



Partnership for the Assessment of Risks from Chemicals

Deliverable D7.1 PARC Data Management Plan V1.0 (DMP) WP 7.1

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 2

Technical References

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

This partnership has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101057014

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 3

Work Package	WP 7 - FAIR Data
Task	T 7.1 - PARC FAIR Data Policy (PFDP) and DMP
Dissemination level ¹	PU
Lead Beneficiary / Responsible AE	Lead Beneficiary: RIVM / Responsible AE: TNO
Contributing Participants	EAA (AT), UG-PL (PL), VITO (BE), UBA (DE), UU-IRAS (NL), ANSES (FR), MU (CZ), ISS (IT), EV-ILVO (BE), UZIS (CZ), AUTH (EL), BRGM (FR), JSI (SI), UL-LACDR (NL), KWR (NL), IMM-KI (SE)
Responsible author(s)	Fred van de Brug / TNO / fred.vandebrug@tno.nl Rob Stierum / TNO / rob.stierum@tno.nl
Co-authors	Sandrine Fraize-Frontier / ANSES / sandrine.fraize-frontier@anses.fr Olivier Frezot / BRGM / o.frezot@brgm.fr Barbara Magagna / GFF / barbara@gofair.foundation Erik Schultes / GFF / eriks@gofair.foundation Cecilia Bossa / ISS / cecilia.bossa@iss.it Penny Nymark / KI / penny.nymark@ki.se Tessa Pronk / KWR / Tessa.Pronk@kwrwater.nl Katarína Řiháčková / MU / katarina.rihackova@recetox.muni.cz Lucie Bielska / MU / lucie.bielska@recetox.muni.cz Richard Hulek / MU / richard.hulek@recetox.muni.cz Jildau Bouwman / TNO / jildau.bouwman@tno.nl Sylvia Le Dévédec / UL-LACDR / s.e.ledevedec@lacdr.leidenuniv.nl Iseult Lynch / University of Birmingham / I.Lynch@bham.ac.uk Ondřej Májek / UZIS / Ondrej.Majek@uzis.cz Jan Theunis / VITO / jan.theunis@vito.be Sylvie Remy / VITO / sylvie.remy@vito.be Peter Von der Ohe / UBA / Peter.VonderOhe@uba.de Ivana Huskova / NIVA / Ivana.Huskova@niva.no

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 4

Reviewers	RIVIERE Gilles / anses / Gilles.RIVIERE@anses.fr
	SANDERS Pascal / anses / Pascal.SANDERS@anses.fr
	nina.vogel / rivmnl / nina.vogel@uba.de
	Achilleas Karakoltzidis / auth / achillek@office365.auth.gr
	Susana Pedraza-Diaz, ISCIII (Invitado)
	Emma Westerholm / kemi / Emma.Westerholm@kemi.se
	Ivana Huskova / niva / Ivana.Huskova@niva.no
	Uhl Maria / umweltbundesamt / maria.uhl@umweltbundesamt.at
	Kirsten Baken / vito / kirsten.baken@vito.be
	Liese Gilles / vito / liese.gilles@vito.be
	Urban Boije af Gennäs / kemi / Urban.BoijeafGennas@kemi.se
	Phillipp Schmidt (UBA) (Guest)
	Marlene Ågerstrand (Stockholm University) (Guest)
	Margaux RIOU / santepubliquefrance / margaux.riou@santepubliquefrance.fr
	Ngo Ondřej Mgr. / mzcr / Ondrej.Ngo@uzis.cz
Maria João Silva / insa / M.Joao.Silva@insa.min-saude.pt	
ROUSSELLE Christophe / anses / Christophe.ROUSSELLE@anses.fr	
Anja Duffek (UBA) / Umweltbundesamt316 /	
Sónia Namorado / insa / sonia.namorado@insa.min-saude.pt	
Henriqueta Louro / insa / Henriqueta.Louro@insa.min-saude.pt	
Lina Wendt - Rasch / kemi / Lina.Wendt-Rasch@kemi.se	
Maria João Silva (Convidado)	
Nancy Georgiou (Guest)	
Due date of deliverable	31 10 2022
Actual submission date	19 10 2022

¹ PU = Public

PP = Restricted to other programme participants (including the Commission Services)

RE = Restricted to a group specified by the consortium (including the Commission Services)

CO = Confidential, only for members of the consortium (including the Commission Services)

