Data Management Plan

Wind at Sea Ecological Programme (Wozep)

Rijkswaterstaat WVL

28 March 2023

Project	Data Management Plan
Client	Rijkswaterstaat WVL
Document	Wind at Sea Ecological Programme (Wozep)
Status	Final version
Date	28 March 2023
Reference	115621/23-005.347
Project code	115621
Project Leader	C.H. Clemens MSc
Project Director	R. Bouw MSc
Author(s)	C.H. Clemens MSc J.H. Boersma MSc
Checked by	C.H. Clemens MSc
Approved by	R. Bouw MSc
Paraph	

Address

Witteveen+Bos Raadgevende ingenieurs B.V.
Leeuwenbrug 8
P.O. Box 233
7400 AE Deventer
The Netherlands
+31 570 69 79 11
www.witteveenbos.com
CoC 38020751

TABLE OF CONTENTS

1

INTRODUCTION

1.1 Wind at Sea Ecological Programme (Wozep)

In 2015, the Ministry of Economic Affairs decided to set up an integrated monitoring Programme to study gaps in our knowledge relating to the impact of offshore wind farms on the ecosystem of the North Sea. This generic Programme was established in response to a recommendation from Rijkswaterstaat, based on the findings from research and monitoring of 'Round 1 and 2 wind farms', that the knowledge gaps are primarily generic rather than specific to individual wind farms. This led to the establishment of the Offshore Wind Ecological Programme (Wozep).

Wozep started in 2016 with 'no regret' research. This includes feasibility studies, international inventories for data and knowledge and pilot research with field observations. The Programme continues as a monitoring and research Programme for the coming years (> 2030). The research focusses on the ecological effects of windmill farms om specific species, such as birds, bats, sea mammals, benthos and fish.

1.2 Data management within the Wozep Programme

This Data Management Plan (DMP) describes the data management for the Dutch Wind at Sea Ecological Programme (Wozep). Under the Programme different projects will generate data. This DMP describes how data management will be applied on all the datasets, including scripts and models, that have been and will be acquired by Rijkswaterstaat during the course of the Programme. The goal is to achieve transparency of research data streams and (future) re-use.

1.3 Data management principles

The following general data management principles apply for the Wozep Programme:

- We encourage to make data F.A.I.R. (Findable, Accessible, Interoperable, Reusable) and openly available where possible.
- All datasets gathered for the Programme, are stored, maintained and in time available as open data for users.
- All data in the data management system (DMS) is processed according to the agreed data management procedures as described in the data management plan (DMP).
- All datasets are described in a specific data management factsheet that includes all details about acquiring, processing and storing.
- All persons who are active in this project are bound by the agreements in the DMP and by the tasks and responsibilities appropriate to their role.

1.4 Reading guide

This Data Management Plan (DMP) contains the following chapters:
- Introduction (chapter 1).
- Organisation: roles and responsibilities (chapter 2).
- Acquiring, processing and storing data (chapter 3).
- Explanation of the data management system (chapter 4).
- Agreements on ownership, reuse and access to data (chapter 5).
- Helpdesk and support (chapter 6).

Appendix I is an A3-version of the schematic overview of data management phasing. Appendix II a dummy version of the Data Management Factsheet.

During the duration of the project, data management will be subject to change. The DMP can therefore be regarded as a living document, which must be evaluated at least once a year and, if necessary, adjusted.

2

ORGANISATION - ROLES AND RESPONSIBILITIES

2.1 Introduction

Under the Wozep Programme different projects will generate data. The parties involved in data management and their tasks are described in this chapter.

2.2 Data management roles explained

2.2.1 Rijkswaterstaat WVL - commissioner data management

Programme team - data management

Ingeborg van Splunder (Programme manager), Kees Borst (project leader data management) and Dagmar van Nieuwpoort (technical manager), from Rijkswaterstaat WVL are responsible for the organization of the data management for the Programme. This includes:

- Organizing centralized data management as part of the Programme.
- Formulating requirements to introduce data management in every project within the Programme.
- Acceptance of the project data for reuse and archiving.
- Authorization of data that's put on-line to users outside of the Programme.

RWS project managers

The projects within the Programme are managed by project managers from Rijkswaterstaat. These project managers are responsible for the contract with the project parties, including implementation of the data management within the project, the quality control of the data delivery combined with the delivery of a report / scientific paper.

