

1. Introduction

ESG provides nature based solutions that are evidence based. Therefore, reliable models and databases are of high importance for us as Science Group and for all researchers working at ESG. Models & datasets play an important role in the research that WENR and WOT perform for clients. The importance of quality assurance of models and datasets has been indicated by Jansen et al. (2004). A translation of these suggestions into an audit process has been described by Houweling et al. (2015). Below a description of the practice of quality assurance for models and datasets as implemented in 2020.

Authors: Hengeveld G.M., van der Grefte-van Rossum J.G.M., de Bie P.A.F.

2. Overview of Quality Assurance

In the process of quality assurance two cycles are important;

1. The self-assessment of and audit on the quality of a specific model or dataset (figure 1).
2. The model-dataset-indicators development and investment planning cycle at institute level (figure 3).

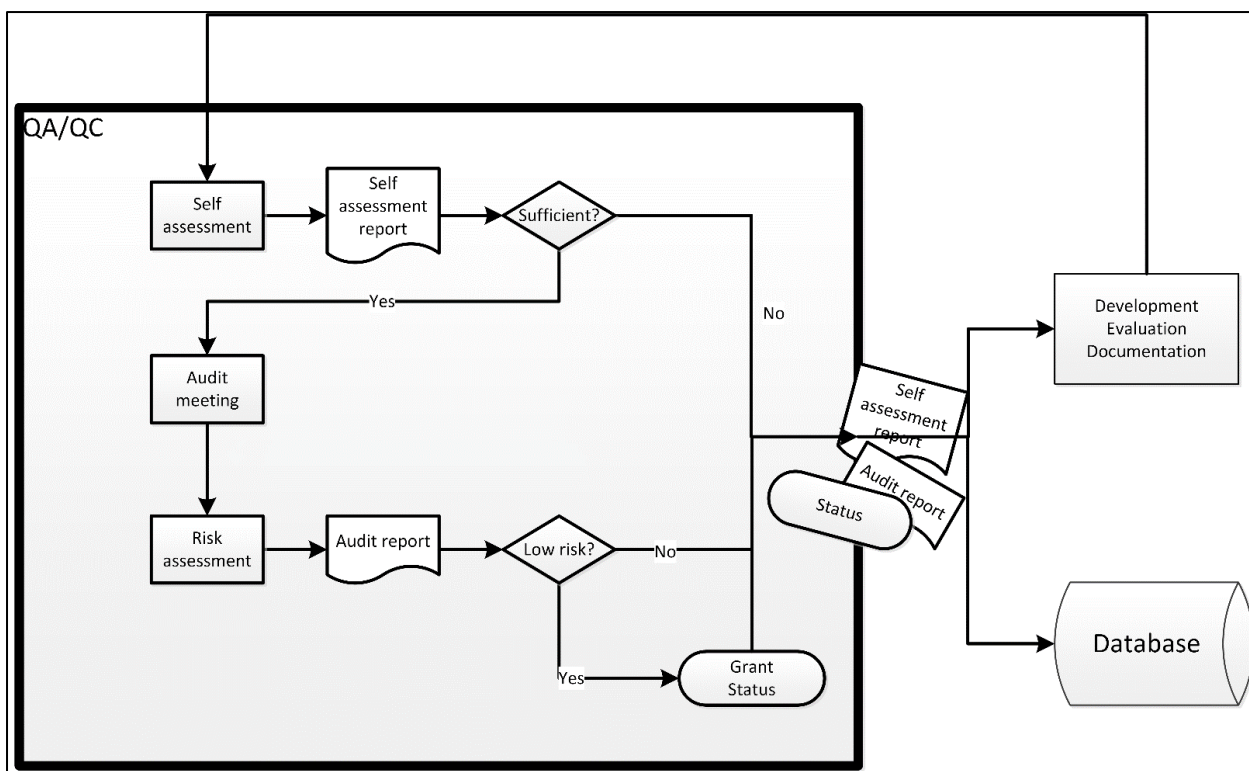


Figure 1: Self-assessment and audit cycle of an individual model/dataset

3. Explanation of the process of quality assurance

3.1 Two phases: self-assessment and internal audit

The assessment of the quality of models and datasets consists of two phases (figure 1). The first phase is a **self-assessment**, a **form** is used¹. In the self-assessment the model/dataset developers provide a reference to the documentation of the model for each item on this checklist. They also assess how well the criteria for the requirement are covered within the provided references. Comments are provided indicating possible actions to cover currently uncovered criteria. The self-assessment is sent to the Models & Data Assurance Manager (MDAM) of ESG and stored in a central archive. It is considered good practice to include the self-assessment form in the development plan and reflect on it. The self-assessment is periodically (depending on model developments and use) updated. Good practice is to re-affirm the self-assessment annually.

¹ The current versions of forms mentioned are available through the intranet group:

<https://intranet.wur.nl/Project/QualityofModelsDataKwaliteitsSlag>

When the self-assessment deems all criteria covered, the team/theme leader and the MDAM can decide to plan an **audit meeting** (second phase). Two auditors, not involved in the development of the model/dataset, receive the self-assessment and the associated documentation. Based on this documentation the auditors re-assess the cover of the criteria per requirement. The findings of the audit team are discussed with the model/dataset developers and can be adjusted based on new insights. For non-conformities, a risk-assessment is performed during the audit meeting using the rubric as set out in WI-0033 of WENR-ESG, i.e., for each non-conformity, the chance & impact of failure of the model/dataset on this requirement are assessed on a 5-point scale. Multiplication of the chance and the impact produce a quantified risk assessment (figure 2).