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 5

Document history

Version	Date	Reviewer name/institution	Short description of changes
0.5	2022-09-16; 2022-09-23	Online workshops with WP71 partners, GFF. MS Teams participants are listed as co-authors in the Technical References table.	Initial draft created by TNO. Changes: Elaboration of context (pre-amble), work processes, numerous refinements.
1.0	2022-09-30	Shared with WP leaders / MB on September 30, 2022 and online MS Teams meeting on October 10, 2022. MS Teams participants are listed as reviewers in the Technical References table.	Comments communicated during the Teams meeting with MB (WP9 training support, elaboration on roles, definition of domain, one invalid link) are resolved on 2022-10-17. The new version has timestamp date in document name of 20221017.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 6

Abstract

PARC is the European partnership for the development of next generation chemical risk assessment methods. In the context of Open Science, the PARC objective OO10 is to implement FAIR data practices and enhance innovation in complex data analysis for chemicals risk assessment. PARC will develop a FAIR data culture enabling open science and provide the technical methods, tools and infrastructure for effective exchange of data and information. Innovation in risk assessment and risk management which is required for the transition to next generation risk assessment can only be achieved in an open and collaborative way.

FAIR and open data sharing according to the 'as open as possible, as closed as necessary' principle will be the default for PARC and will be applied to all research outputs (including reports, data, software, guidelines, method descriptions, formats, templates, semantic artefacts, ontologies, vocabularies etc.).

This deliverable presents version 1 of the PARC overarching data management plan (DMP). The domain specific guidance in version 1 is based on the analysis of DMP's of a selection of projects covering research data domains PARC is concerned with: exposome research, biomonitoring, innovations in (eco-)toxicological hazard assessment, e.g., via in vitro new approach methodologies, nanomaterial safety and metabolism disrupting chemicals.

The DMP contains guidance to support PARC researchers in multiple ways: by providing additional information; by providing generic texts and by suggesting possible domain specific or generic choices.

Keywords

Data management plan, FAIR, risk assessment, open science, exposome, biomonitoring, hazard assessment, nanomaterials, safety

“Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.”

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 7

Table of contents

Technical References	2
Document history	5
Abstract	6
Key Words	6
Table of contents	7
Glossary	9
Preamble to PARC Initial Data Management Plan	12
Introduction	13
1. Data Summary	16
1.1 What is the purpose of the data collection/generation and its relation to the objectives of the project?	16
1.1.1 The overarching PARC DMP and the relationship with project-level DMPs.....	17
1.2 What types and formats of data will the project use?	20
1.2 Which data(sets) will the project generate or collect?	21
1.3 Will you re-use any existing data and how? What is the origin of the data?	22
1.4 What is the expected size of the data?	22
1.5 To whom might it be useful ('data utility')?	23
2. FAIR data	24
2.1. Making data findable, including provisions for metadata	24
2.1.1 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)?.....	24
2.1.2 What naming conventions for files do you follow?	24
2.1.3 Will search keywords be provided that optimize possibilities for re-use?	25
2.1.4 Do you provide explicit machine-readable version numbers?	25
2.1.5 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.	25
2.2. Making data openly accessible	27
2.2.1 Which data produced and/or used in the project will be made openly available as the default?.....	27
2.2.2 How will the data be made accessible (e.g. by deposition in a repository)?.....	28
2.2.3 What methods or software tools are needed to access the data?	28



D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 8

2.2.4 Is documentation about the software needed to access the data included?	28
2.2.5 Is it possible to include the relevant software (e.g. in open source code)?	28
2.2.6 Where will the data and associated metadata, documentation and code be deposited?.....	28
2.2.6 Have you explored appropriate arrangements with the identified repository?	29
2.2.7 If there are restrictions on use, how will access be provided?.....	29
2.2.8 Is there a need for a data access committee?	29
2.2.9 Are there well described conditions for access (i.e. a machine readable license)?	29
2.2.10 How will the identity of the person accessing the data be ascertained?	29
2.3. Making data interoperable.....	30
2.3.1 Are the data produced in the project interoperable?	30
2.3.2 What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?	30
2.3.4 Will you be using standard vocabularies for all data types present in your data set, to allow interdisciplinary interoperability?	32
2.3.5 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?.....	32
2.4. Increase data re-use (through clarifying licences).....	33
2.4.1 How will the data be licensed to permit the widest re-use possible?.....	33
2.4.2 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.	33
2.4.3 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.....	33
2.4.4 How long is it intended that the data remains re-usable?	34
2.4.5 Are data quality assurance processes described?	34
3. Allocation of resources	35
3.1 What are the costs for making data FAIR in your project?.....	35
3.2 How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).	35
3.3 Who will be responsible for data management in your project?	35
3.4 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)?	35
4. Data security.....	36
4.1 What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?.....	36
4.2 Is the data safely stored in certified repositories for long term preservation and curation?	37
5. Ethical aspects	38
5.1 Are there any ethical or legal issues that can have an impact on data sharing or data visiting?	38
5.2 Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?	38

