HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

SOP: Using Elements to manage marine image data

The **purpose** of this Standard Operating Procedure (SOP) is to describe to all **administrators, data managers and users** of the federated MareHub Elements environment how to add and manage research image data in the **Elements environment**.

The **goal** of this document is to **enable stakeholders** to use Elements environment optimally and **consistently** to support its purpose as the image data provider for other tools (PANGAEA, BIIGLE, ..).

The **scope** of this SOP includes the steps necessary to prepare the Elements environment for incoming data, add new data to it, manage and curate image data inside the Elements environment and provide access to the image and metadata to external users and tools.

Completing the SOP will **result** in **well-curated image** data that adheres to open standards and which is **accessible to selected users** through the web interfaces of the Elements environment **for analysis, visualization and advertising.**

The *Elements* environment is a Media Asset Management System (MAMS) and consists of hardware and software components. It is a commercial product and has been chosen as the infrastructure for FAIR marine images in the MareHub of GEOMAR, AWI and Hereon. The environment provides PB-scale storage space for marine images, a REST API to access the data (e.g. through the marine-data.de portal or through the BIIGLE annotation software) and a browser-based interface to work with the data and to share data with colleagues. The physical storage is split between large-scale servers operated at the institutes and mobile versions that can be taken on expeditions. The data management team at GEOMAR, the MareHub working group on Videos/Images and the international Marine Imaging Community develop metadata profiles and standard operating procedures towards making marine images FAIR using the Elements infrastructure.

Terms used in the Elements environment

Term	Usage	Comment
Production	Recommended: **One *Production* per expedition / project** (e.g. SO268). Deprecated: one *Production* per camera system.	*Productions* are the fundamental data structure in the Elements environment. *Productions* are the main structuring element for image data! Default access rights can be specified per *Production*.
Workspace	**One *Workspace* per camera platform** (e.g. ROV-KIEL6000). Deprecated: one Workspace per deployment/acquisition run.	The secondary data structure in the Elements environment. *Productions* contain at least one or many *Workspaces*. *Workspaces* can be mounted by the Elements desktop client (through SMB, NFS, etc.). Individual access rights can be specified per *Workspace*.
Media Library	**One *Media root* for each workspace.**	Collaborative editing of footage in the browser, sharing of material, commenting on media. Belongs to a *Workspace*. Default custom fields can be defined for each media library.
Shares	May be used to exchange selected parts of image data.	A folder share into the file system with selectable protocol (SMB, NFS).
Automation	MarIQT methods will be implemented as *Automations*.	Build workflows of processing and QA/QC steps. Access *Automation* workflows from the desktop client. Cron jobs to execute regularly.
Workspace templates	**One *Workspace template* per camera platform** (e.g. ROV-KIEL6000).	A set of default settings to be applied to *Workspaces* (access rights, default folder structure, etc.). Essentially these are empty *Workspaces* from which new, empty *Workspaces* can be created.
Custom fields	**Use selected iFDO core fields as *Custom fields*.**	Belong to all *Media libraries*. Can be populated at media ingest time from metadata in the media files. *Custom fields* can be activated/ deactivated individually per *Media library*.

For administrators

1. Admins make the Elements hardware accessible to data managers and users. → FAIR

Result: *An accessible **URL** for web access and mount options (**NFS, SMB**, …) for network drives.*

Info: To provide access to the Elements environment for data managers and users it is essential to make the hardware accessible through the web interface (for configuration, exploration, sharing and file upload) and through mount options such that stakeholders can access the file systems like a network drive.

Tip: Also consider to be able to provide a physical access to the hardware to a limited group of users (e.g. the data managers) to make data transfer of Terabytes into the Elements environment feasible.

2. Admins create the set of required iFDO core fields as custom fields in the Elements media library. → FAIR

Result: *The standardized set of **iFDO core fields** is available as **custom fields** in the Elements instance.*

Info: The Elements media library collects metadata in a database. It is possible to populate those fields during ingest or afterwards. In either case it is essential that the respective fields have been made available as custom fields in the respective Elements instance. The set of required fields are the following:

iFDO core field	Data type	iFDO core field	Data type
image-set-uuid	Text field	image-set-handle	Text field
image-set-name	Text field	image-uuid	Text field
image-hash-sha256	Text field	image-filename	Text field
image-datetime	Date field	image-latitude	Number field
image-longitude	Number field	image-depth	Number field
image-event	Text field	image-platform	Text field
image-sensor	Text field	image-pi	JSON field
image-creators	JSON field	image-license	Test field
image-copyright	Large text field	image-abstract	Large text field

Tip: The media library database can be filled automatically from metadata incorporated into the media files during media ingest. This requires very well-maintained image metadata.

