DELIVERABLE 1.2

Report on requirements for transparency and quality control

Version 1.0

WP 1
Deliverable 1.2
Lead Beneficiary: ICES
Topic: Shaping ecosystem based fisheries management
Grant Agreement No: 101000318
Dissemination level: PU
Date: 05.10.2022

Please cite this work as: Millar, S., Jacobsen, N. S., Rindorf, A. and Taylor, M. (2022): SEAwise Report on requirements for transparency and quality control. Technical University of Denmark. https://doi.org/10.11583/DTU.21221432

Contact: Anna Rindorf, https://orcid.org/0000-0002-4290-3983, email: ar@aqua.dtu.dk

SEAwise has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000318

SEAWISE

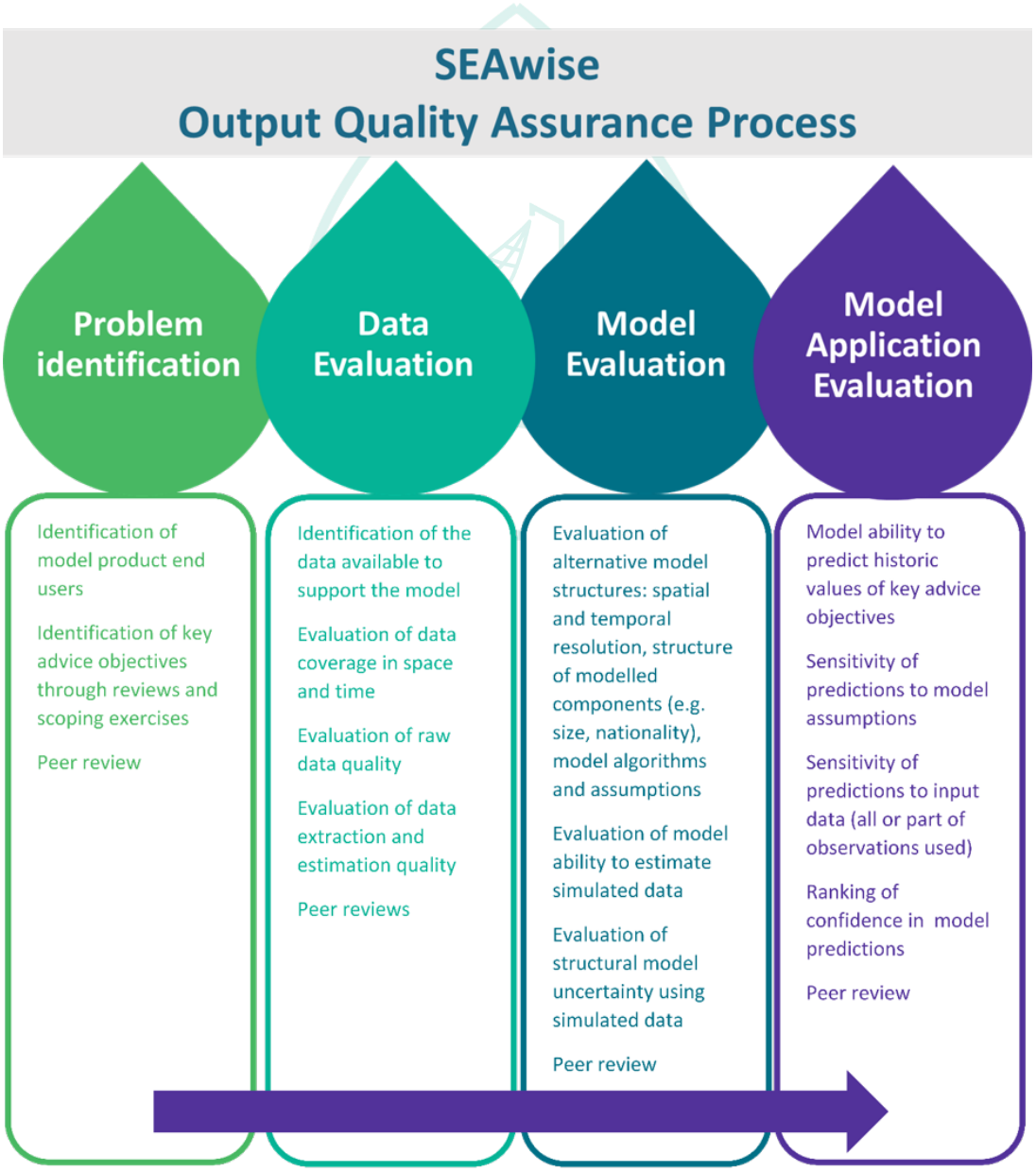
Executive summary

This report describes the common protocol for producing openly-available, quality assuring data and methods used in SEAwise. SEAwise data and methods are all data collected, processed, used or stored in connection with the SEAwise project. Making quality assured data openly available is a key priority in SEAwise in accordance with the FAIR principles: Ensuring that data are findable, openly accessible, interoperable and available for increased re-use. The processes in place to ensure this are described in this report.

SEAwise data are divided into four types:

- Quantitative data are organised in the SEAwise quantitative metadatabase, listing primary data sources and responsible parties for these
- Qualitative data from interviews and workshops are organised in the SEAwise qualitative metadatabase, listing primary data sources and responsible parties for these
- Software code will be available at the SEAwise Github after the first year of the project.
- Reports and other publications (scientific and otherwise) are available on the SEAwise website.

All data and models are evaluated using the SEAwise Output Quality Assurance Process.



Contents

1. SEAwise background 5
1.1 The role of this deliverable 6
1.2 Authors 6
2. Types of Research Data 6
3. Transparent data structures 6
3.1 Findable data 6
3.2 Open access to data 7
3.3 Interoperable data 7
3.4 Increase data re-use 7
4. Quality assurance processes 8
4.1 Quality control of data and models 8
5. Databases 10
6. Peer review 11
7. References 11
8. Annex A: ICES Transparent Assessment Framework 12
8.1 Data Profiling and Data Accreditation 13
8.2 Data Quality and Data Policy 13
9. Document Information 14

1. SEAwise background