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 9

6. Other issues 40

- 6.1 Do you make use of other national/funder/sectorial/departmental procedures for data management?
If yes, which ones?..... 40

Glossary

COMPSAFENANO	CompSafeNano drives the development of integrated and universally applicable nanoinformatics models, with broad domains of applicability across nanomaterials compositions and forms, that are directly usable by industry, especially SMEs, regulators for risk assessment and decision making (COMPSAFENANO).
Data Champion	Data Champions are appointed by each PARC project to “champion” the need for proactive data management and feed issues that arise in the data management back to WP7. FAIR data champions are scientific experts and are hands-on in the field of FAIR data. The Champions work as FAIR ambassadors, sharing FAIR implementation stories, enhancing synergies, contributing to training activities and webinars, and encouraging cross-domain engagement with FAIR (see also section 1.1.1 in this document and the WP7 Role description).
Data Liaison	Data Liaisons from within WP7 have overarching data management knowledge (and receive additional training) and preferably domain-specific knowledge that are appointed to “support” the PARC projects and Data Champions in developing their data management plans. The data liaison is a key person in use case identification based on the PARC project descriptions of the WPs 4,5,6 and 8 (see also section 1.1.1 in this document and the WP7 Role description).
Data Steward	The research data steward, positioned at the research institution, supports and works in close collaboration with the main data producers and users in academia: the researchers, ranging from undergraduate students to full professors. The data steward advises researchers, makes sure data is handled in a manner compliant with the institute’s policy and may also perform hands-on work in a project (ELIXIR).
Domain	The term domain refers to the scientific domain of activity, e.g., environmental monitoring versus human biomonitoring, toxicity / ecotoxicity, etc. If projects are clustered per scientific domain, then the term domain closely aligns with a cluster of projects.

EUROPEAN PARTNERSHIP



Co-funded by
the European Union

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 10

EPHOR	The Exposome Project for Health and Occupational Research will lay the groundwork for evidence-based and cost-effective prevention for improving health at work, by developing a working life exposome toolbox (EPHOR).
European Open Science Cloud (ESOC)	The ambition of the European Open Science Cloud (EOSC) is to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and reuse data, tools and services for research, innovation and educational purposes (EOSC).
Fair Implementation Profile (FIP)	The FIP is a collection of machine-actionable FAIR implementation choices made by a community of practice for each of the FAIR Principles. Community specific FAIR Implementation Profiles are themselves captured as FAIR datasets and are made openly available to other communities for reuse (GO-FAIR).
FAIR Implementation Taskgroup (FIT)	The tasks and responsibilities of the FIT will include all of the activities outlined above for Data Liaisons plus the following additional aspects: Receive significant training on the technical processes including FAIR Implementation Profiles (FIPs) and Metadata for Machines (M4M), such that they gain confidence to facilitate and support others in application of these approaches. Development of domain-specific and use-case FIPs, in collaboration with GO FAIR Foundation, WP7 colleagues and domain experts. Development of domain-specific and use-case metadata, in collaboration with GO FAIR Foundation, WP7 colleagues and domain experts. Provision of expertise to support PARC-projects in the delivery and implementation of their DMP, including where needed, recommendations for new Use cases to facilitate development of bespoke solutions to address needs / gaps. Lead the FIP development and implementation for new Use-cases identified from projects (WP7 Role description).
GOLIATH	Beating Goliath: Generation Of Novel, Integrated and Internationally Harmonised Approaches for Testing Metabolism Disrupting Compounds (GOLIATH).
HBM4EU	Human Biomonitoring Initiative in Europe is coordinating and advancing human biomonitoring in Europe and so provide better evidence of the actual exposure of citizens to chemicals (HBM4EU).
NanoCommons	NanoCommons will create a community framework and infrastructure for reproducible science, and in particular for in silico workflows for nanomaterials safety assessment and beyond (NanoCommons).
PARC	Partnership For The Assessment Of Risks From Chemicals (PARC).