3. Admins add all required Elements automations to their system.

Result: *A set of **image curation algorithms** is available for the users to execute as Elements **automations**.*

Info: The Elements environment provides a computation framework to conduct repetitive tasks. The AG Videos/Images of the MareHub develops automations that can help with image curation tasks (see https://gitlab.hzdr.de/datahub/marehub/ag-videosimages/). Fetch and install those that are beneficial for your instance and use cases.

4. Admins create templates for the platforms that will provide data repeatedly.

Result: *A set of **template** folder structures and settings to be re-used for repeatedly occurring image data sets.*

Info: Image data in the Elements environment needs to be structured by project / cruise, camera platform and acquisition event / station. As camera platforms are likely to re-occur, it is suggested to create those as templates to ease later image ingests. It is suggested to provide one empty dummy folder in the template as a basis for all events of that Workspace. This dummy folder should contain the subfolders to organize the data according to the iFDO recommendations (“raw”,”intermediate”,”processed”,”products”,”external”,”protocol”).

Tip: Use unique and human-readable identifiers for the camera platforms and thus workspace names that can be linked to either the AWI sensors catalogue, the GEOMAR equipment catalogue or another gear database.

5. Admins create proxy profiles for their media files.

Result: *A set of proxy **profiles** for previews of the image data (e.g. for web **thumbnails**, **streaming**).*

Info: Elements uses automations to compute proxies for image data (using ffmpeg). These proxies can be lower-resolution versions or different file formats. Proxy profiles can be defined per Elements instance and activated per Media library.

Tip: Add the proxy profiles described at https://gitlab.hzdr.de/datahub/marehub/ag-videosimages/ to keep your data compatible to the marine-data.de portal.

6. Admins create a BIIGLE user and API token for annotation access (optional).

Result: *A BIIGLE instance can access the Elements environment to display images for annotation.*

Info: Making image data available for annotation is the final step in providing access for scientific use.

Tip: Push the MareHub team to quickly develop an Elements automation that can make a Production or Workspace available for annotation in BIIGLE by single click. Or develop this feature yourself! It would be highly appreciated.

For data managers

7. Data Managers create a production for a new expedition / project.

Result: *An **empty production** with a unique name.*

Info: This step has to be conducted once per expedition / project.

Tip: Use the cruise acronym or project acronym to keep things short and concise.

8. Data Managers create a workspace for each platform.

Result: *An **empty workspace** for each camera platform used in the expedition / project.*

Info: This step has to be conducted once per camera platform used in the expedition / project.

Tip: Use workspace templates to have everything prepared for the scientists.

9. Data Managers create the media library for each workspace.

Result: *An **empty media root** for each camera platform workspace.*

Info: The media library is collecting metadata and providing views to the data. Create one for each workspace by navigating to the workspace's setting's "Advanced" tab and click on "Create Media Root".

10. Data Managers grant access to the workspaces and media libraries to the BIIGLE user (optional).

Result: *The **BIIGLE** user has **read access** to the workspace and media library.*

Info: Add the BIIGLE user to the workspace by clicking on "Workspaces" in the left menu, clicking on the respective workspace, then navigating to the "Permissions" setting tab and clicking on the "+ Add Permissions" button. Search for the BIIGLE user and click "Select". For the media library access click on "Media Library" in the left menu, then click on the respective library in the list and then on the "Configure" button in the top right. Go to the "Permissions" tab and click "+ Add users and groups" and search for the BIIGLE user and click "Select". In the next modal only select "Download Original Files" and "Download proxies" from the "Permissions" list.

11. Data Managers grant access to workspaces and media libraries to all necessary users.

Result: *All necessary users have access to the workspaces and media libraries.*

Info: Add a user to the workspace by clicking on "Workspaces" in the left menu, clicking on the respective workspace, then navigating to the "Permissions" setting tab and clicking on the "+ Add Permissions" button. Search for the respective user and click "Select". For the media library access click on "Media Library" in the left menu, then click on the respective library in the list and then on the "Configure" button in the top right. Go to the "Permissions" tab and click "+ Add users and groups" and search for the respective user and click "Select". In the following modal select the access rights the user shall have.

Tip: You can specify groups or even connect to an Active Directory and use the groups therein to manage access rights in Elements.

12. Data Managers ingest the iFDO metadata into the media library of the workspace.

Result: *All **iFDO metadata** relevant for data discovery are available to the **Elements search** functions.*

Info: Use the Elements automation to incorporate iFDO metadata into the media library for that specific workspace.