2.2.2 Witteveen+Bos - data stewardship

Witteveen+Bos (W+B) has been commissioned by Rijkswaterstaat to carry out the data management for the duration of the Programme (2020 - 2024). The Witteveen+Bos data stewardship team is responsible for the following aspects:

- Realisation And Maintaining The Data Management System (DMS).
- Write And Update A Data Management Plan (DMP), With All Organizational, Technical And Operational Aspects Of The Data Management.
- Facilitate Data Suppliers In Providing An Helpdesk, And Give Support And Training. This Will Be Done By A Data Stewardship Team From Witteveen+Bos That Is Available To Help Data Suppliers With The Preparation Of The Data For Use In The DMS And In Line With The DMP.
- Support The Data Suppliers In Delivering And Storing Data In The DMS.
- Check Compliance Of Every Data-Delivery With The Agreements On Data-Delivery, Structuring And Processing Towards Standard Data As Described In This DMP.

2.2.3 Project parties - data supplier (GIV)

Within the Programme the data is gathered by a contracted project party named a ‘data supplier’ or GIV. A data supplier has the following responsibilities:

- The data supplier is responsible for gathering and delivery of the by Rijkswaterstaat requested datasets and/or models to a data repository (a location to store datasets).
- Delivery of documentation of the data (in the data management factsheet) such as reports or scientific papers.
- Delivery of scripts/software/models used to transform data to project data or a requested standard.
- Delivery of the data to the data stewardship team, or storing the data in the DMS.

2.2.4 Data storage parties (internal and external)

The Programme wants to store the data gathered in an open access database that is continuously managed, to make sure the data can be reused. The long term storage is provided by RWS-CIV or by external data management parties. The external partiers are institutes with specialized data-platforms that are contracted by RWS to store Wozep-datasets outside the Wozep DMS.

2.2.5 Data users (internal and external)

The data that is generated within the Wozep Programme and considered ‘open data’ and can be used outside of the Wozep Programme, by other organizations, such as the Marine Data and Information Center (Informatiehuis Marien), scientist or consultants. A data request can be sent to the helpdesk (see chapter 6, Helpdesk and support).

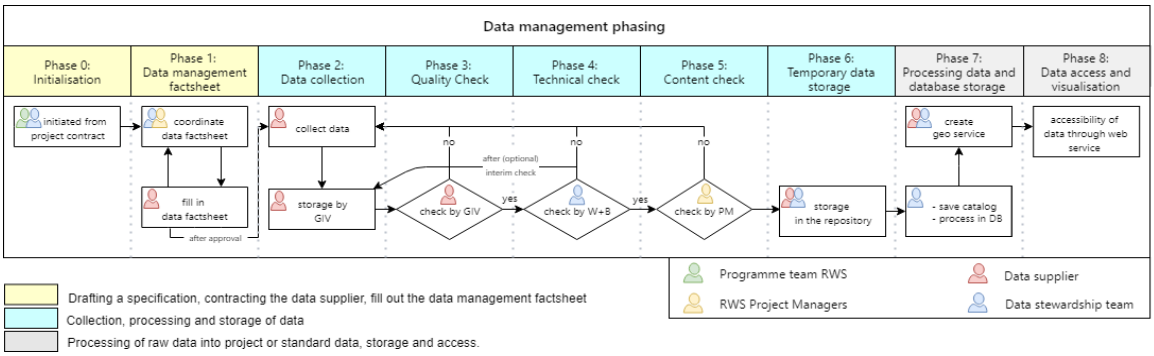
3

ACQUIRING, PROCESSING AND STORING DATA

3.1 Introduction

This chapter describes the phases of data collection, processing, documentation and storing for a project within the Programme. It describes the phases that the data must go through to be suitable for reuse. A general overview of these phases is presented in figure 3.1. A more detailed description of the steps within the data management phases and a larger version of the schematic overview is available in Appendix I.

Figure 3.1 Schematic overview of data management phasing



The phases of data management and corresponding actions are described in the remainder of this paragraph. Note that phases 2 to 8 could be an iterative process where data is delivered in batches *during* the data collection phase. The data management process can be divided into three parts:

- The start of the data management process (phase 0 and 1). For each dataset that is obtained (from an external party, a regular database of Rijkswaterstaat or collected on behalf of Rijkswaterstaat), a data fact sheet must be filled in for accountability.
- Collection, processing and initial storage (phases 2 to 6), i.e. the collection process. In these phases, the measurement parties and support staff in particular work on the collection of raw data and initial storage in version control system (repository) that is part of the DMS.
- Phases 7 and 8 are aimed at further processing the data into project data or standard data for a specific purpose and user: for example, the realization of a project dashboard, or access via a website.

3.2 Start-up of the data management process

The Wozep Programme works with a multi-year plan. For each years there is an overview of contracts that are issued. After the procurement process produces a data supplier for the execution of the contract and is finished, the data management process starts.