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 11

RISK-HUNT3R	“ <u>RISK</u> assessment of chemicals integrating <u>HU</u> man centric <u>Ne</u> xt generation <u>Te</u> sting strategies promoting the <u>3Rs</u> ”. RISK-HUNT3R will provide a full sustainable framework for NGRA that is human-relevant, fully based on non-animal approaches, and fit for implementation through engagement with chemical safety regulators (RISK-HUNT3R).
SbD4nano	Safe-by-Design for Nano (SbD4Nano) project is to create a novel e-infrastructure for the definition, performance testing and implementation of Safe-by-Design (SbD) approaches in the nanotechnology supply chains (SbD4nano).
ToxRisk	Integrated European ‘Flagship’ Programme Driving Mechanism-based Toxicity Testing and Risk Assessment for the 21 st century. EU-ToxRisk will drive the required paradigm shift in toxicological testing towards a toxicological assessment based on human cell responses and a comprehensive mechanistic understanding of cause-consequence relationships of chemical adverse effects (ToxRisk).
Use cases	Specific use cases (projects within PARC) will enable the multi-faceted FAIRification process to be developed gradually and pragmatically.

EUROPEAN PARTNERSHIP

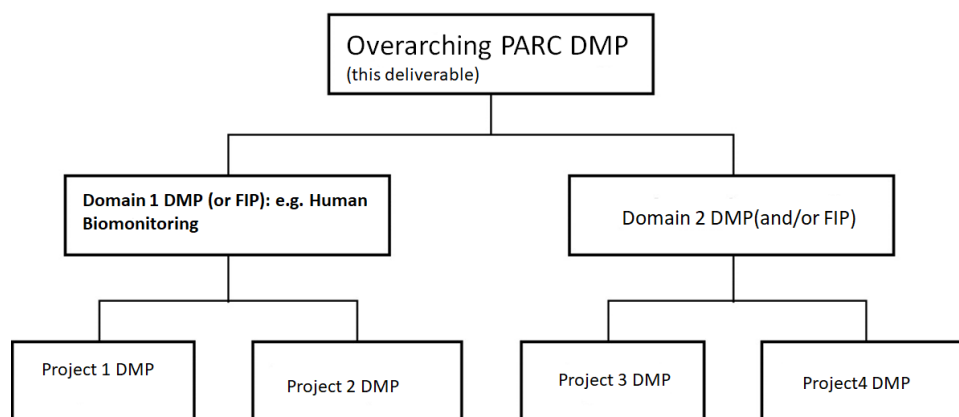


**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 12

Preamble to PARC Initial Data Management Plan

With 200 partners and a budget of 400 Million euro, PARC is operating at a scale where it has huge complexity and where each WP is itself larger than most projects, and runs over a 7 year period based on rolling workplans. As a means to manage this complexity, the work is organised into **projects** of different duration and different configurations of partners, which are clustered by topic or domain of activity, e.g., environmental monitoring versus human biomonitoring, toxicity / ecotoxicity, etc. Each project and domain may have specific needs not common across all domains / projects. We thus see the DMP as set of a nested DMPs of increasing specificity and refinement, starting from this overarching one which lays down the broad principles, but which need to be checked for their implementation in each of the project. The “nested DMPs” approach is shown schematically here, and is envisaged as a “living document” that will be continuously updated over the course of the project, including the two formal updates at mid-terms and at the end of the PARC funded period.



Schematic representation of the nested DMPs concept that will be followed in PARC.

As we progress from this **initial PARC DMP**, which mostly identifies the key issues and lays out the **processes by which solutions to these issues will be developed and shared**, towards the mid-term PARC DMP, the level of domain-specific specification of the exact processes will be included in the nested DMP which as the Mid-terms DMP will include the project-level DMPs of all projects to date.

