Zenodo using the IOTC Secretariat token are thus the property of the IOTC.

IOTC registered an ORCID, as it was found to be beneficial for bibliographic tracking and metadata.

4) IOTC publishes the final DOIs on the IOTC website

As a final step in the publication process, the IOTC adds the Zenodo DOI to each document's IOTC website, including a hyperlink to the Zenodo page. Likewise, the Zenodo publication page for each IOTC document includes a hyperlink of the reference name of the document, linked to the IOTC website (e.g. [Figure 2](#)). To facilitate future publication of the DOI on the Drupal-based IOTC website, a Drupal-DOI plug-in was identified to aid in managing the workflow (https://github.com/NCAR/drupal_DOI).

Results

The filtering protocol was successfully applied to all IOTC documents. Of the 6124 documents derived from the IOTC website and defined as an appropriate document type for the scientific community (i.e. meeting reports, executive summaries, meeting documents, national reports, publications, datasets, and expert consultations), 3314 documents were retained for subsequent publication on Zenodo following the selection criteria ([Table 3](#)). Validation of the documents was limited due to time constraints of the Secretariat and the duration of the OpenAire project (30 June 2019). Within the timeframe of the project, 601 documents, including meeting reports and executive summaries, were published on Zenodo (~18% of the total filtered documents).

Table 3. Summary of the IOTC documents of the six main document types deemed appropriate for the scientific community, and their publication status.

Publication order	IOTC approval	Document type and link to IOTC metadata	Total number of docs	Rows retained for	Rows removed	Published on Zenodo	Date published on Zenodo	Date IOTC metadata downloaded
-------------------	---------------	---	----------------------	-------------------	--------------	---------------------	--------------------------	-------------------------------

				publication				
1	YES	MEETING REPORTS	444	237	207	237	25/06/2019	11/4/19
2	YES	EXECUTIVE SUMMARIES	386	364	22	364	3/7/2019	11/4/19
5		MEETING DOCUMENTS	4433	2050	2383	NO		11/4/19
3		NATIONAL REPORTS	298	281	17	NO		11/4/19
4		PUBLICATIONS	33	29	4	NO		11/4/19
		DATASETS	457	260	197	NO		11/4/19
6		Expert consultations	103	93	10	NO		
		TOTAL	6154	3314		601		

The image shows a Zenodo publication page for a report from the Indian Ocean Tuna Commission. The page layout includes a header with the Zenodo logo, search bar, and navigation links. The main content area features the document title, author information, and an executive summary. A central viewer displays a PDF of the report cover, which includes the IOTC logo and the title 'Report of the Eighteenth Session of the Indian Ocean Tuna Commission' held in Colombo, Sri Lanka, from June 1-5, 2014. To the right, a sidebar provides metadata such as view and download counts, the OpenAIRE logo, publication date, DOI (10.5281/zenodo.3255729), keywords, related identifiers, alternate identifiers, communities, and license (Creative Commons Attribution Non Commercial 4.0 International). Below the viewer, a file list shows the PDF document (1.5 MB) with preview and download options. A 'Citations' section at the bottom indicates no citations are present.

zenodo Search Upload Communities Log in Sign up

June 13, 2014 Report Open Access

Report of the Eighteenth Session of the Indian Ocean Tuna Commission

IOTC, Secretariat

Executive summary The Eighteenth Session of the Indian Ocean Tuna Commission (IOTC) was held in Colombo Sri Lanka from 1 to 5 June 2014 Chaired by Mr Daroomalingum Mauree (Mauritius). A total of 172 delegates attended the Session composed of 134 delegates from 25 Contracting Parties (Members) of the Commission 3 delegates from 2 Cooperating Non-Contracting Parties and 35 delegates from 14 Observers to the Commission (including 1 from FAO and 12 invited experts)) The Commission adopted the IOTC IUU Vessels List as provided in Appendix X. (para. 66) The Commission granted the status of Cooperating non-Contracting Party until the close of the 19th Session in 2015 to Djibouti Senegal and South Africa. (paras. 69 75 77) The Commission adopted the budget for and the scheme of contributions for 2014 and 2015 as outlined in Appendix XII and Appendix XIII respectively with the understanding that areas of potential savings continue to be explored. (para. 96) The Commission adopted 7 Conservation and Management Measures in 2014 consisting of 6 Resolutions and 1 Recommendation as follows: · Resolution 14/01 On the removal of obsolete Conservation and Management Measures · Resolution 14/02 For the conservation and management of tropical tunas stocks in the IOTC area of competence · Resolution 14/03 On enhancing the dialogue between fisheries scientists and managers · Resolution 14/04 Concerning the IOTC record of vessels authorised to operate in the IOTC area of competence · Resolution 14/05 Concerning a record of licensed foreign vessels fishing for IOTC species in the IOTC area of competence and access agreement information · Resolution 14/06 On establishing a programme for transshipment by large-scale fishing vessels · Recommendation 14/07 To standardise the presentation of scientific information in the annual Scientific Committee report and in Working Party reports