probability of problem occurring	likely / unknown	5	10	15	20	25
	4	4	8	12	16	20
	3	3	6	9	12	15
	2	2	4	6	8	10
	small	1	1	2	3	4
		1	2	3	4	5
		Exploratory research	Research paper		Policy advice	Policy support
		impact of problem occurring				

Figure 2: Risk assessment in line with WI-0033. Risk is calculated as the product of the probability of occurrence of a problem, and the impact of the occurrence of a problem. Unknown probability is ranked as high risk. Types of use are shown as guidance for the potential impact of a problem occurring.

The findings of the audit team and the risk assessment combined form the audit report. If none of the non-conformities receives a risk above 6, and the cumulative risk across all requirements is below 10, the model is assessed as Status A, A+, or AA, depending on the depth of all criteria covered. The audit report is stored in the database and actively shared with responsible team leaders and theme leaders. The quality status A, A+ or AA is valid for a limited period of 5 years, and needs to be re-affirmed after this period if use of the model/dataset is still anticipated.

In situations where non-conformities can be solved within one month, a re-assessment can be scheduled.

3.2 The database of models and datasets

The database of models and datasets is maintained by the Models & Data Assurance Manager (MDAM) and the model steward, both ESG.

Two activities are scheduled annually.

1. The MDAM has a meeting the WENR-team leaders and WOT-Theme leaders. In this meeting updates to the database are discussed and planned audits are prioritized.
2. The model steward reminds the maintainers of models and datasets to update the metadata and self-assessment forms on their model or dataset in the database.

The current database of models and datasets contains the metadata of models, indicators and datasets for which WENR is the owner and that a) generate a substantial revenue for WENR or b) are used for WOT Natuur & Milieu tasks. The WENR-database is connected to the WUR Model and Data search database to be hosted by the Wageningen Modelling Group at the Wageningen Data Competence Centre (WDCC). Full lists of the models and datasets included in the database will be linked through the [intranet group](#).