The number of research domains that PARC’s research spans is also wide, from analytical sciences to human exposure monitoring, to ecotoxicity and omics analysis, each of which has its own customs, norms and best practice, and each of which are in different stages of their journey towards becoming FAIR. Thus, a key piece of work identified already at this point is that PARC partners will need to agree what is our level of FAIR ambition for PARC data – should all data reach the same level of FAIR maturity or can different domains progress based on their current level of FAIR maturity? The initial PARC DMP won’t have the answers to this question, but lays out the issue, and a process for reaching consensus and implementing the targets agreed across all types of PARC data.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 13

Introduction

[Open Science](#) is the standard method of working for EU-funded scientific practice and is a policy priority for the European Commission. To realise this ambition, [FAIR](#) (Findable, Accessible, Interoperable, Reusable) and open data sharing will become the default for EU-funded scientific research. PARC is the European partnership for the development of next generation chemical risk assessment methods. In the context of Open Science, the PARC objective OO10 is to implement FAIR data practices and enhance innovation in complex data analysis for chemicals risk assessment. PARC will develop a FAIR data culture enabling open science and provide the technical methods, tools and infrastructure for effective exchange of data and information. Innovation in risk assessment and risk management which is required for the transition to next generation risk assessment can only be achieved in an open and collaborative way.

This deliverable presents version 1 of the PARC overarching data management plan (DMP). The compulsory implementation of the DMP to foster the sharing and re-use of existing data within and beyond PARC includes support for data stewards in making domain specific choices for their projects for effective exchange of data and information by serving as a reference document. A knowledge sharing infrastructure for this purpose will subsequently be developed.

The DMP will evolve over time when it takes on board inputs from data stewards, data champions and data liaisons after PARC projects (smallest entity where research is actually being implemented and data generated) have started, been executed and closed. During the PARC partnership issues will be solved, while on the other hand new issues will emerge. The PARC DMP is a living document that will be updated over the course of the project with formal deliverables in M6, M42 and M72.

This version 1 is based on the [H2020 DMP template](#) and aims to be overarching regarding the width of the PARC domain, while it also aims to provide domain specific options for PARC researchers to adopt into their detailed task- or project-related DMPs. This deliverable will be publicly available, via the project website and the EC portal, and can be reused as such. The initial PARC DMP will be made available online via DMP Online (<https://dmptool.org/>), the Data Stewardship Wizard (DSW, <https://ds-wizard.org/>) and potentially at the OpenAIRE and European Open Science Cloud (EOSC)-backed Argos tool (<https://argos.openaire.eu/home>) and the FAIR Connect website (<https://www.adoro.net/fair/v01>), in order to both test the flexibility and suitability of each, as well as the partner's (and WP's) familiarity with and utilisation of the different tools.

The DMP contains guidance to support PARC researchers and data stewards (see glossary), working together in collaborative teams. The guidance in the DMP is not intended as a tick box exercise, but rather to trigger thoughtful choices and considerations. Researchers and data stewards are encouraged to seek advice from their institutional information specialists, data protection officers and information security officers. In addition, training of PARC FAIR data stewards and data champions (train-the trainer) is foreseen, which will ensure compliance with the aims of the PARC DMP. The guidance includes generic text and other resources, as well as domain (and project) specific options, which PARC researchers can choose to adopt the DMP for their specific project. The reason to provide domain specific options is to, wherever possible, drive convergence on the

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 14

stewardship and implementation choices being made towards common agreements within the chemical risk assessment stakeholder community that will support efficient data sharing and re-use. The domain specific guidance in version 1 is based on the analysis of DMPs from the EU projects EPHOR, HBM4EU, NanoCommons, ToxRisk, RISK-HUNT3ER, COMPSAFENANO, Sbd4nano, GOLIATH. These projects cover some of the main (research) data domains PARC is concerned with: i.e., exposome research, biomonitoring, innovations in hazard assessment, e.g., via *in vitro* new approach methodologies, nanomaterial safety and metabolism disrupting chemicals. Additional domain specific guidance from other initiatives will be included in further updated PARC DMPs.

The PARC DMP provides guidance for the researcher in multiple ways:

- By providing information on the key points to be addressed.
- By providing generic texts which can be adapted or elaborated to meet project-specific purposes. Keep in mind that the generic texts are generally short, but that the amount of information provided by the user should be in appropriate balance with the scale of the project deliverables and budget, such as providing details per work package.
- By converging choices made in projects towards preferred options regarding both domain specific or generic implementation of, for example, metadata schemas, standard controlled vocabularies, data standards, data repositories to enable other users to build on the knowledge gained during the project. It is the intention to develop comprehensive lists of commonly agreed and preferred standards in PARC, considering data type or subdiscipline. A unified system of FAIR standards will make PARC projects more reproducible and interoperable. If researchers want to deviate, they should therefore justify why this is really necessary, and describe how the deviations will provision equivalent levels of FAIRness.