4 views 4 downloads See more details...

Indexed in OpenAIRE

Publication date: June 13, 2014
DOI: DOI 10.5281/zenodo.3255729
Keyword(s): IOTC IOTC Secretariat Indian Ocean Commission Meeting Reports
Related identifiers: Identical to urn:IOTC-2014-S18-RE (URN)
Alternate identifiers: https://iotc.org/documents/report-eighteenth-session-indian-ocean-tuna-commission
Communities: Fisheries and aquaculture Indian Ocean Tuna Commission
License (for files): Creative Commons Attribution Non Commercial 4.0 International

Preview

Page: 1 of 151 Automatic Zoom

Indian Ocean Tuna Commission
 Commission des Thons de l'Océan Indien
 iotc ctoi

IOTC-2014-S18-R[E]

Report of the Eighteenth Session of the Indian Ocean Tuna Commission

Colombo, Sri Lanka, 1–5 June 2014

Files (1.5 MB)

Name	Size	Preview	Download
IOTC-2014-S18-RE.pdf	1.5 MB		

md5:e3ad7b786c898dc3723558a080df6e7a

Citations

Show only: Literature (0) Dataset (0) Software (0) Unknown (0) Citations to this version

No citations.

Versions

Version	Date
Version 1	Jun 13, 2014

10.5281/zenodo.3255729

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3255728. This DOI represents all versions, and will always resolve to the latest one. Read more.

Share

Cite as

IOTC, Secretariat. (2014, June 13). Report of the Eighteenth Session of the Indian Ocean Tuna Commission. Zenodo. http://doi.org/10.5281/zenodo.3255729

Figure 2. An example of a Zenodo publication page. The document title, authors, and abstract are given. A pdf of the document is available for direct viewing or download. The menu to the right gives information on the date of publication, the DOI, keywords, and related identifiers. These identifiers link the user back to the document's page on the IOTC website. Links are also given to the Zenodo communities that list this publication. The license and version number are listed. Sharing information in the way of a recommended citation is listed and the document can be shared directly on several social media platforms.

Discussion

The short-term objective of this project remains the complete validation by the IOTC and publication on Zenodo of the documents identified by this project as appropriate for the scientific community. The long-term objective of this project is to publish IOTC documents each year, following a standard protocol. The protocol and codes for the filtering and publication steps are complete and published (code available at [aenieblas/Zenodo_DOI_IOTC](#)). The objective of this paper is to introduce working party participants to the concept of DOI, the benefits of DOIs and the workflow of assigning DOIs to future IOTC documents of working parties and meetings. A next step is to then explain to working party participants how to assign their own DOI from Zenodo, get author IDs from ORCID, or alternatively, to provide their document's metadata to a google spreadsheet to enable the Secretariat to then assign their DOIs.

Further enrichment of the metadata associated with each document would be useful for the larger community and lead to greater document and data discovery (e.g. refining keywords). The IOTC provided a detailed meeting registration list that can be used to give information about authors. Indeed, the author list for each document should include the full first and last name, a middle initial if used, an ORCID should be encouraged, the affiliation of the author, and the contact details of the author (email address). This list of contacts/authors could be compiled in a google document, which could then be used in Rmarkdown to get the contacts listed in the report.

For future working parties, users could :

- 1) Learn how to assign a DOI to their research object.
- 2) Populate a google spreadsheet with the core metadata elements required for their document for a DOI to be reserved by the IOTC (preferably via a Drupal plug-in workflow).
- 3) Register their name to an ORCID.