The metadata contained in the database consist of

1. The WUR default metadata (the current definition is found on the [intranet group](#)).
2. WENR specific metadata including
 - a. Classification of Under development, Experimental, Active, Historic (box 1), see below.
 - b. Quality status (Undefined, Self-assessed, Status A, Status A+, Status AA)
 - c. Self-assessment reports
 - d. Audit reports

Updates in the database occur through

1. **Addition of new models.** Addition of new models to the database follows on the annual work-meeting of the WENR-team leaders, WOT-Theme leaders resp., and the MDAM. Based on developments within the team or theme.
2. **Updates in the classification.** Updates in the classification of the models or datasets occurs either when the WENR-team leaders, WOT-Theme leaders resp. do not anticipate current or future use (Classification to 'Historic'), or when the quality status changes (see point 3).
3. **Updates in the quality status.** Updates in the quality status are associated to the actions in the assessment of the quality of the model/dataset (section 2) and are accompanied by an addition of a self-assessment or audit report.
4. **Updates in the WUR default metadata.** Updates in the WUR default metadata can be made continuously by the maintainer of the model/dataset. The project leader quality assurance reminds the maintainer on an annual basis to check these metadata.

Box 1. The classification of models as defined by the anticipated use of the model/dataset, and the quality status

Quality status	Anticipated use	
	Yes	No
Status A/A+/AA (Self)-assessed	Active	Historic
Undefined	Experimental Under development	Historic

The MDAM consults the WENR-ESG QAM-manager on an annual basis to discuss

- the quality of the self-assessment and audit reports in the database
- the efficiency of the workflow
- the desired quality levels

3.3 Use of quality information in application of models and datasets

The information on the quality of a model or dataset is used in different stages

1. When discussing models/datasets to be used with clients, the minimum required quality level can be established. In principle the aim is for a minimum level of status A. As per clients request the quality level aimed for in models/datasets used in projects for that client can be increased to A+, or AA.
2. At the start of each application project, the project leader of that project takes notion of the quality status of the models and datasets for which use is foreseen through the most recent audit report or an (updated) self-assessment report. Based on this information the project leader determines the suitability of the use of the model or dataset in the project. When necessary, the project leader decides on additional activities in development, analysis and documentation.
3. In the discussion between the model/dataset development team and the team leader the assessment of the quality status and its components can help formulate short term and long-term investment needs as written down in annual activity plans and management plans for the model/dataset.

4. Activities, roles & responsibilities

This section takes out the activities mentioned in the previous three sections and indicates the roles of the involved people. Figure 3 shows the model-dataset-indicators (MDI) development and investment planning cycle at institute level. Table 1 shows the roles of the involved people.