The DMP is organised along the FAIR principles ([Go-FAIR](#)). The aim is that all data are made as FAIR as possible, under the existing requirements and constraints. In this respect, the PARC FAIR Data Policy, will aim at a realistic level of ambition towards FAIRification of all PARC-related data. The FAIRification will start with the use cases in WP7.2.

FAIR and open data sharing according to the ‘as open as possible, as closed as necessary’ principle will be the default for PARC and will be applied to all research outputs (including reports, data, software, guidelines, method descriptions, formats, templates, semantic artefacts (ontologies, vocabularies etc.). Also note that FAIR is not the same as Open Data ([link to detailed explanation on Go-FAIR](#)). A second note here to make is that FAIR principles apply both to open and restricted data. Open Data is openly accessible, exploitable, and editable, and can be shared or visited by anyone for any purpose, including commercial use. However, there may be legitimate reasons to restrict access to data under well-defined conditions (licenses, restrictions, costs), because of regulatory and jurisdictional reasons, such as consent, privacy, confidentiality or Intellectual Property Rights (IPR). When there are such restrictions on the reuse of individual datasets (e.g., regarding openness), those restrictions and protocols for access should be made explicit to both humans and machines. This will include implementations around FAIR Principles A1.2 and R1.1.

IP rights on data and sharing of data among the consortium and the projects defined within PARC are covered via the PARC Grant Agreement and PARC Consortium Agreement. PARC data can be shared during the course of PARC with external parties as well.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 15

All partners agreed to contribute to, and implement the data management workflows laid out in, the Data Management Plan, as agreed in the PARC Consortium Agreement Article 8.5 ‘Open science research and data management’. All partners will take the following actions (citation from Consortium Agreement):

- “Contribute to the Data Management Plan (and regularly update it): The Data Management Plan will consist of generic parts that are applicable to all Data, and specific parts that are applicable to specific subsets or types of Data. Each Participant is responsible that management of the Data it generates and/or uses, is properly and timely addressed in the Data Management Plan, and will provide the necessary inputs to that extent, making use of the tools provided for that purpose by the Partnership. Each Participant has the responsibility to timely bring to the attention any issue that may impede adherence to it. Each participant will update parts of the Data Management Plan that specifically pertain to the Data it generates and/or uses, without delay, whenever significant changes arise. This includes, but is not limited to: the generation of new data, changes in data access provisions or curation policies, attainment of tasks (e.g. datasets deposited in a repository, etc.), changes in relevant practices (e.g. new innovation potential, decision to file for a patent).”

The project-level DMPs must be developed in alignment with this overarching PARC DMP (this Deliverable and the subsequent iterations) and will become appendices to the PARC intermediate and final DMPs. Thus, the project level DMPs should follow the same template and ensure alignment with, and updating of the PARC DMP as needed. The Data Liaison and Data Champions will have overall responsibility for implementing this alignment process. However, **each partner** in the specific project is responsible for **implementing the project DMP** steps and for ensuring that their data is managed in accordance with the project-level DMP.

Future versions of the PARC DMP will thus include inputs from specific PARC projects executed within the WPs. The planned convergence, coordinated by WP7, of the FAIR implementation choices made by the various PARC communities will provide additional guidance to researchers on implementation of the FAIR principles. Future versions of the DMP will also consider the ethics framework, which will be developed within Task 1.4, and the principles of IP rights management as planned to be developed in the first year Annual Work Program (WP1).

Future versions of the DMP will also accommodate the numerous and ongoing developments in the area of Open Science and FAIR data, such as the European Open Science Cloud (EOSC) and in the innovation of risk assessment, such as the DG ENV’s Common Open Platform on Chemical Safety Data (COPCSD).

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 16

1. Data Summary

Points to be addressed

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify the existing data that is being re-used
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful
- Metadata preservation policy as requested by Principle A2 (not mentioned below)?!

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

PARC's general objective is to consolidate and strengthen the EU's R&I capacity for chemical RA to protect from impacts on human health, biodiversity and human and natural environments at large. Linked to this general are three specific objectives (SO), around which the 9 Work Packages (WP) of PARC are structured and for which 13 realistic, measurable, achievable and verifiable operational objectives (OO) are developed. The WPs and OOs are further described in Part B of the Project Proposal.