The SEAwise project works to deliver a fully operational tool that will allow fishers, managers, and policy makers to easily apply Ecosystem Based Fisheries Management (EBFM) in their own fisheries. With the input from advice users, SEAwise identifies and addresses core challenges facing EBFM, creating tools and advice for collaborative management aimed at achieving long-terms goals under environmental change and increasing competition for space. SEAwise operates through four key stages, drawing upon existing management structures and centered on stakeholder input, to create a comprehensive overview of all fisheries interactions in the European Atlantic and Mediterranean. Working with stakeholders, SEAwise acts to:

- Build a network of experts - from fishers to advisory bodies, decision makers and scientists - to identify widely-accepted key priorities and co-design innovative approaches to EBFM.
- Assemble a new knowledge base, drawing upon existing knowledge and new insights from stakeholders and science, to create a comprehensive overview of the social, economic, and ecological interactions of fisheries in the European Atlantic and Mediterranean.
- Develop predictive models, underpinned by the new knowledge base, that allow users to evaluate the potential trade-offs of management decisions, and forecast their long term impacts on the ecosystem.
- Provide practical, ready-for-uptake advice that is resilient to the changing landscapes of environmental change and competition for marine space.

The project links the first ecosystem-scale impact assessment of maritime activities with the welfare of the fished stocks these ecosystems support, enabling a full-circle view of ecosystem effects on fishing productivity in the European Atlantic and Mediterranean. Drawing these links will pave the way for a whole-ecosystem management approach that places fisheries at the heart of ecosystem welfare. In four cross-cutting case studies, each centered on the link between social and economic objectives, target stocks and management at regional scale SEAwise provides:

- Estimates of impacts of management measures and climate change on fisheries, fish and shellfish stocks living close to the bottom, wildlife bycatch, fisheries-related litter and conflicts in the use of marine space in the Mediterranean Sea,
- Integrated EBFM advice on fisheries in the North Sea, and their influence on sensitive species and habitats in the context of ocean warming and offshore renewable energy,
- Estimates of effects of environmental change on recruitment, fish growth, maturity and production in the Western Waters,
- Key priorities for integrating changes in productivity, spatial distribution, and fishers' decision-making in the Baltic Sea to create effective EBFM prediction models.