3.2.1 Phase 0: Initialisation

The data management process is initiated by the Programme team (RWS) and Witteveen+Bos based on a contract between RWS and the GIV. During this phase RWS provides documentation with background information about the project (e.g. contracts, implementation plans, planning, etc.) to W+B.

3.2.2 Phase 1: Immediately after award

Data management factsheet - draw up a first version

To get an up to date overview of data generated in the projects under the Programme, it is necessary to fill out a project-specific data management factsheet. See the dummy data-factsheet in Appendix II. The data management factsheet describes the project details, contact persons and specifications about each dataset used, such as data format, additional use of datasets from other sources and required storage. This ensures that datasets can be traced back to the project and contract number, and provides an overview of datasets to deliver to the DMS. The factsheets are used as a logbook, so if during the project something changes that is relevant for the data, this is included in the factsheet. W+B will fill out a first version of the data management factsheet in preparation of the start up meeting.

Plan a start-up meeting (PSU)

W+B plans a project start up meeting (PSU) about data management with the GIV and the client (RWS) (project manager for the contract and manager data management).). The following people should be present:

- RWS: the project lead; Kees Borst.
- GIV: project manager and the contact person for data delivery.
- W+B: two data stewards.

Approval of the final data-factsheet

Prior to or after the PSU, W+B provides the GIV a draft of the data management factsheet per email, based upon the documentation received in phase 0. The GIV is requested to provide all additional information required to complete the data management factsheet. Completion of the factsheet is an iterative process between W+B and the GIV, coordinated by the W+B data manager. When the final version is ready, W+B will request approval from the project manager and the GIV.

3.3 Collection, processing and initial storage

3.3.1 Phase 2: Data collection phase

Data is collected by the GIV throughout the project. Storage of the data is in accordance with the data management factsheet. For the storage of data and scripts a subversion repository is available within the DMS. W+B provides the GIV with a standard format to fill out by the GIV for delivery of the data.

The data supplier (GIV) has to deliver the data during or at the end of the project to the SVN-server or (for large files) to the FTP-server. The data manager views/checks the (large) datasets/scripts as placed in the Azure File Share through the FTP-server and after approval uploads the data to the SVN-server.

3.3.2 Phase 3: Quality check

The acquired data is quality checked by the GIV. Before entering phase 4 the GIV makes sure that the (batch of) raw data, accompanying metadata, processed (tailored) data and scripts are delivered in accordance with the data factsheet. It is expected that additionally a technical report is available. Both the datasets, scripts and the report is delivered by the GIV to W+B for phase 4.

3.3.3 Phase 4: Technical check

W+B checks whether the delivered data (and scripts) matches the data factsheet. This phase asserts that the correct file formats are delivered, that the data is in accordance with the agreed formats, that the documentation of metadata is available and correct, and that the metadata matches the supplied data.

In addition the data manager checks the reprocessing scripts and adjust where necessary. Both scripts and data are placed in a dedicated folder within the tree-view structure of the SVN-repository.

3.3.4 Phase 5: Content check

The project manager performs the final check on the contents of dataset before it is approved for definitive storage in phase 6.

3.3.5 Phase 6: Data storage

The GIV and W+B make sure that the (batch of) data is stored in the WOZEP repository, or that a reference is made if the data is stored elsewhere. W+B makes sure that the data ends up at the correct location and structure within the repository.

3.4 Standardization and storage for access

3.4.1 Phase 7: Processing data (not applicable to all datasets)

Selected datasets are upgraded during this phase. For instance, upgrades may involve transforming data to a specific data format in order to make it available through a GeoServer. The selection and upgrading process is a joint task between W+B and the GIV.

3.4.2 Phase 8: Data access (not applicable to all datasets)

Preferably the data is made available through services and added to a viewer such as ‘Waterinfo extra’ or ‘Informatiehuis Marien’. Coordination and realization of this phase is still required.

3.5 Deletion of data

The phasing of data is included from initialisation of a WOZEP project, up to the storage and accessibility of the data in the repository. The concluding step of archiving the data at the final end of WOZEP is not included in the scheme, but could be considered ‘phase 9’.

At the end of the data management by W+B for Wozep the repository is archived. Preferably the archive storage medium would be the same as for Waterinfo-extra and Informatiehuis Marien. Coordination and realization of this phase is still required.