PARC's first specific objective (SO1) is that EU and national risk assessors and regulatory entities come together with the scientific community in a cross-disciplinary network to set priorities for R&I in chemical RA. The implementation of FAIR data and services as directed under this DMP can foster these discussions.

Specific objective 2 (SO2) is that European and national RA entities and their scientific networks carry out a joint R&I programme to respond to the agreed priorities in chemicals RA. In this programme, various types of data are being generated, from existing and novel technologies (e.g., on hazard assessment/toxicology, exposure assessment, risk assessment, see in detail 1.2). The purpose of this data collection is to help the development of innovative methods for risk assessment that can be ultimately applied for regulatory purposes. This will allow to speed up the risk assessment process, and to potentially use less animal data. Also, some datatypes are generated to fill in data gaps, e.g., conventional *in vivo* toxicology data.

Specific objective 3 (SO3) is that European risk assessors, their scientific network and the wider stakeholder community have access to the R&I capacities required to implement innovative chemical RA.

The implementation of collected data compliant with the PARC DMP will result in an increase in the FAIRness of chemical risk assessment related data, thus the potential access to and interoperability among the data generated within PARC.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 17

1.1.1 The overarching PARC DMP and the relationship with project-level DMPs

Given the size and scale of PARC, much of the research will be performed with smaller units of fixed duration, referred to as projects. These are developed within (or across) WPs and reviewed for the fit to the overall PARC regulatory remit by WP2 and then supported in the development and implementation of their project-level DMPs. These DMPs are specific to the type and scale of data being generated in the project, and aligned to the domain-specific norms and standards (where these exist). A key aspect of the initial PARC DMP (this deliverable) is to lay out the process for integration and alignment of Data Management into the projects. Here we lay out the details of this process (see Figure 1 and Figure 2), the roles and responsibilities within WP7 and with the PARC projects, and the initial process agreed for development, review, updating and alignment of project-level DMPs and the overarching PARC DMP.

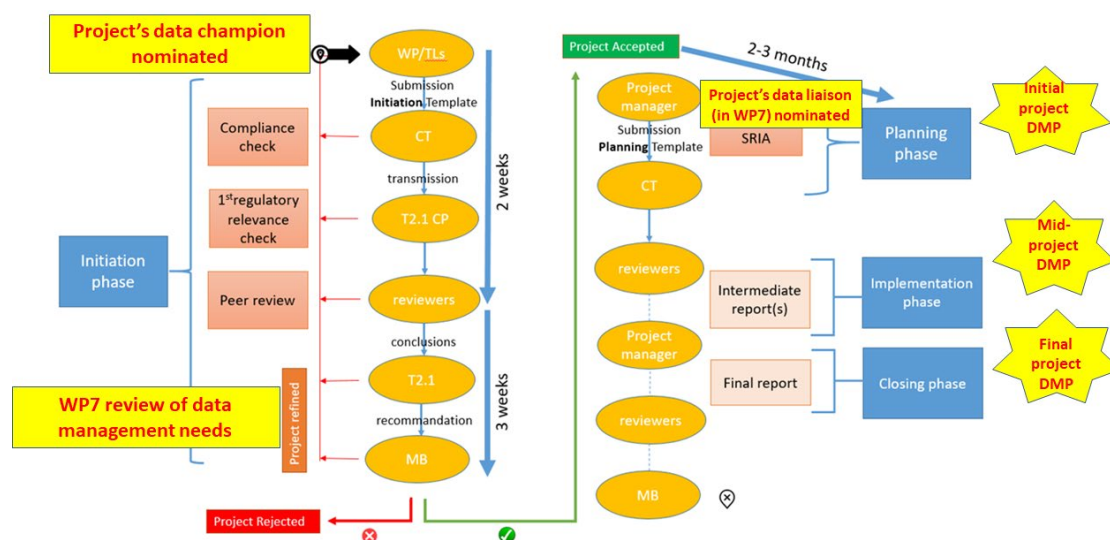


Figure 1: Summary of the process for development, review, approval and implantation of PARC projects, and the involvement of WP7 in the development, review and updating of the project-level DMP. The roles of Data Champion and Data Liaison are described briefly in the glossary, and a fuller description is presented below.