For scientists

13. Scientists discuss storage needs with an RDM team that operates an Elements server

Result: *A **docx file in a cloud folder** that the RDM team can access.*

Info: To prepare the Elements environment, information on the incoming data is required. There should be a data management plan for this information already somewhere. Should storage demands exceed the available capacity or exceed the respective "fair-use" policy of the institute, solutions regarding acquisition of additional storage capacity need to be found early.

Tip: Re-use an existing data management plan for the project within which the image data is being created. That way you prevent duplicate work and the RDM team is well-informed of the project scope and incoming data.

14. Scientists provide a list of project members that need access to the data. → FAIR

Result: ***Access rights** in the Elements environment have been set.*

Info: Access to the image data in the Elements environment is governed by the respective RDM team. Provide a list of users that need read / write access to the data.

15. Scientists migrate the curated image and metadata to the Elements server. → FAIR

Result: *Curated data is stored with **backup** in the centrally maintained server infrastructure **and accessible** through network folders, web URLs and handles.*

Info: This step includes: a) documenting each image acquisition by an acquisition protocol; b) Adding a data processing report to each image data set; c) Creating an iFDO for each image data sets; d) collecting all curated image data and metadata on a portable storage device and e) physically migrating the files to the Elements environment.

Tip: See image curation SOP for details.

16. Scientists update the data management plan. → FAIR

Result: ***Additional sections in the DMP** which is stored as a living document in a cloud.*

Info: Update the DMP with all relevant information from the acquisition process and its documentation. Add the Elements production and workspace information to your DMP.

Tip: Let the RDM team create OSIS deliverables from the DMP to monitor your project's data requirements.

Appendix:

References:

- SOP “Image curation and publication”: https://doi.org/10.5281/zenodo.5704846
- SOP “Image curation and publication” supplement: https://doi.org/10.5281/zenodo.5704844
- SOP “iFDO creation”: https://doi.org/10.5281/zenodo.5681429
- SOP „iFDO creation“ supplement: https://doi.org/10.5281/zenodo.5683082
- FAIR marine images: https://marine-imaging.com/fair
- Elements media library at GEOMAR: https://medialib.geomar.de
- Elements media library at AWI: https://medialib.awi.de
- BIIGLE image annotation at GEOMAR: https://annotate.geomar.de/

Glossary:

DMP: Acronym for Data Management Plan – a living document that collects information on planned and acquired datasets as well as models, software and more

FAIR: Acronym for Findable, Accessible, Interoperable, Reusable – describing the leading principle in data management to increase the value of data

iFDO: Image FAIR Digital Object – a standardized format for the description of image metadata

Image: Photo (still image) or video (moving image)

RDM: Acronym for Research Data Management – referring to the process and also the team of highly-trained people at GEOMAR

SOP: Acronym for Standard Operation Procedure – a static or dynamic document describing a sequence of tasks acting on data entities to reach a defined goal

Information about this document:

Title: Using Elements to manage image data

Authors: Timm Schoening

Month, Year: 2021/11

Abstract: This Standard Operating Procedure document describes how to use the Elements Media Asset Management System (MAMS) as an environment for FAIR marine research data.

Note: This SOP targets admins, data managers and users of Elements infrastructure of the MareHub. It includes information and references to AWI, GEOMAR and Hereon infrastructure that may not be available to outside users.

DOI:

Keywords: Image, curation, publication, photo, video, iFDO

Licence: CC-0

Related Identifiers:

Communities: MareHub of the Helmholtz Association

Revisions:

Version	Date	Author	Comment
v1.0.0	2021/11	Timm Schoening	Initial draft of a public text version of this SOP. Compiled from discussions in the MareHub AG Videos/Images.

image-project	
image-platform	
image-sensor	
image-creators	

Done	Who	When	What
Preparation phase			
	Scientists	Before project start	Discuss storage needs
	RDM team	Before image ingest	Create Production
	RDM team	Before image ingest	Create Workspace(s)
	RDM team	Before image ingest	Create Media Libraries
	Scientists	Before image ingest	Provide list of project members
	RDM team	Before image ingest	Grant access to all necessary users
	RDM team	Before image ingest	Grant access to BIIGLE user
Ingest phase			
	Scientists	Ingest	Migrate curated data to Elements
	Scientists	After Ingest	Update Data Management Plan
	RDM Team	After Ingest	Populate custom fields from iFDO metadata

Note: Lines in green are recommended steps that make your image curation and publication life easier.