Any future publication protocol will include a buffer period of about 2-3 months after the working party or meeting and prior to final publication of the DOI on Zenodo. This period will enable enough time to ensure stable versions of the documents or datasets are provided to the IOTC by the participants.

Acknowledgments

This project was funded by the European OpenAIRE-Connect H2020 project (grant No. 731011). We would like to thank the Zenodo team for their technical support and flexibility in the IOTC use case. We thank Sylvain Bonhommeau for his helpful discussions and inputs in the development and undertaking of this project.

References

Emmanuel Blondel, Julien Barde, & Wilfried Heintz. (2019, April 3). geoflow: R engine to orchestrate and run geospatial (meta)data workflows (Version 0.1-beta). Zenodo. <http://doi.org/10.5281/zenodo.3138921>.

Blondel, Emmanuel, & Barde, Julien. (2019, August 27). zen4R: R Interface to Zenodo REST API (Version 0.3). Zenodo. <http://doi.org/10.5281/zenodo.3378733>.

Appendix 1. Workshop agenda

Workshop OpenAIRE Zenodo: IOTC use case

Draft agenda

29 April 2019

Location: IOTC meeting room, Victoria, Mahe, Seychelles

9am.

Presentation of the objectives of the OpenAIRE project, IOTC use case

Explanation of Zenodo, R workflow

10:00am. Break

10:30am.

Definition of a simple-use case

Criteria to filter documents

12:00 pm. Lunch.

1:00pm.

Criteria to filter documents

3:00pm. Break

3:30pm.

Pending questions

Strategies for future working parties

5:00pm. End of workshop.

Appendix 2. Workshop presentation

[Link](#) to the IRD workshop presentation

https://docs.google.com/presentation/d/1tJh4rlaSXt4J6st-naLdGb_6ImRnrMLazikhNYKapy4/edit?usp=sharing

Appendix 3: IOTC DOCUMENT FILTERING PROTOCOL

1. df\$IRD_reference is now added to the data frame and will align reference names to the [naming convention](#) agreed upon in the Workshop Meeting Report.
2. Keep document types, including:
 - a. Meeting reports
 - b. Meeting documents
 - c. Executive summaries
 - d. Datasets
 - e. National reports
 - f. Publications
 - g. [Expert consultations]
3. Filters by document type, including:
 - a. **MEETING DOCUMENTS ([KEPT](#)/[REMOVED](#))**
 - i. Reassign
 1. National reports (121)
 2. Executive summaries (80)
 3. Expert consultations (94)
 4. Meeting reports (46)
 - ii. Add 'English' where language metadata are blank
 - iii. Replace long “-” for short “-” in the reference names
 - iv. Keep only high priority meetings
 1. Scientific committee (SC)
 - i. IOTC-2012-WPB10-R[E] : df\$Title=='FINAL'
 1. Decision : retitled and moved to Meeting Reports
 2. Working Party on Temperate Tuna (WPTmT)
 3. Working Party on Data Collection and Statistics (WPDCS)
 4. Working Party on Tropical Tunas (WPTT)
 5. Working Party on Billfish (WPB)
 6. Working Party on Methods (WPM)
 7. Working Party on Ecosystems and Bycatch (WPEB)
 - i. IOTC-2015-WPEB11-37 Rev_1 :Metadata disorganised due to “;” in title. Fixed at source
 8. Working Party on Neritic Tunas (WPNT)
 9. IOTC ad hoc Working Group on FADs (WGFAD)
 - i. Single document note by Secretariat to exclude already excluded: IOTC-2017-WGFAD01-09 Rev_3
 10. Indian Ocean Tuna Tagging Symposium (TagSym)
 11. Commission
 - v. Remove rows that are complete duplicates (i.e. across all columns): 0
 - vi. Remove agreed-upon IOTC [Secretariat documents](#)
 1. Agenda/Ordre du jour/
 2. Proposed schedule/Proposition de calendrier
 3. Progress report/Progrès Accomplis
 4. List of documents/Liste des documents