Each of the four case studies will be directly informed by expert local knowledge and open discussion, allowing the work to remain adaptive to change and responsive to the needs of advice users.

1.1 The role of this deliverable

This deliverable aims to describe a common protocol for producing openly available, quality assuring data and methods used in SEAwise. SEAwise defines Research Data as all data collected, processed, used or stored in connection with the SEAwise project. Making quality assured data openly available is a key priority in SEAwise in accordance with the FAIR principles: Ensuring that data are findable, openly accessible, interoperable and available for increased re-use. The processes in place to ensure this are described in this report.

1.2 Authors

Nis Sand Jacobsen, Sarah Miller, Anna Rindorf and Marc Taylor

2. Types of Research Data

SEAwise divides Research Data into the following main types:

- Quantitative data collected and owned by partners or collaborators outside SEAwise that are not Person confidential, e. g. data collected on scientific survey, in scientific experiments and data publicly available.
- Quantitative data collected and owned by partners or collaborators outside SEAwise that are Person confidential, e. g. data from logbooks, VMS[1] data and data collected by fisheries observers.
- Qualitative data collected and owned by partners or collaborators in SEAwise that are Person confidential, e.g. data from interviews and workshops
- Primary data collected in SEAwise, e.g. recordings of interviews or workshops
- Software code developed or further elaborated in SEAwise.
- Reports and scientific publications

3. Transparent data structures

3.1 Findable data

Data used as input or produced as output in SEAwise are registered in the SEAwise metadatabase. The database is divided into four parts:

- Quantitative data are organised in the SEAwise quantitative metadatabase, listing primary data sources and responsible parties for these
- Qualitative data from interviews and workshops are organised in the SEAwise qualitative metadatabase, listing primary data sources and responsible parties for these
- Software code will be available at the SEAwise Github after the first year of the project.
- Reports and other publications (scientific and otherwise) are available on the SEAwise website.

The SEAwise quantitative and qualitative metadatabases contain metadata ensuring that data used in studies can be identified and a source of further information about the data is available. The quantitative database registers the

[1] VMS (vessel monitoring system) data is detailed position data generated by commercial fishing vessels on a regular basis with information on vessel ID, speed, position and direction.

component which the data pertains to (e.g. common sole), a data description including variables, a link to the data collection protocol, extent and resolution of fisheries data, Temporal and Spatial extent and resolution, the data source or provider, a data link where data are publically available, an extraction code link, an output data link, the SEAwise contributor(s) to contact, the file format, person submitting the data, comments and the version number. Findability of data is ensured through live links to the source or the doi of published data.

The qualitative metadatabase registers the events supporting the social science in SEAwise. They are categorised according to the type of contact, focus of the contact, date, selection of participants, the number of people participating, the duration of event/interviews, the topic(s) discussed, contact person and contact email as well as links to interview guide/workshop plan and aggregated data where this is consistent with protecting privacy.

3.2 Open access to data

The data used in publications and other products of SEAwise are published every 6 months in the SEAwise metadatabases including data descriptions and an associated open link to data in all cases where this is consistent with GDPR. Where partners are not able to host such data, the data will be hosted at DTU. 'Personal Data' is available only by request and then only in an aggregated format after an individual confidentiality declaration has been signed (See Annex A for an example). The proper use and publication of data collected under Council Regulation (EC) No 199/2008 are clearly described by the regulation and must be adhered to in all use.

The SEAwise data are stored in a variety of formats, with a preference for simple formats that are widely useable such as txt and csv for numerical data. All data links are live links. To avoid using unnecessary space to store large data, large input data sets may not be available by doi. However, links to where data can be obtained, output data (in aggregated or metadata format where this is required for GDPR considerations) and links to the extraction code used is always available in the SEAwise database. Additional large data will be available by contact to the contact person listed in the metadatabase. Quality assured code will be openly available, while code under development and quality assurance will remain restricted access.