4

EXPLANATION OF THE DATA MANAGEMENT SYSTEM

4.1 Introduction

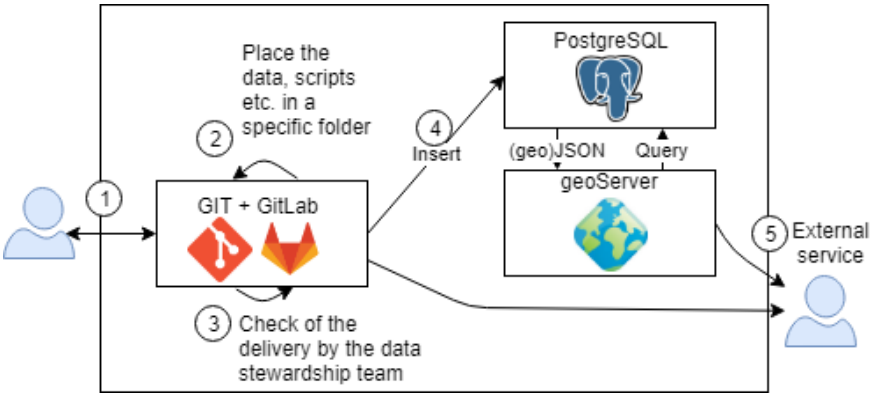
This chapter describes the Wozep Data Management System (Wozep-DMS). Main goal of the Wozep-DMS is to store the data in a central location, that includes version control, for the duration of the Wozep Programme. This to get a transparent routing from (raw) data to analysis and reporting, reproducibility of results and good opportunities for multi-thematic use and future reuse.

4.2 The Wozep-DMS Design explained

The design of the Wozep-DMS (reference date November 2022) is schematically shown in figure 4.1. The DMS is provisioned on a virtual machine in the Microsoft Azure Cloud. The following enumeration is an explanation for the numbers in the figure:

1 The Wozep-DMS is accessible through a website (see section 4.2).
2 The data supplier can get access to a GIT-repository by signing in. After signing in the data supplier can place data and scripts in a specific folder (see figure 4.3).
3 The delivery is then checked by the W+B data stewards.
4 The scripts are executed to transfer data to a PostgreSQL-database.
5 A data user (or external service) can request datasets via a Geoserver, or consult the data by using GitLab.

Figure 4.1 The Wozep Data Management System (DMS)