5. Outcomes/Résultats
 6. Recommendations
 7. Report on the IOTC xxx/Rapport du Secrétariat/Report of the Secretariat
 8. Capacity building
 9. Program(me) of work/Programme de travail
 10. Previous decisions of the Commission
 11. Update on the performance review
 12. Rules for the appointment of an invited expert
 13. df\$Title=='credentials'
 14. List of participants
 15. Etat des PAN
- vii. Remove information papers
 - viii. Remove withdrawn papers
 - ix. Remove presentations
 - x. Further analyse duplicates in the Reference name to identify which metadata are different between the files
 1. Meeting session: Keeps the most recent, even if multiple
 2. Language: Keep the latest version of each language
 3. Authors:
 - a. How to handle different authors between duplicated references ? Author flags for meeting docs are currently dealt with using other filters; however, for future reference, for duplicated references with different authors :
 - i. Choose the most recent doc
 - ii. Where the authors are :
 1. 1) the Secretariat and 2) Anonymous = assign both to the Secretariat; and
 2. Where 1) Secretariat and 2) empty = assign both to the Secretariat
 - iii. CASE BY CASE : With same reference name, diff language, AND different authors, E.g., IOTC-2016-SC19-11 : english version = 'iotc secretariat' and the french version = 'DFARD'
Decision : XXXX (changes to be performed by the IOTC Secretariat and eventually re-incorporated once everything has been fixed)
 4. Abstract: Keep the most recent doc(s), if more than one is most recent, choose that with an abstract, assuring that Language is not the other reason for duplication
 - xi. Remove Ad-hoc documents (e.g., IOTC-2009-WPB-Adhoc7, IOTC-2009-WPB-Adhoc2,...)
 - xii. Remove Annex documents (e.g., IOTC-2006-WPB-11-ANNEX, IOTC-2006-WPB-10-ANNEX,...)
 - xiii. Duplicates that pass through the filters:

- i. IOTC-2012-SC15-08: this is a duplicate based on date availability (older version removed), but it also seems to be part of a draft executive summary - include with executive summaries?
 - 1. Decision : most recent version reclassified as executive summary
 - ii. IOTC-2007-WPTT-03: seems to be 2 of the same paper with very slight differences in metadata.
 - 1. Decision : kept the first entry, removed the other
 - xiv. Check files against identifiers
- b. MEETING REPORTS (KEPT/REMOVED)**
 - i. Reclassify any national reports (0), executive summaries (0), or expert consultancies (3)
 - ii. Replace long “-” for short “.” in the IRD_reference
 - iii. Add ‘English’ where language metadata are blank
 - iv. Keep only high priority meetings (see [2.a.iii](#))
 - v. Remove rows that are complete duplicates (i.e. across all columns): 17 (this number is likely due to the previous reclassifying step)
 - vi. Remove information papers: 0
 - vii. Remove withdrawn papers: 0
 - viii. Remove presentations: 0
 - ix. Further analyse duplicates in reference names to identify which metadata are different between the files
 - 1. Meeting: For duplicates before 2012, where there is a report from the Scientific Committee and the WP, remove the SC document.
 - i. The meeting for IOTC-2017-WPTT19-R[FR] is listed as ‘WPDCS’, which is incorrect. Meeting metadata changed to WPTT19.
 - 2. Meeting session: Keeps the most recent, even if multiple
 - 3. Language: Keep the latest version of each language
 - 4. Abstract: Keep the most recent doc(s), if more than one most recent, chose that with an abstract, assuming that Language is not the other reason for duplication
 - 5. Authors: duplicate documents with different authors are flagged. No action currently taken to remove/change them (NONE for meeting reports)
 - 6. Files : are the links to the file documents the same?
 - a. FIPL/R551(Bi) and FIPL/R564(Bi): separate entries for French and for English, but same file, and same document (translation within the document)
 - i. The French entry for each has been removed and the English entry for each has had its identifier changed to include both language codes, i.e. ‘EF’.
 - x. Check files against identifiers