When publishing data collected and processed in SEAwise the users must respect any and all restrictions on the use or reproduction of data, such as restrictions on data usage for commercial purposes and Person sensitive data, which are governed by GDPR. Scientific publications and reports should properly acknowledge data by citation. The citation must include as a minimum a reference to the database where the data were extracted from and the year in which the database was accessed.

3.3 Interoperable data

Ensuring that quantitative data can be widely used requires agreement on relevant resolution and format of the data. Decision on these are made regionally (Mediterranean Sea, Western waters, North Sea and Baltic Sea). Area identification is either through lat/lon coordinates or through references to ICES, GFCM or FAO statistical areas. Fish species are identified by their scientific name and WORMS code.

3.4 Increase data re-use

Data licence will be CC BY 4.0 where-ever this is consistent with the guidelines given by the owner of the raw data used. Data in the metadatabase are made publically available with an associated doi no later than at the time at which the results are published unless the data are protected by GDPR considerations. Data can be reused after the end of the project following the conditions under CC BY 4.0. Primary data can be attained where this is compatible with GDPR by contacting the contact person listed in the relevant SEAwise metadatabase.

4. Quality assurance processes

Data used in SEAwise are continuously screened for errors by expert teams working together to ensure reproducibility and reliability of the results. Data in the metadatabases are made publically available for a period no less than 5 years after the termination of the project (i.e. at least until 2030) allowing time for person outside the project to reproduce any results. All data have an associated contact person to which any data problems encountered can be reported. Primary data from interviews and workshops are stored and quality assured at the partners responsible for data analysis.

- Software code will be available at the SEAwise Github after the first year of the project.
- Reports and other publications (scientific and otherwise) are available on the SEAwise website.

4.1 Quality control of data and models

Data and models analysed and applied in SEAwise undergo thorough quality assurance to ensure that model results are suited to the objectives of the analyses and robust to uncertainties in model formulation. The quality assurance includes a review of data and model application against the criteria derived from the approaches of model life cycle evaluation of NRC (2007), ICES Transparent Assessment Framework (Fig. 1, Annex A) and ICES WGSAM (ICES, 2021), adding principles from HELCOM (2016) on confidence.

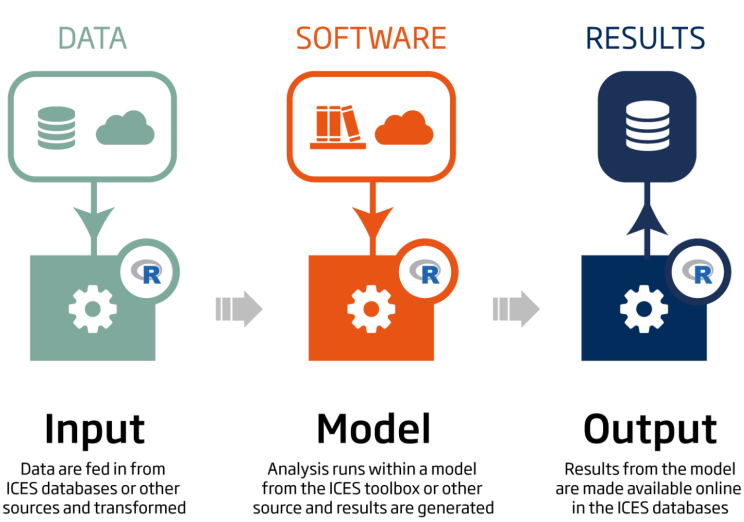


Figure 1: The process in ICES TAF

The aim is to ascertain that the model is suitable for the question being investigated. The end user (e.g., stakeholder, scientist, etc.) is important here, as the complexity and transparency of the model output must fit it's intended users. The process has four steps (Fig. 2):

- Problem identification
- Data evaluation
- Model evaluation
- Model application evaluation

Three of the points are similar to the ICES Transparent Assessment Framework steps (Data evaluation/Data, Model evaluation/software, Model application evaluation/Output).