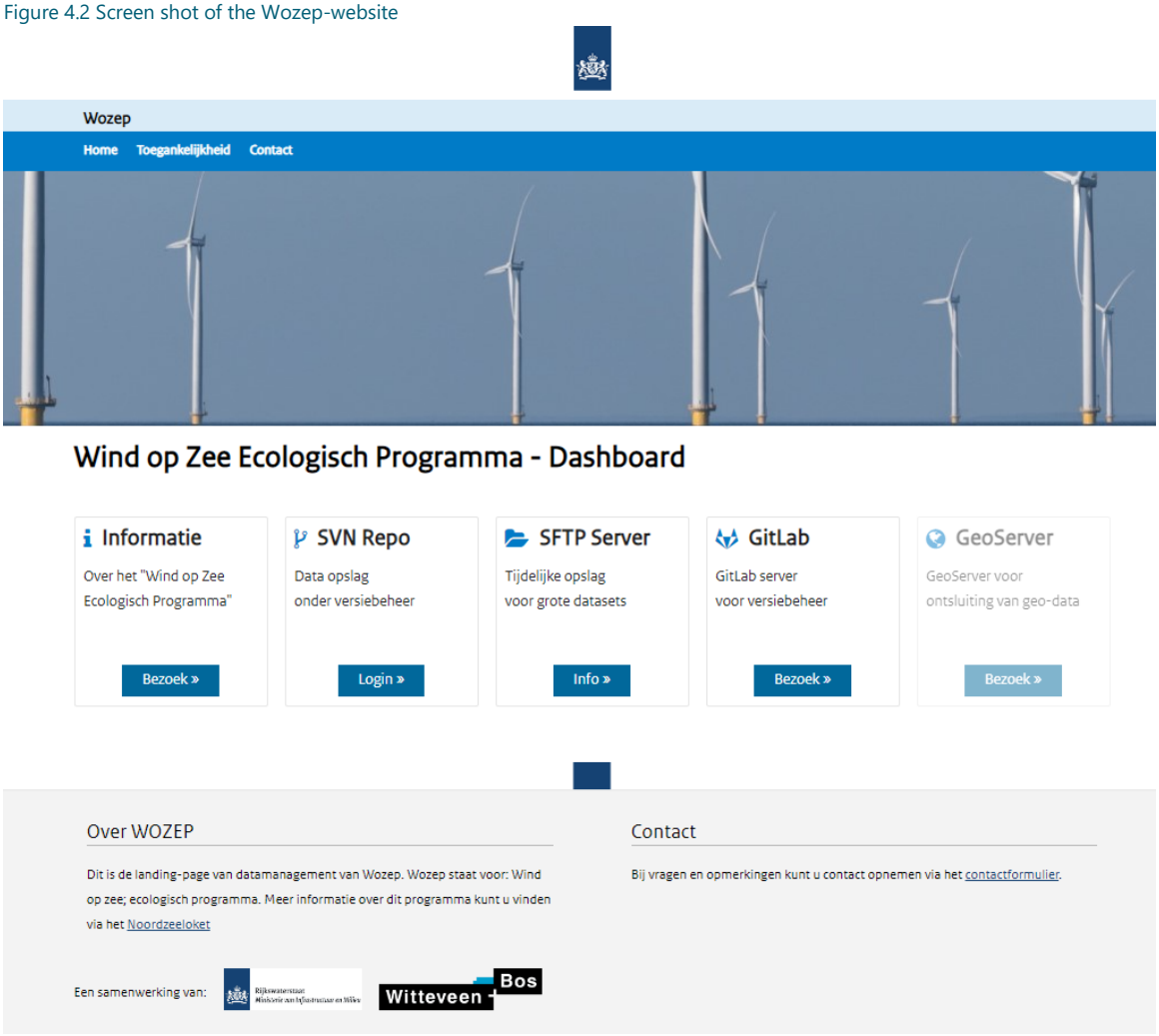


4.2.1 Website

The Wozep-DMS is accessible via https://wozep.nl/ Via this route you can find buttons to get access to GitLab and the previous version of the SVN-repository. The Geoserver is not in use.

We are preparing an additional page with additional information about the data that is available from the Wozep Programme.

Figure 4.2 Screen shot of the Wozep-website

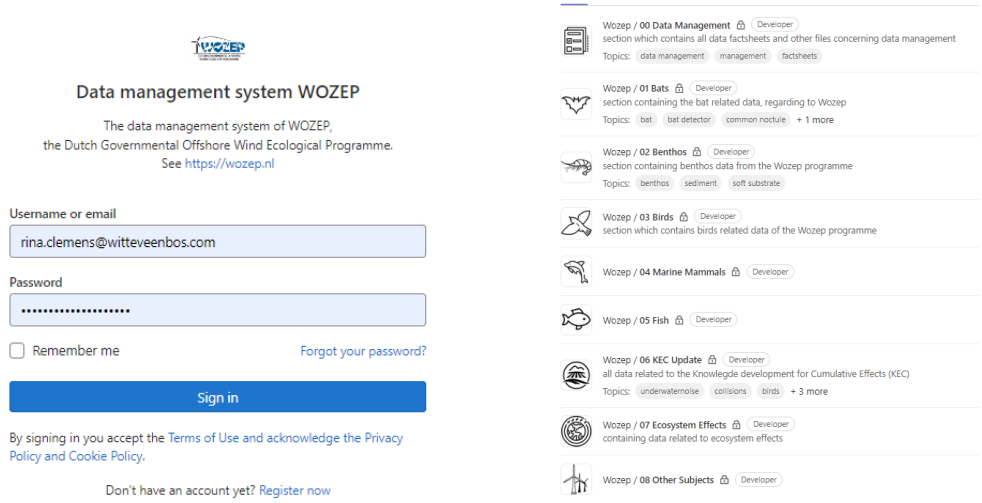


4.2.2 GitLab server

The GitLab server provides the functionalities: data storage, data processing and version control. The repository forms the core of the data management system. In this repository the data (raw, project, standard, model), scripts and documentation are stored and maintained.

The GitLab server can be accessed by the users (data supplier, RWS Project Manager) for the delivery of data, scripts and documentation. To get access you will need an account. This account can be provided by Witteveen+Bos. You can then log-in to access the folder structure, as shown in figure 4.3.

Figure 4.3 Access to and folder of the GitLab-repository



4.2.3 PostgreSQL-database

The PostgreSQL database server provides the functionality: data management. Datasets that are available via the Git repository are loaded according to the data model. The database server is set up with user management and linked to GeoServer.

4.2.4 Geoserver

The Geoserver provides the functionality: offering other data files via web services. Via the GeoServer, the data in the PostgreSQL database is accessed as WMS and/or WFS services. It is also possible to query the PostgreSQL database using a REST API. This allows users to retrieve specific data from the database and receive the information in a processing tool of their choice. The Geoserver is not in use.

4.3 Data storage on hard drives

In some projects extremely large and unique datasets are acquired. There is a procedure to make sure these datasets are placed on separate hard drives and then stored in the safe of RWS-CIV. When applied within a project, this will be described in the data management factsheet.

4.4 Information security, cloud-management and maintenance

Information security

Information security is implemented. Every user receives their own and unique login (username - password) for access to the DMS. This ensures that the environment is safely accessible via the internet. Passwords are stored encrypted on the server and can therefore not be retrieved by the administrator or the user. The password can be reset by both the administrator and the user.

