Horizon Europe

Grasshopper

Data Management Plan

Deliverable number: 1.3
Version 1.0

This project is funded by the Horizon Europe / Marie Sklodowska-Curie Staff Exchange under Grant Agreement No. 101182922.

Project Acronym:	GRASSHOPPER
Project Full Title:	Jumping forward from preclinical development of pediatric cancer drugs to clinical implementation
Call:	HORIZON-MSCA-2023-SE-01 (MSCA Staff Exchanges 2023)
Topic:	HORIZON-MSCA-2023-SE-01-01
Type of action:	HORIZON-TMA-MSCA-SE
Grant Number:	101182922
Project URL:	https://grasshopperconsortium.com
Authors:	Linda Schild; Princess Máxima Center for Pediatric Oncology Celina Szanto; Princess Máxima Center for Pediatric Oncology
Work package:	WP1
Deliverable nature:	Report (R)
Dissemination level:	Public (P)
Contractual Delivery Date:	31-05-2025
Actual Delivery Date:	28-05-2025

HISTORY OF CHANGES		
Version	Publication date	Changes
1.0	28.05.2025	Initial version

Table of Contents

1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

The GRASSHOPPER project aims to boost the implementation of new therapeutic interventions for pediatric cancer through international staff exchange and collaborative research. Data collection and generation are central to achieving the project's scientific and translational goals across four interlinked research areas:

1. **Immune interventions** – Data from clinical and laboratory research will be collected to identify and validate novel immunotherapeutic targets and pathways.
2. **Molecular interventions** – High-throughput omics data (e.g., genomics, transcriptomics) will be generated and analyzed to uncover new molecular drivers of pediatric cancer.
3. **Drug repurposing** – Data from pharmacological screenings and preclinical studies will support the identification of existing drugs with potential pediatric oncology applications.
4. **Clinical validation** – Clinical and biological data will be collected from patient cohorts to validate promising therapeutic interventions in diverse European and international settings.

Data are generated during secondments and reflect real-time knowledge exchange between academia, clinical centers, and industry. The integration and harmonization of these data across partner institutions enable:

- Preclinical and clinical validation of therapeutic strategies
- Development of interoperable research datasets
- Strengthening the evidence base for new pediatric cancer treatments
- Capacity building across institutions through shared data practices

Thus, data collection and management are not just technical requirements, but essential enablers of GRASSHOPPER's mission to accelerate innovation and collaboration in pediatric oncology.

Data will be generated during each secondment and harmonized between projects. This is necessary to achieve the objectives. We will collect and harmonize data at the different beneficiaries across Europe. The data collection will comply with all national and European Union legal and ethical requirements.

What types and formats of data will the project generate/collect?

The GRASSHOPPER project will generate a diverse range of research data, including:

- **Clinical Data**: Patient demographics, treatment outcomes, and clinical assessments.
- **Biological Sample Information**: Metadata associated with biological samples collected from patients.
- **Genomic and Transcriptomic Data**: Whole Genome Sequencing (WGS), RNA sequencing (RNAseq) data.
- **Imaging Data**: Histopathology and radiology images.

- **Immunophenotyping Data**: Flow cytometry and other immunological assays.
- **Pharmacological Screening Results**: Data from drug testing and efficacy studies.
- **Computational Data**: Analysis scripts, models, and software tools, code.
- **Model Systems**

Data will be collected in standard, open formats such as CSV, FASTQ, VCF, BAM, to ensure interoperability, reusability, and long-term accessibility across the consortium. FASTQ, VCF and BAM files for genomic data and txt/tsv for clinical data. Possible hail tables and SQL. Clinical data: medical report (pdf). Translational research data: raw and analyzed (vcf).

Will you re-use any existing data and how?

Existing data from previous research projects performed at the beneficiaries' labs will be used together with newly generated data. Additionally, reference gene databases and existing institutional databases (e.g., Biobank data, compound CCI/DKFZ database, R2) will be employed.

The consortium encourages making existing data available for research within and outside the GRASSHOPPER consortium, adhering to the FAIR principles.

What is the origin of the data?

The data is collected at the beneficiaries that are located across Europe.

What is the expected size of the data?

At this moment it is not possible to estimate the size of the data that will be generated within Grasshopper. Members of the Grasshopper consortium do not expect generation of large datasets through data-intensive techniques or formats because the program funds short-term exchanges of staff members.

To whom might it be useful ('data utility')?

Datasets generated by the GRASSHOPPER consortium can be interesting for:

- Scientist in the field
- Clinicians working in the field of paediatric oncology
- Other research projects that are related to work within the GRASSHOPPER consortium

2. FAIR data

2.1 Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

Yes, the GRASSHOPPER consortium will work with unique sample and dataset identifiers. All patient data will be pseudonymized, and unique IDs will be assigned to biological samples to ensure secure traceability. At time of publication, datasets will be made available via public repositories (e.g., Zenodo, EGA, GEO), which assign persistent identifiers such as DOIs. Details on identifier structures will be described in institutional-level data workflows.

What naming conventions do you follow?

Members of the GRASSHOPPER consortium will use the following naming conventions:

- E.g. Project name, subject, date of creation and version number: e.g. GRASSHOPPER_WP7_20210501_v1.csv
- E.g. The file names will be short (max 25 characters) without spaces
- E.g. The file names are independent of the location
- E.g. The file naming convention will also be included in the data management plans of the Grasshoppers.

Will search keywords be provided that optimize possibilities for re-use?

Yes, the search terms are described e.g. in institutional-level data workflows

Do you provide clear version numbers?