In the first step of the **problem identification**, it is also vital that the models employed in SEAwise are determined appropriate for the problem for which they are used. Problem identification in SEAwise is based on the combined results of the systematic reviews and scoping exercises.

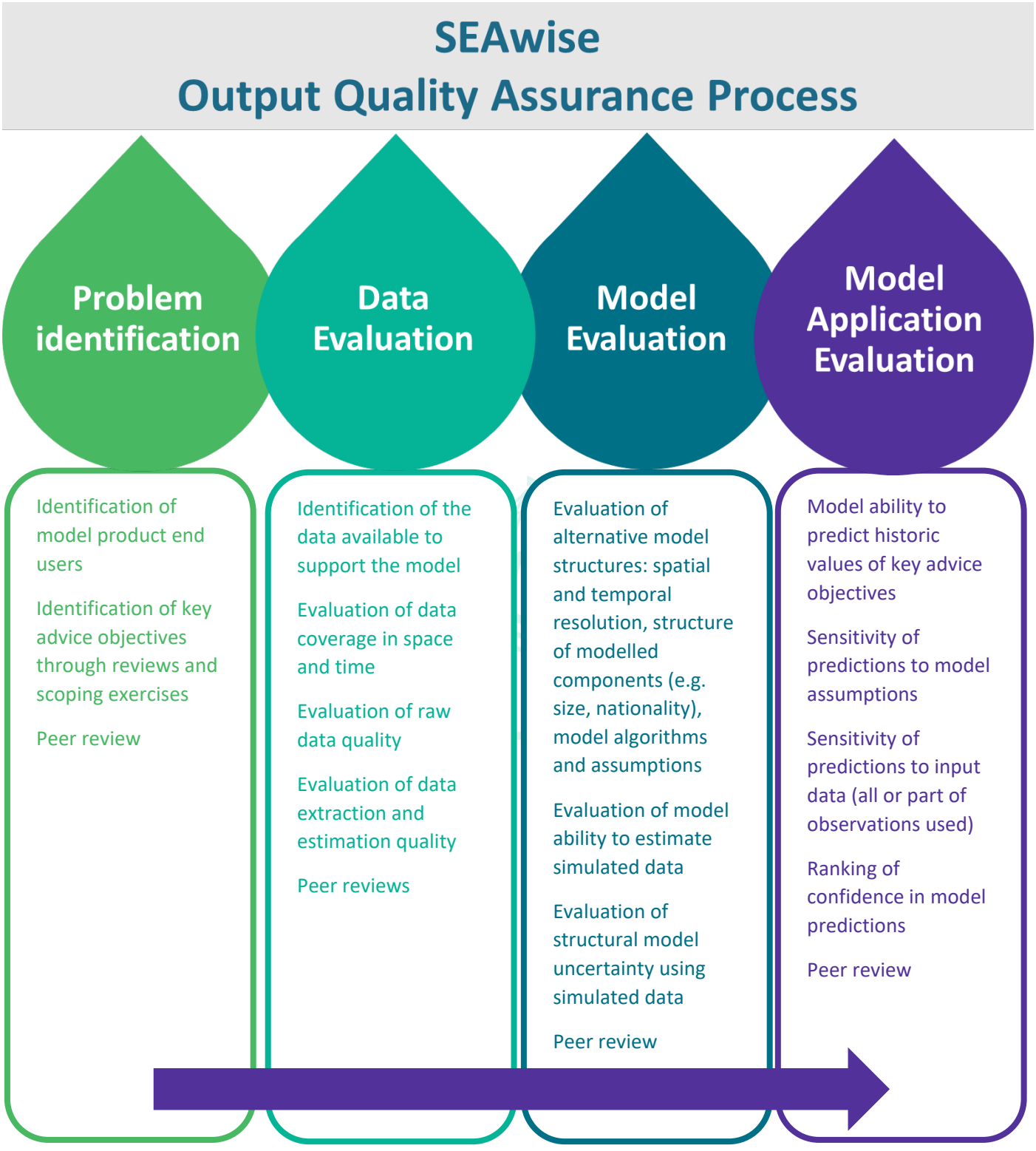


Figure 2: SEAwise Output Quality Assurance Process

The **data evaluation** in SEAwise investigates the spatio-temporal coverage of data and the quality of raw and processed data for use in the model. Criteria include whether the spatiotemporal resolution in the data sufficient to describe the extent and evolution of the model component, whether the data was collected and processed in a consistent manner with estimates of data uncertainty, including uncertainties of the data collecting scheme and whether the data are publically available in a useable format. The evaluation is supplemented by using the ICES Data profiler tool for relevant ICES data (Annex A, https://www.ices.dk/data/tools/Pages/Data-profiler.aspx). Data used in SEAwise are continuously screened for errors by expert teams working together to ensure reproducibility and reliability of the results. Data in the databases are made publically available for a period no less than 5 years after the termination of the project (i.e. at least until 2030) allowing time for person outside the project to reproduce any results.

The **model evaluation** investigates alternative model structures, the model spatial and temporal resolution, the structure of modelled components (e.g. size, nationality), model algorithms and assumptions, the ability of models to predict simulated data and to predict sensible future development from these data. The evaluation also includes evaluation of the ability of the model to estimate simulated data (level as well as temporal development) and an evaluation of structural model uncertainty using simulated data.

Under the **model application evaluation,** the ability of models to predict observed data and to predict sensible future development from these data is evaluated. For ecosystem models, points that can be considered for model appropriateness include (ICES, 2021; Kaplan & Marshall, 2016):

- Agreement with historical data sources or assessments:
    - Focal species should match biomass and catch levels and temporal changes over the historical time period. Suggested tests include modelling efficiency, RMSE. Note that model CV will not reflect model structural uncertainty, so this indicator of fit should be used with care.