The virtual machine runs locally-redundant, with full backup functionality. The environment is backed up daily, weekly and monthly during the project. This backup is located at multiple locations in the Netherlands. In addition all network traffic to and from the DMS is secured.

Cloud-management and maintenance

Maintenance is continuously carried out by a cloud service provider. This includes, among other things, a 24/7 helpdesk, monthly assessment of the performance of the DMS and proactive advice to prevent technical failure of the environment.

5

AGREEMENTS ON OWNERSHIP, REUSE AND ACCESS TO DATA

This chapter describes the administrative and legal requirements for datasets that are collected as part of the Programme, including ownership, reuse and publication of monitoring data.

5.1 Data ownership

The ownership rights of the collected data lie with the Commissioner data management (RWS) and/or financier of the measurement effort. This means that the data ownership rights of the data often lie with Rijkswaterstaat. Sometimes this ownership is shared. This also applies to standard and reprocessed data, as well as the means developed to arrive at these products, which have been reprocessed on behalf of or financed by Rijkswaterstaat. In addition data is also purchased for single use. The ownership of this data lies with third parties. The owner of the data is described in the data management factsheet, and in the metadata of a dataset.

5.2 Availability of data

In principle, all data are only available for use by the parties within the partnership (Rijkswaterstaat, Data Manager and Data supplier (GIV)-parties of the various subprojects) or by explicit permission granted by the Project Manager of Rijkswaterstaat. Eventually all data will become publicly available. In addition, tailor-made agreements can be made about earlier public availability or later (e.g. in connection with the time required for publication).

5.3 Data use rights

Right of use is distinguished for three roles:

a. Data owner.
b. Data supplier (GIV).
c. Third parties.

a. The data owner has the unlimited (ownership and) right of use at all times.
b. The data supplier (GIV) party is not a data owner. In the period that data is not yet publicly available, the data supplier has the unlimited right to use all Rijkswaterstaat data that is necessary and available for the research carried out on behalf of Rijkswaterstaat in the context of the Wozep Programme.
c. During the period that data is not yet publicly available, the data management system (DMS) shall keep a list of persons who have access to the data. Access to the data (e.g. to students) is granted for work within the framework of Wozep or by explicit permission given by Rijkswaterstaat for other initiatives. The permission is given for a set period of time.

The data is exchanged through the DMS. A Data requests (access to the DMS) can be made by the data management team of Witteveen+Bos (see the contact details in Chapter 6), as they will ask for permission to give access from the Project Manager of Rijkswaterstaat. When permission is given, the data will be made available for re-use. The following terms and conditions then apply for the party that requested the data, regarding the use and publication of data as mentioned:

- It's not permitted to use the collected data for projects other than WOZEP prior to the general availability (open access) of the data.
- In addition, you do not acquire any right (including copyright) to ownership, use and publication of data (raw, but also processed data by into standard and processed data).

5.4 Sharing and publications of data

Rijkswaterstaat and GIV parties have entered into a partnership to carry out demand-driven research in an academic context. Agreements on publication are made between RWS and GIV parties. It's important that datasets are open access (with a DOI). A GIV can make a request to publish the data in a scientific data publication. This will then be registered in the data management factsheet. In this case we request the GIV to also mention the Programme and Rijkswaterstaat as the data-owner.

6

HELPDESK AND SUPPORT

Witteveen+Bos provides a helpdesk for the duration of the Wozep Programme, that includes support and training. A data supplier can contact the Witteveen+Bos datastewardship team with any question regarding data management:

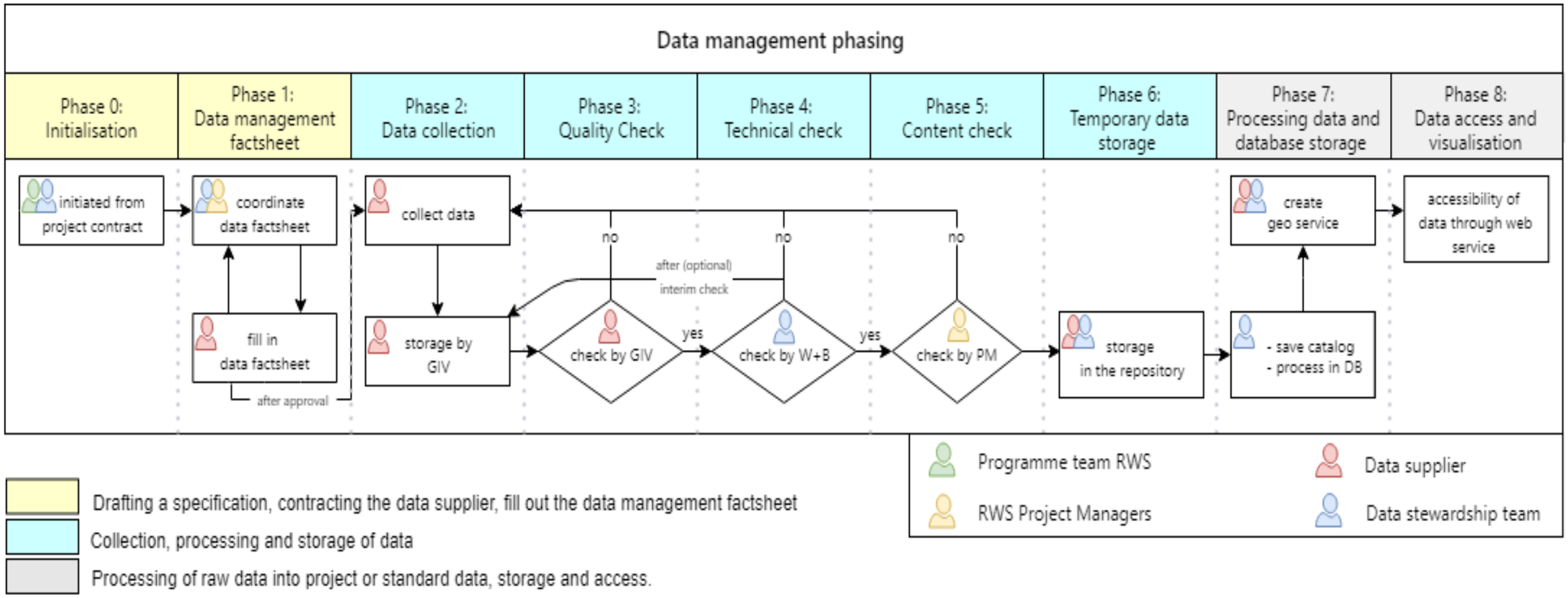
- Jochem Boersma (jochem.boersma@witteveenbos.com, +31 (0)6 86 80 16 49).
- Frank Klein Schaarsberg (frank.klein.schaarsberg@witteveenbos.com, +31 (0)6 86 88 04 14).
- Hillianne de Jonge-Buist (hillianne.de.jonge@witteveenbos.com, +31 (0)6 18 01 12 80).

For any additional questions please contact the Witteveen+Bos project manager:

- Rina Clemens (rina.clemens@witteveenbos.com, +31 (0)6 50 22 18 90).

Appendices

APPENDIX: SCHEMATIC OVERVIEW OF DATA MANAGEMENT PHASING (A3-VERSION)


Data management phasing
Phase 0: Initialisation
Phase 1: Data management factsheet
Phase 2: Data collection
Phase 3: Quality Check
Phase 4: Technical check
Phase 5: Content check
Phase 6: Temporary data storage
Phase 7: Processing data and database storage
Phase 8: Data access and visualisation
initiated from project contract
coordinate data factsheet
collect data
create geo service
accessibility of data through web service
no
no
no
after (optional) interim check
fill in data factsheet
after approval
storage by GIV
check by GIV
yes
check by W+B
yes
check by PM
storage in the repository
- save catalog
- process in DB
Programme team RWS
Data supplier
RWS Project Managers
Data stewardship team
Drafting a specification, contracting the data supplier, fill out the data management factsheet
Collection, processing and storage of data
Processing of raw data into project or standard data, storage and access.

APPENDIX: DUMMY DATA MANAGEMENT FACTSHEET

DATA MANAGEMENT FACTSHEET

00. Project details

Programme name	Wind at Sea Ecological Programme (Wozep)
Project name	
RWS-contract nr. (zaaknr)	00000000
Contact details client: - Rijkswaterstaat	<name (position)> <email address>
Contact details contractor: - <name>	<name (position)> <email address>
Contact details data manager - Witteveen+Bos	<name (data manager)> <email address>
Short description of the project	
Project duration	
Factsheet number + status	v0.1 || concept
Date of the factsheet	

01. Data collection

Describe each dataset that will be created/collected and used during the project.