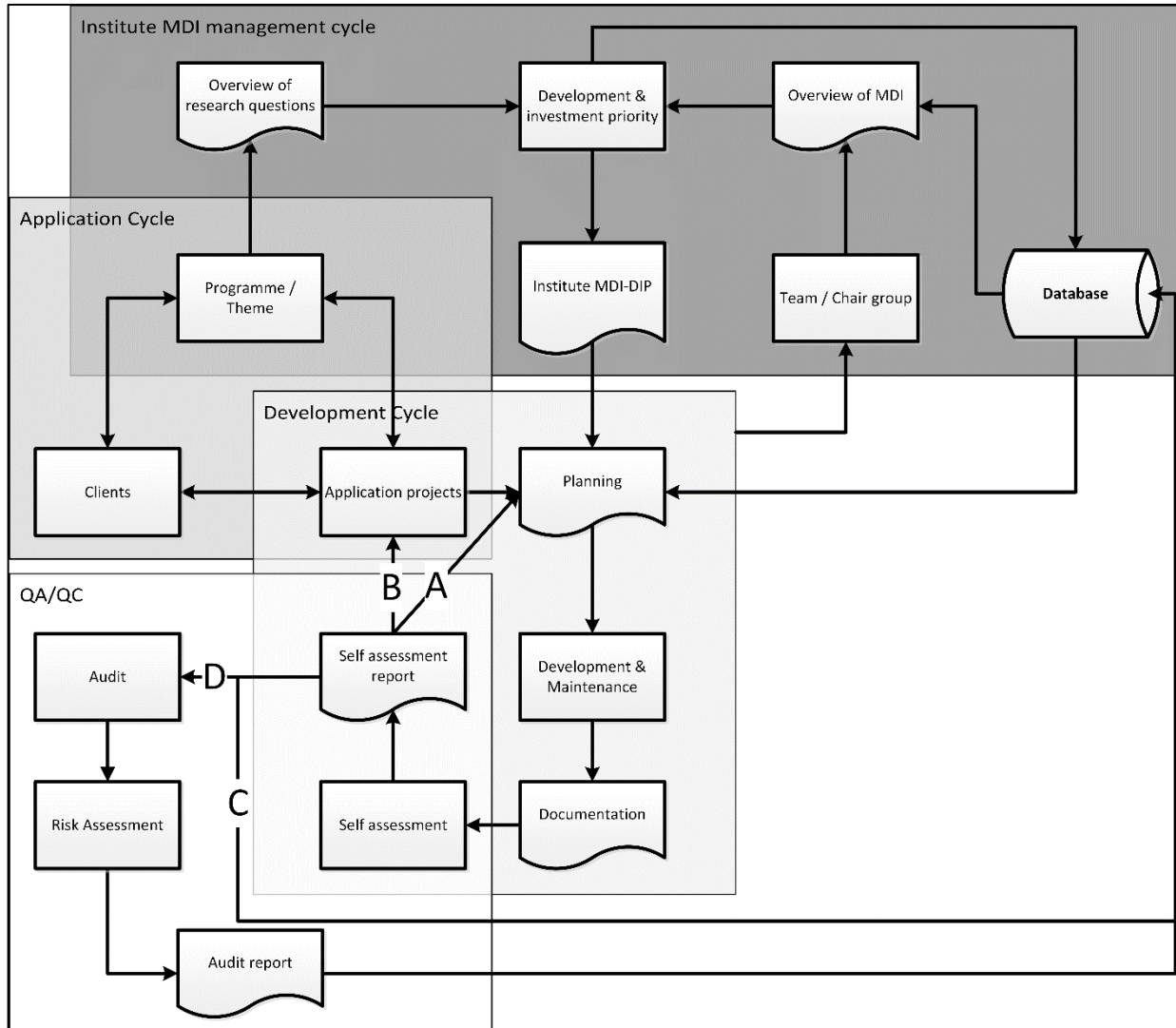


Figure 3: The quality assurance of models and datasets embedded in the general operation at institute level and the development and application cycles of models and datasets. Terms and responsibilities are explained in table 1.

Table 1: The roles of the involved people based on figure 3.

Component	Description	MDAM	Model Steward	Institute Auditors	Biometris Auditors	Model Development Team/ Model Coordinator	Project Leader Application Project	Chair Holders/ Team Leader	Programme Leader/ Theme Leader	Kam/QAM
Institute Mdi Management Cycle										
Assuming MDI management at institute level with institutes as matrix organisation, the MDI management cycle combines overview of demand (from programmes) with an overview of capacity and development need (MDI and HR) into an MDI development and investment plan										
Overview Of Research Questions	From the bundled interests of clients and development in society an overview of research questions for now and in the future is extracted									Provide overview of research questions
Overview Of Mdi	An overview of the MDI within the institute/BU/team is made. Including quality status, investment needs, development options	From the database provide an overview of the quality status of MDI	Support extraction from database					Provide an overview of models within the BU/team and their needs/desires for investment and development		
Development & Investment Priority & Mdi-Dip	At institute level it is decided where to place emphasis, which models need auditing, which investments in model development are needed. The result of this is laid down in the MDI-DIP (model-datasets-indicators development and investment plan)	Discuss & advice						Discuss & write	Discuss & write	