Yes, we will apply file versioning. By using the naming conventions as listed above each file name will contain a creation date and version number.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

One or more metadata files are generated for each dataset. The metadata are identified by the same unique ID of the related dataset, with a different suffix/extension.

2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

- Repositories**:** Datasets will be deposited in trusted repositories such as the European Genome Archive (EGA), GHGA (German Version of EGA), national and institutional repositories, Secureshare.
- Access Control**:** Access to sensitive data will be controlled through data access committees and will comply with ethical and legal requirements.

DKFZ: EGA, institutional repositories, GHGA (German Version of EGA)

IC: Clinical data are shared during discussion at the Molecular Tumor Board internal meeting – Institutional repositories

RIGS: national and institutional repositories

NKUA: institutional repositories

IBRI: institutional

CCI: Secureshare, One Drive, local infrastructure

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

As indicated in the Consortium Agreement, each beneficiary has the right to publish the research data that they generate. This implies that partners have to agree with the publication of datasets to which they have contributed. GRASSHOPPER will comply with open access as stated in the Grant Agreement following the HE principle to make all publications open access. This means that the default approach is to make datasets public.

Datasets will be made accessible is possible by deposition in repositories as mentioned above.

What methods or software tools are needed to access the data?

R, SPSS, Microsoft Office, R2, Python

Is documentation about the software needed to access the data included?

Is it possible to include the relevant software (e.g. in open source code)?

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

EGA and GitHub

Have you explored appropriate arrangements with the identified repository?

If there are restrictions on use, how will access be provided?

Yes, this needs to be determined by the beneficiary providing the data. The coordinator request review by the BDAC.

Is there a need for a data access committee?

Are there well described conditions for access (i.e. a machine readable license)?

How will the identity of the person accessing the data be ascertained?

As the DMP is a living document, this section will be added throughout the course of the GRASSHOPPER program. We need to get more information on the datasets before we can discuss how to access the data.

2.3. Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

All data will be made available in standard/open formats compliant with commercial/open software in order to allow as much as possible data exchange between researchers and institutions.

Standard vocabulary for metadata description will be used, in case this will not be possible a mapping of more common ontologies will be provided.

2.4. Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest re-use possible?

Access to the data will often be restricted and needs to be checked by the data access committee of the Beneficiary.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Data will be made available for re-use after publication.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

Yes, if the reuse purpose matches the purpose mentioned in the Informed Consent(s). Data cannot be used for all purposes. Commercial purposes that will be approved are likely to be limited.

How long is it intended that the data remains re-usable?

This depends on the time that is listed in the Informed Consents and might differ between projects. Most data can be re-usable for 10-15 years.

Are data quality assurance processes described?

The DMP is implemented at both the central and local level. At central level, the coordination team (grasshopper@prinsesmaximacentrum.nl) will update the DMP and oversee compliance. At local level, the grasshoppers will add details on datamanagement in their secondment mentoring report.

The supervisors will ensure that the report is aligned at the local level and all GRASSHOPPERs are informed on the DMP.

Further to the FAIR principles, DMPs should also address:

As the DMP is a living document, this section will be added throughout the course of the GRASHOPPER program. We need to get more information on the datasets before we can discuss how to increase the data re-use.

3. Allocation of resources

What are the costs for making data FAIR in your project?

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).

The costs associated with data management activities, including data collection, validation, documentation, archiving, and sharing, will be covered by the research budgets of the participating institutions. Costs related to open access publications can be partly covered by the GRASSHOPPER consortium.

Who will be responsible for data management in your project?

The DMP is implemented at both the central and local level. At central level, the coordination team (grasshopper@prinsesmaximacentrum.nl) will update the DMP and oversee compliance. At local level, the grasshoppers will add details on datamanagement in their secondment mentoring report. The supervisors will ensure that the report is aligned at the local level and all GRASSHOPPERs are informed on the DMP.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

At this moment, no specific arrangements have been made for long term data preservation within the consortium. However, every researcher/PI of each Beneficiary is responsible to take this aspect along in their local data management plan following the infrastructure at the hosting institutes.

4. Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Is the data safely stored in certified repositories for long term preservation and curation?

Besides the GDPR, the Grasshopper members regard privacy and data protection as a fundamental principle and will apply a strict policy on this matter.

We will pseudonymize all patient data, meaning that no data can be directly traced back to the patient.

Data collected during the Grasshopper project will be stored at the beneficiaries. Depending on the type of data multiple storage systems are available. These systems can be accessed from outside the beneficiaries' institute and are backed up by the institutes regularly.

In addition, data will be stored in repositories such as EGA that provide long term preservation and curation.

Multiple local Data Protection Officers (DPO) are involved in data privacy and ethics:

Institute	**Name**	**Contact**
Máxima	Functionaris voor Gegevensbescherming	fg@prinsesmaximacentrum.nl
DKFZ	Ursula Weber	u.weber@dkfz-heidelberg.de
IC	Madeline Casse,	madeline.casse@curie.fr
RIGS	Berit Schwartz	Berit.schwartz.01@regionh.dk
BKUS	Privacy and ethics questions are under the responsibility of CCUH Education and Science Division. A Data Steward will be appointed at CCUH.	
NKUA	Mr. Themistoklis Sioros & Mr. Haris Amorgiano	dpo@paidon-agiasofia.gr
IRBI	Amy DeFelice Amy DeFelice	adefelice@indianabiosciences.org

5. Ethical aspects

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Grasshopper will comply with the EU regulation regarding the protection of personal data. In addition, legal aspects are described in the Consortium Agreement.

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

Yes, this is included in the informed consent.

6. Other issues

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

Each beneficiary has their own departmental procedures for data management.