Checklist to make a description of each dataset:

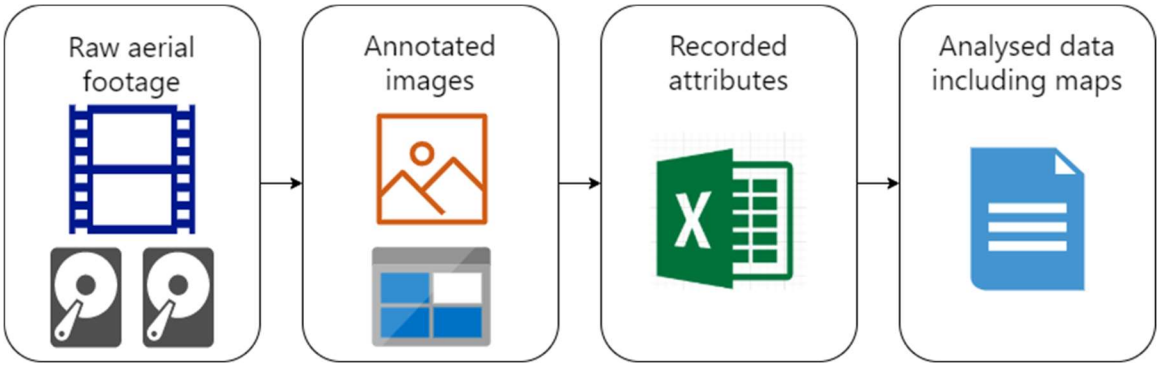
1 What type of (raw)data will be collected? Will you also use pre-existing data? From where?
2 How will data be collected? Can you describe the data collection process? Will it be reproducible? What would happen if it gets lost or becomes unusable later?
3 What is the estimated size of the data? In what file formats?
4 Do the data have a specific character in terms of confidentiality (e.g. privacy), etc.? What does this mean for the management of the data?
5 Which data-standard will be used?
6 Which standards for the metadata will be used? Describe in detail which, and state in which databases/files these will be included.
7 Which tools or software are needed to create/process/visualize the data?
8 How do you handle version control to maintain all changes that are made to the data?
9 State in which existing repository or database the data is stored and which type this is. If available mention the URL.

02. Data storage and back-up

description	type of data	storage medium and location*	file extension
dataset description	raw data, standard data (processed), model data, code, etc.	e.g. wozep repository: /08_domain_name/dataset_name *or:* HDDs at company X	For instance: .csv.

* describe the location in detail. If data is stored in Wozep repository, it should contain the file location.

<include a scheme of the dataflow, to describe the cohesion of datasets, if multiple datasets are collected, for example:



03. Data documentation / metadata records

Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This documentation preferably includes details on the methodology used, analytical and procedural information, and data specific information (parameters and/or variables used, column headings, codes/symbols used, etc.).

This should include metadata records along with data files to describe or contextualize the data, such as: Title, Date of creation, Creator(s), Keywords, Description, and under what conditions it can be accessed. We request the (meta)data documentation to be delivered as a Readme.txt file (a plain text file) to the folder where the corresponding datasets are saved.

04. Analytics / code

In addition to the data and documentation: describe the analytics used (a.i. software) or code that was used. It is required to also deliver the code that was used to standardize the raw data. Please describe the code and in what format this will be delivered.

05. Quality and risks

05.1. Quality

Add a description of the quality checks that will be done on the dataset by: 1) the contractor, 2) data manager W+B and 3) the client. Include how quality checks are documented.

Before the data is send towards the client, the contractor will perform the following checks:
- check on completeness, will all requested data be provided?

After receiving the data, W+B will perform a technical check. This are, e.g.:
- consistent file names and (if applicable) in line with previous deliveries;
- (if applicable: checks on coordinates and reference system);
- further technical checks.

After the technical check from W+B, the project lead of RWS will perform (with assistance from W+B) checks concerning the contents:
- is the data interpretable;

05.2. Risks

Describe specific risks for the data-management.

06. Project specific agreements

Describe if the project receives/uses data that is under confidentiality or other restrictions, so these datasets are identified. This includes third party data with specific limitations for distribution (based on IP or agreements for obtaining the data).

07. Data-access and ownership conditions

Please describe the necessary access conditions for the deposited (meta)dataset, code and documentation. You can choose one off the following access conditions:

- **Open access**: there are no additional restrictions on access to the data or publication of results;
- **Embargo period**: you can request an embargo period, whereby no access to the data would be permitted outside the Wozep-programme, until after the data that is specified.

The conditions under which the data will be made available from the data-repository to other researchers is described in [ref: the memorandum].
Describe the aim and purpose of follow-up usage of data. What is the intentional reusability of this data? Which datasets will be exposed as an open web-service?

08. Planning for delivery

When the data will be delivered to the repository:

No.	Delivery	Delivery date
01		
02		
03		

09. Logs

Log of changes made to this factsheet or dataset:

No.	Description	Date
v0.1 || concept	*Initial factsheet*	2022-03-01

10. Documentation used for this factsheet

Please describe the documentation used for this factsheet, e.g.
1. specification
2. implementation plan
3. reports

ATTACHMENTS