1. Flagged for removal
 - a. IOTC-2004-WPT-RE linked to [IOTC/WPT/04/INF-3](#)
 - i. Doc found
 - b. IOTC-2010-WPDCS-R[F] : [404 FILE NOT FOUND](#)
 - i. Doc found
 - xi. Added mandatory keywords specified by the Secretariat “IOTC, IOTC Secretariat, Indian Ocean” to all documents
- c. EXECUTIVE SUMMARIES ([KEPT/REMOVED](#))**
- i. Reclassify any national reports (0), executive summaries (0), or expert consultancies (0)
 - ii. Add ‘English’ where language metadata are blank
 - iii. Keep only high priority meetings
 - i. Note: this removes “Commission” meetings (3) and Workshop on Connecting the IOTC Science and Management Processes (1)
 - iv. Remove rows that are complete duplicates (i.e. across all columns): 0
 - v. Remove information papers: 1
 - vi. Remove withdrawn papers: 0
 - vii. Remove presentations: 0
 - viii. Further analyse duplicates in reference names to identify which metadata are different between the files
 1. Meeting session: 0
 2. Language: 1
 3. Abstract: 0
 4. Authors: 0
 5. File urls: 2 (in flagged files below)
 - ix. Remove ‘draft’ from the titles of the executive summaries where there are no “final” copies.
 - x. Many .docx files instead of .pdf (n=97) TO BE UPLOADED AS .DOCX AND REPLACED WITH PDF WHEN AVAILABLE.
 - xi. Assumed that ‘Anonymous’ and blank author entries are ‘IOTC Secretariat’
 - xii. Removed any authors that are not the IOTC (Marsac, Fonteneau).
 - xiii. Removed all Southern Bluefin Tuna reports (author = CCSBT)
 - xiv. Flagged files:
 1. IOTC-2017-SC20-ES20 : French language and title, but linked to English file (Suggest removing French entry except if you have the French version) : REMOVED
 2. IOTC-2016-SC19-ES13F : French language and title, but linked to English file (IOTC found this doc)
 3. IOTC-2015-SC18-13F/E : Proposals for improved figures in the tropical tunas executive summaries (REMOVED FROM EXEC SUMMS)
 4. IOTC-2017-WPTT19-23_Rev3 : Proposals to revisions to the IOTC Tropical Tuna Executive Summaries (supporting information) (REMOVED FROM EXEC SUMMS)
 5. IOTC-2017-WPM08-16 : Revision of Supporting information for IOTC species Executive Summaries (REMOVED FROM EXEC SUMMS)

- xv. Added mandatory keywords specified by the Secretariat “IOTC, IOTC Secretariat, Indian Ocean” to all documents
- xvi. Add species keywords (French and English) to the keywords of all executive summaries.

d. NATIONAL REPORTS (KEPT/REMOVED)

- i. Reclassify any national reports (0), executive summaries (0), or expert consultancies (1)
- ii. Add ‘English’ where language metadata are blank
- iii. Keep only high priority meetings
 - i. Note: this removes “IOTC Working Party on the Implementation of Conservation and Management Measures (WPICMM)”, which was considered a ‘medium’ priority meeting.
- iv. Check for complete duplicates of rows (i.e. across all columns): 0
- v. Remove information papers: 97
 - 1. These papers were listed as info papers in the meeting documents. Then the duplicates in Title and Meeting.year are FLAGGED (n=5 are then checked manually)
 - i. n=1: 2x Korea national reports 2009 - one is actually an INF paper (Scientific questions and recommendations following the 2009 yellowfin IOTC stock assessment) which is removed.
 - ii. n=2 language duplicate (no change needed)
 - iii. n=3:5 Rev_1 were kept, old versions removed
- vi. ‘Authors’ changed to match the country in the Title for National Reports (i.e. replaced empty values or ‘IOTC Secretariat’).
- vii. Remove withdrawn papers: 0
- viii. Remove presentations: 0
- ix. Further analyse duplicates in reference names to identify which metadata are different between the files
 - 1. Meeting session: 1
 - i. Reference name is the same (IOTC-2015-SC18-NR01) for different files: IOTC–2015–SC18–NR01; IOTC–2016–SC19–NR01.
 - 1. Decision: Changed the Reference to match the file.
 - 2. Language: 1
 - 3. Abstract: 0
 - 4. Authors: 0
- x. Authors are listed in the Creator field with the prefix “nation”, e.g., “nationSaudi Arabia”; which then lists Saudi Arabia as the last name of the author.
- xi. File flags:
 - 1. No file : IOTC-2015-SC18-NR33 Bangladesh – National Report (2015)

2. Reference name in metadata : IOTC-2010-SC-Inf03 linked to file with [IOTC-2009-SC-Inf03](#) as reference name inside.
3. Reference name in metadata : IOTC-2009-SC-INF12 linked to file with [IOTC-2007-SC-INF12](#) as reference name inside.
4. Reference name in metadata : IOTC-2013-SC16-NR05 National Report – Eritrea linked to file [IOTC-2013-SC16-NR04](#) (Comores national report)

e. PUBLICATIONS (**KEPT/REMOVED**)