    - Natural mortality decreases with age for majority of groups.
    - Age and length structure match expectations for majority of groups.
    - Diet predicted qualitatively matches empirically derived diet composition.
    - Spatial distribution from model predictions match reference datasets for spatial models
    - Ecosystem indicators match reference data if needed for the identified problem.
- Model stabilizes for an unfished, unperturbed 80–100 year run with functional groups persisting in the ecosystem

Further, a ranking of model prediction confidence is completed.

5. Databases

Data used in SEAwise are collated in a *quantitative* and a *qualitative* Excel data base. The quantitative database includes information from each data type on:

- Case study area
- Spatial and temporal resolution and extent
- Work packages the data is being used for
- Coarse description of the data
- Where to access the data
- Which SEAwise contributor to contact for access and information.

The qualitative database includes information on

- Type of interview contact

- Focus of the interview
- Participant selection
- Duration
- Topics
- Contact persons
- Link to the raw data.

The databases will be continuously updated throughout SEAwise to reflect the continuous expansion and development of the project.

6. Peer review

Models and their associated data will be reviewed by external experts in the ICES community, as well as in internal SEAwise benchmark workshops.

7. References

European Commision. 2016. General Data Protection Decision Making. Regulation No 2016/679.

HELCOM. 2016. Development of a biodiversity assessment tool. Helsinki. https://portal.helcom.fi/meetings/HOLAS II 5-2016-347/MeetingDocuments/4-1 Developement of a biodiversity assessment tool.pdf.

ICES. 2021. Working Group on Multispecies Assessment Methods. Copenhagen.

Kaplan, I. C., and Marshall, K. N. 2016. A guinea pig's tale: learning to review end-to-end marine ecosystem models for management applications. ICES Journal of Marine Science, 73: 1715–1724.

NRC. 2007. Models in Environmental Regulatory Decision Making. National Academies Press. 104–170 pp.

Regional Coordination Group of the North Sea and Eastern Arctic. 2016. Report of the Regional Co-ordination Group for the North Sea and Eastern Arctic (RCG NS&EA) 2018. Lyngby. https://datacollection.jrc.ec.europa.eu/documents/10213/1239599/2018_RCG+NSEA.pdf/25defcc6-3ef3-45aa-920e-7c67b858cd91.