Component	Description	MDAM	Model Steward	Institute Auditors	Biometris Auditors	Model Development Team/ Model Coordinator	Project Leader Application Project	Chair Holders/ Team Leader	Programme Leader/ Theme Leader	Kam/QAM
Database	An institute level database is maintained. The database contains metadata, self-assessment reports, audit reports, risk assessment, links to documentation and management/development plans	Oversight & Provide overview of MDI (quality) status	Maintain database - ensure all documents are in			Provide own metadata & documentation				Oversight over quality assurance
Application Cycle										
Within the application cycle MDI are used in application projects. This is where MDI meet clients.										
Application Projects						Evaluate model performance in application	Apply models in projects			Overview of applications and client needs
Development Cycle										
Within the development cycle a model is developed. This can be part of one or multiple (pre-application) projects.										
Planning	The planning is a crucial component of the Plan-Do-Check-Adjust cycle. During planning the MDI-DIP and other funding options are considered, the self-assessment and possible other information in the MDI-database are used to motivate plans and plan further model evaluation					Make & adjust the plan				
Development & Maintenance and Documentation	The development cycle follows a 'plan-do-check-adjust' cycle. For this the development entails documentation and quality assurance activities	Support in self-assessment (content)	Support in self-assessment (administration)			Develop model, test model, evaluate model, write documentation		Take note of model needs & ensure self-assessment		

Component	Description	MDAM	Model Steward	Institute Auditors	Biometris Auditors	Model Development Team/ Model Coordinator	Project Leader Application Project	Chair Holders/ Team Leader	Programme Leader/ Theme Leader	Kam/QAM
Self-Assessment & Self Assessment Report	The self-assessment is performed by the model development team. The model steward and MDAM can provide support. The self-assessment report is drafted by the model steward based on the filled in form	Support filling in of self assessment.	Support filling in of self assessment. Generate report. Require periodic update			Perform self assessment				
Self Assessment Report A - D	The self assessment report is the basis for the planning phase, the application projects, the database and the audit	D: If self assessment indicates sufficient quality. Plan audit	C: Store self assessment report in MDI database			A: Use self assessment as input for planning	B: Use self assessment report as gate-keeper. Based on self assessment judge if model is applicable and of sufficient quality			
QA/QC										
The QA/QC of a model is partly done during model development (self-assessment), partly done externally. During this cycle the model quality is first assessed by the development team, then audited by external auditors										
Audit	During the audit the self-assessment is checked by two independent auditors	Plan audit, process overview	Support process	Perform audit	Perform audit	Be audited				
Risk Assessment	During the audit for each non-compliance the risk of non-compliance is discussed			Discuss risk	Discuss risk	Discuss risk				
Audit Report	Based on the audit and risk-assessment an audit report is drafted and quality status is granted	Grant status	Store audit report in MDI database	Draft report	Draft report	Comment on report				

5. Background on quality of models and datasets

5.1 Need and perspectives

The need for thorough and complete documentation of models/datasets has been discussed in many places (e.g. van Voorn et al. 2016, and references therein). For this purpose, many checklists have been designed. The current checklist is aimed at quality assurance of models developed within Wot/Natuur & Milieu and Wageningen Environmental Research – both WUR. For in depth discussion on the background of the predecessors of this checklist, please check Jansen et al. 2004, Tiktak et al. 2013 and Houweling et al. 2015. Following Houweling et al. (2015), the main aim of this checklist is to provide input to a quality assurance system that minimises the risks of:

- A. Public discussions on the quality of models applied for a specific application.
- B. Inefficiencies in the (policy) advice process due to shortcomings in modelling.

Three perspectives on quality have been deduced from this:

- 1. Precise scientific and technical documentation, testing and validation of the model/dataset operation.
- 2. Alert organisation of the management of the model/dataset.
- 3. Transparent communication on the applicability, use and interpretation of the model/dataset and results.

These three perspectives partly overlap, but take a different angle at the quality of the model or dataset. For perspective 1 the audience is internally within the institute and scientific community. Working towards this aim the model developers ensure that model/dataset results are scientifically and technically sound and that technical work and model/dataset results can be easily shared within the organisation or the scientific community at large.

Perspective 2 focusses on the project-planning around the model/dataset. With this aim the organisation anticipates future use and development of the model/dataset, by taking ownership, by ensuring continuity where external or internal dependencies demand so and by systematic planning and quality assurance of the development and exploitation.

For perspective 3, the target audience are those people that make use of the model/dataset results down the chain of the (policy) advice process. This can include people running a script or making use of the model/dataset through a graphical user interface, and extends to stakeholders reading numbers from calculations in a policy brief. From the documentation this audience should be able to interpret the results and assess the applicability and (un)certainity associated with the model/dataset-based advice.