- i. Checked for secretariat documents (1: performance review)
- ii. Double-check that these do not already have a DOI (manual check, n=33)
 - a. FLAG: ID-THONS (EN); and TUNA-ID-CARDS (EN) seem very similar
 - i. Decision : ID-THONS (EN) does not have a file associated, thus it has been removed.
 - b. FLAG: [IOTC-2013-Pilot data collection](#) (seems like a repeat of meeting document : [IOTC-2012-SC15-38\[E\] \(English\)](#)), but files are different. Suggest removing the meeting doc version
 - i. Removed meeting document version.
 - c. FLAG: IPTP-90-WP-20 -Jan (google search finds it labeled as a thesis and a book), suggest removing
 - i. Removed

f. EXPERT CONSULTATIONS (**KEPT/REMOVED**)

- i. Reclassify any national reports (6), executive summaries (0), or meeting reports (0)
- ii. Add 'English' where language metadata are blank
- iii. Replace long “-” for short “.” in the reference names
- iv. Keep only high priority meetings
- v. Remove rows that are complete duplicates (i.e. across all columns): 0
- vi. Remove agreed-upon IOTC [Secretariat documents](#) : 0
- vii. Remove information papers: 0
- viii. Remove withdrawn papers: 0
- ix. Remove presentations: 0
- x. Further analyse duplicates in reference names to identify which metadata are different between the files
 1. Meeting session: 0
 2. Language: 0
 3. Abstract: 0
 4. Authors: 0
- xi. Manual check through files (n=99) :
 1. EC601-11 : Country report - Mauritius
 - a. Decision: reclassified as National Report
 2. EC7-07 : Country report of small tunas and seerfish, Saudi Arabia
 - a. Decision: reclassified as National Report

3. REPORT OF THE Xth EXPERT CONSULTATION ON INDIAN OCEAN TUNAS (is this a meeting type? In which case, is it a priority meeting?)
 - a. Removed

g. DATASETS ([KEPT](#)/[REMOVED](#))

- i. Reclassify any national reports (0), executive summaries (0), or meeting reports (0)
- ii. Add 'English' where language metadata are blank
- iii. Replace long “-” for short “-” in the reference names
- iv. Remove rows that are complete duplicates (i.e. across all columns): 0
- v. Remove agreed-upon IOTC [Secretariat documents](#) : 0
- vi. Remove information papers: 0
- vii. Remove withdrawn papers: 0
- viii. Remove presentations: 0
- ix. Remove code (low priority - needs to be formalised) : 1
- x. Remove model parameterisations : 6
- xi. Remove data catalogues : 24
- xii. Remove equations : 25
- xiii. Remove data forms : 3
- xiv. Remove dataset availability documents : 14
 1. Not to be confused with 'Available data '
- xv. Remove reference data / CE REF : 60
- xvi. Remove stock assessment data from Secretariat : 27
- xvii. Remove Standardized CPUE from the Secretariat : X
 1. “Joint” CPUE files to be removed
 - a. E.g., [IOTC-2017-WPTT19-DATA19](#) : Joint CPUE indices for Bigeye tuna
- xviii. Keep: nominal catch, 3 catch and effort (sep), size frequency,
 1. Remove the zip file of the combined catch and effort for 'all gears' AND 'all vessels' : 37
- xix. In regards to dataset Titles, does the Secretariat suggest to change them to standard names?
 1. e.g. 'Catch and Effort - vessels using pole and lines or purse seines' to 'Catch and Effort - surface fisheries'?
 - a. Though these have different TITLES, do their Reference names match?

Appendix 4: Naming conventions

- Meeting Documents : IOTC-YYYY-WPXX#-DOC#-[LANG]
- Meeting Reports : IOTC-YYYY-WPXX#-R[LANG]
- National Reports : IOTC-YYYY-SC##-NR##-[LANG]
- Executive Summaries : IOTC-YYYY-WPXX#-DOC#-[LANG]
 - Working party is either WP or SC
- Publications : No naming convention?
- Expert consultation : IOTC-YYYY-EC-##
- Revisions should be added to the end of the reference name as “_Rev#”