8. Annex A: ICES Transparent Assessment Framework

ICES developed the Transparent Assessment Framework (TAF) in 2016 after there was an increasing demand from ICES experts for a user-friendly framework to organise their data, methods, and results. The aim of TAF is, foremost, to assure quality, improve efficiency, and ensure transparency of data and analyses that are used in the ICES advisory processes, this is done by giving stock assessors and reviewers the tools to routinely document data, methods and disseminated results used in ICES assessments.

Assuring quality, is a key element to provide assessments and advice products, and this is why TAF is implemented as an essential tool in the overall quality assurance framework at ICES, where various quality control checkpoints are included and documented within TAF, to create common data processing and protocols that are easily accessible and replicable, even with new data. TAF thereby provides a way to track changes in input data and scripts from year to year and allow alternative approaches to be archived (i.e., allow multiple assessment runs).
TAF is used to document the workflow in a series of R scripts (icesTAF, R package), that have been developed to prepare reports of plots and tables from the input data, describing where they came from (licencing/ownership) and what was done to them (data cleaning), before they were entered into the assessment model (toolbox).

The general principles of the ICES TAF R package assessment workflow is scripted in a series of R scripts:

Core scripts	Purpose
data.R	pre-process data and write TAF data tables
model.R	run analysis and write model results
output.R	extract results of interest and write TAF output tables

Optional scripts include report.R (prepare plots and tables for the report) and utilities.R (custom functions used in the above scripts).
The outputs are model specific results that can be extracted as text files/tables which can be used in (ICES) ecosystem advice and other assessment processes. These outputs can then be fed directly into assessment tools, such as the stock assessment database (SAG: https://sag.ices.dk).

Access to TAF assessments can either be from GitHub or the TAF web application (https://taf.ices.dk/app/about), which allows for easy access to view, store and run the assessment analyses. The outputs of TAF are usually R-based (but it can also use R as a wrapper) and have user manuals together with adequate documentation. ICES uses GitHub as their code hosting platform for transparency of methods for TAF assessments, expert groups and created R packages and other codes/scripts used for assessment purposes or data extraction from ICES databases. For this purpose, GitHub is a great open-sourced platform that provides version control and collaboration of repositories, where the users contribute code and upload datasets to TAF.

This open framework, together with the TAF app, then enables anyone to easily find, reference, download, and run the assessment from any stage in the process leading to the published ICES advice for a given stock. All TAF assessments are stored and accessible on the ICES TAF GitHub page (https://github.com/ices-taf/).

ICES strives to make sure that TAF works for all analyses that support ICES advisory products and allow analyses to be rerun quickly also ensuring that outputs can be replicated and also be used in future analyses with minimal changes. The benefits of using TAF has mainly improved time efficiency and reduced the workload of the ICES expert groups, implementing high-quality science, by making it online peer-reviewed, reproducible and keeping an open and structured workflow for high-quality science.

8.1 Data Profiling and Data Accreditation

Data profiling is the process of reviewing source data, understanding structure, content and interrelationships, and creating useful summaries of data. The process yields an overview that aids in the discovery of data quality issues, possible risks, and overall trends. To be able to create a framework where data and products, which are outside of the normal data management flow of ICES, the Data Profiling Tool (DPT) was developed (https://www.ices.dk/data/tools/Pages/Data-profiler.aspx). The DPT strives to provide a standardised way to gather information about ad hoc datasets and visualisations, aiding experts in evaluating the completeness of the supporting information for a data flow or a data product. The DPT is a checklist for those data flows and data products that primarily feeds scientific and/or advice outputs through ICES working groups, asking questions regarding data sharing, categorisation, storage and access, quality and format. Thereby, the aim of the DPT is to both document the data flow or product, by using the answers as a gauge for completeness of the data flow and to document ICES efforts to quality assure all aspects of its advice production.