5.2 Levels of Quality

This work instruction acknowledges five levels of 'quality' of a model/dataset.

Undefined is the level of quality for models for which no self-assessment has been performed. No information is available about the quality status of the model.

Self-assessed is a level of quality for a model/dataset of which the model/dataset developers have provided references for all requirements in the checklist and assessed the coverage of the criteria as sufficient. No external assessment of these references with respect to the criteria set has been performed, or the provided references are assessed in an audit to be insufficient to meet the quality criteria for Status A.

Status A is the minimal level for a model/dataset to be deemed 'ready for use'. This level assures that there is a basic scientific documentation of the model/dataset, that the documentation is clear about the level of evaluation of the model/dataset and its implementation, that the organisation is aware of the existence of the model/dataset and that the end users receive guidance in the interpretation of the results of the model/dataset.

Status A+ (Jansen et al. (2004), elaborates on status A with requirements set by specific stakeholders. One example is the set of requirements set by the Netherlands Environmental Assessment Agency (Tiktak et al. 2013) for models/datasets used in environmental assessments. To obtain the level A+ a predefined set of AA level requirements must be met.

Status AA elaborates on A. This level assures further precision in the scientific and technical documentation, more in depth evaluation of model/dataset performance, cyclic development processes considering results from previous evaluations and formalisation of the organisation around the model/dataset.

The quality status A, A+ and AA are to be reaffirmed after 5 years. In these 5 years, scientific and technical developments are expected to affect the use of the model/dataset.

5.3. Checklist structure

The checklist is structured following the three perspectives as the main themes:

1. Science and Technology
2. Development and Organisation
3. Interpretation and Use

For each of these themes, requirements are formulated within several topics. Each of the requirements addresses an issue to be covered by documentation. This has resulted in a total of 22 requirements split over 7 topics. Each requirement can subsequently have an operationalisation at A and AA level. The balance between the themes is uneven: the theme science and technology are covered by 4 topics and 14 requirements, the theme interpretation and use by only 1 topic and 2 requirements.

The choice has been made to present the criteria operationalising the requirements in an elaborate way. Each criterion consists of several items to be covered in the model/dataset documentation. This choice was made to avoid an overkill at first view of the checklist. The current checklist and self-assessment form can be found on the intranet group page.

5.4 Documentation

The requirements in the checklist refer to documentation on specific issues about the model/dataset. In general, we aim at six types of documentation: Scientific documentation, Technical documentation, Development plan, Management plan, Interpretation guide, User manual. Each of these addresses a specific community of stakeholders. These documentations need not be separate documents and need not be singular, and frequent cross-referencing can be beneficial. However, high fragmentation of documentation beyond these six types can hinder readability and comprehension.

6. Glossary

- MDAM Models & Data Assurance Manager
- MDI Models, Datasets and Indicators
- QAM Quality Assurance Manager
- QA/QC Quality Assurance / Quality Control

7. References

- Houweling, H., G.A.K. van Voorn, A. van der Giessen & J. Wiertz (2015). Kwaliteit van modellen voor wettelijke onderzoekstaken. WOt-paper 38. WOT Natuur & Milieu, Wageningen; Biometris, Wageningen; PBL, Bilthoven.
- Jansen, J.M.L., J.M. Halbertsma, J.A.P. Heesterbeek, H. Houweling & M.J.W. Jansen (2004). Kwaliteitsborging databestanden en modellen - Balanceren tussen chaotische dynamiek en geordende stilstand. Alterra-rapport 956. Alterra, Wageningen.
- Tiktak, A., A. Beusen, R. Loke, J. van Vliet M. de Vos & P. Cleij (2013). Het PBL-Normenkader voor modellen. PBL, Den Haag.
- van Voorn, G.A.K., R.W. Verburg, E.-M. Kunseler, J. Vader & P.H.M. Janssen (2016). A checklist for model credibility, salience, and legitimacy to improve information transfer in environmental policy assessments. *Environmental Modelling & Software* 83, 224-236.