In 2021, ICES Data Centre has achieved accreditation under the Core Trust Seal (CTS - https://www.coretrustseal.org/), which is an international, community based, non-governmental, and non-profit organization promoting sustainable and trustworthy data infrastructures. This accreditation supports data repositories complying with certain standards to ensure data quality, usefulness and archiving, having many benefits, such as building stakeholder confidence, enhancing the reputation of the data flow, and demonstrating that the data flow follows good practices. For ICES, this means that the data flows managed by the ICES Data Centre, are recognised in the wider international data management community as state of the art.

8.2 Data Quality and Data Policy

Data used in ICES data products and for advice purposes need to go through rigorous quality control and quality assurance to be able to pass the data quality verification process, following these 5 principles: accuracy, completeness, reliability, relevance, and timeliness. The ICES working group on the Governance of Quality Management of Data and Advice (WGQuality) analyses existing quality management processes within advice production and evaluate their coherence with the objectives of the ICES advisory plan, ensuring an end-to-end quality assurance framework, meeting international standards, and that all advice products are based on data that adhere to the FAIR principle (Findable, Accessible, Interoperable and Reusable). Data, which is submitted to the ICES databases, go through various validation (https://www.ices.dk/data/tools/Pages/data-validation.aspx), and quality (https://www.ices.dk/data/tools/Pages/quality-control.aspx) checks, making sure that the data quality standards are met. ICES databases are governed by experts groups from the ICES community, ensuring that data can be submitted in the correct format and with a controlled vocabulary (https://www.ices.dk/data/vocabularies/Pages/default.aspx), which is governed on a regional scale.

The ICES data policy states that all public data are under the Creative Commons (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). It applies to data managed by ICES and other data intended to facilitate the production of science-based advice and status reports, and serve the scientific community, making all data products, by default, publicly available, including those derived from restricted data. The existing restricted data licenses in ICES (RDBES, VMS access and VME/Birds and Seals) are also following the same language, definitions and headings as the CCBY license (https://www.ices.dk/data/guidelines-and-policy/Pages/ICES-data-policy.aspx).

9. Document Information

EU Project	No 862428	Acronym	SEAwise
Full Title	Shaping ecosystem based fisheries management		
Project website	https://www.seawiseproject.org/		

Deliverable	N°	D1.2	Title	Report on requirements for transparency and quality control
Work Package	N°	1	Title	Knowledge exchange
Work Package Leader	Anna Rindorf, DTU, ar@aqua.dtu.dk			
Work Participants	Sarah Millar, ICES, sarah-louise.millar@ices.dk Nis Sand Jacobsen, DTU, nsja@aqua.dtu.dk Anna Rindorf, DTU, ar@aqua.dtu.dk Marc Taylor, TI, marc.taylor@thuenen.de			

Lead Beneficiary	Sarah Millar, ICES, sarah-louise.millar@ices.dk
Authors	Sarah Millar, ICES, sarah-louise.millar@ices.dk Nis Sand Jacobsen, DTU, nsja@aqua.dtu.dk Anna Rindorf, DTU, ar@aqua.dtu.dk Marc Taylor, TI, marc.taylor@thuenen.de
Reviewers	Marc Taylor, TI, marc.taylor@thuenen.de

Due date of deliverable	31.03.2022
Submission date	31.03.2022
Dissemination level	PU[2]
Type of deliverable	R[3]

Version log			
Issue Date	Revision N°	Author	Change
31.03.2022			First version
05.10,2022	1	Anna Rindorf	Added doi and suggested citation

[2]Dissemination level (DELETE ACCORDINGLY): **PU:** Public, **CO:** Confidential, only for members of the consortium (including the Commission Services), set out in Model Grant Agreement, **CL:** Classified, information as referred to in Commission Decision 2001/844/EC

[3] Nature of deliverable (DELETE ACCORDINGLY): **R:** Report, **DEM:** Demonstration, pilot, prototype, plan design, **DEC:** Website, patent filing, market studies, press & media, videos, **Other:** Software, technical diagram, etc., **Ethics:** Ethics deliverable