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 18

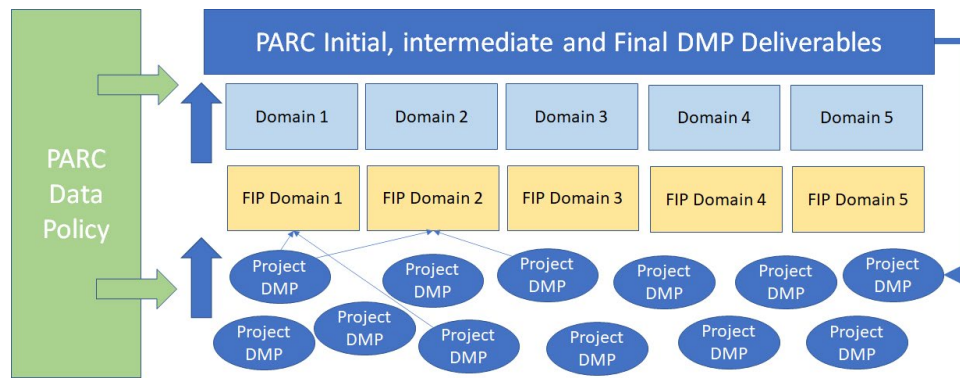


Figure 2: Schematic illustration of the relationship between the PARC Data Policy, the overarching PARC DMP and the project-level DMPs. We define also a set of Domains of relevance to PARC, e.g., environmental and human exposure, environmental and human hazard, chemical properties, omics data, computational data etc., and the clustering of projects within and across these domains to help ensure harmonisation of the project-level and PARC DMPs. Additionally, we foresee the development of Fair Implementation Profiles (FIPs) initially at the domain level, and for the Use-cases that will be defined within WP7.

Specifically, in an interactive process (Figure 3), PARC project leaders are required to describe the data management section of the project initiation template, and to communicate in an early stage to WP7 for all projects for which data will be generated or used, if reuse is foreseen (for input into 1.3), what the expected data size is (for input into 1.4) and to whom the data may be useful (for input into 1.5). WP7 assigns a Data Liaison to the project, and PARC project leaders (preliminarily) assign a Project Data Responsible (“Data champion”). This person organizes an exploratory meeting with the Data Liaison (for which we can use (parts of) the actual data management template as guide for the discussion) and includes conclusions of this exploratory meeting in the Project Initiation Template. These will serve the input towards this DMP.

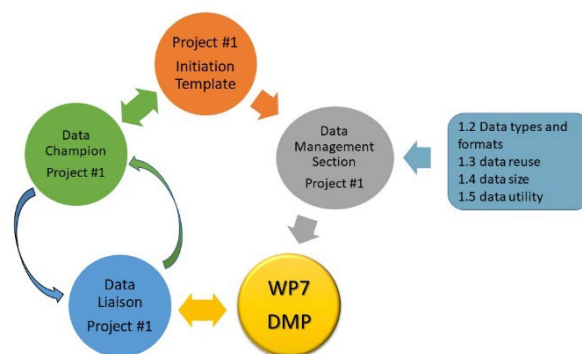


Figure 3: The iterative process by which PARC projects leads communicate on data types, data formats, data reuse, data size and data utility within their project / task /WP and with WP7.

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 19

The [roles and responsibilities](#) of the Data Champion, Data Liaison and Fair Implementation Task group are defined, and the processes for how they are working is being defined. (Please see the roles and responsibilities via the link for now, as WP7 partners are still refining this – the polished version will be dropped in here when ready). Figure 4 describes the tasks, as currently defined, for the Data Champion, Data Liaison and FAIR Implementation Taskgroup (FIT). A key aspect of each of the 3 data support roles in PARC is that they will receive FAIR awareness training and be upskilled in data management. WP9, task 9.4, will support the organisation of training regarding data (re)use (WP7).

Data Champion	Data Liaison	FAIR Implementation Taskgroup (FIT)
<ol style="list-style-type: none"> 1. Identified by Tasks / Projects from within their membership 2. Work closely with WP7 to define existing data management approaches & specific needs of the project 3. Ensure data management always high on agenda 	<ol style="list-style-type: none"> 1. Experts from within WP7 covering different domains as contact points for data champions 2. Domains include tox / ecotox / HBM / informatics / ontologies / IT aspects 3. Groups of projects to be assigned a liaison 	<ol style="list-style-type: none"> 1. A sub-set of the Data Liaisons (& Champions) who will receive further training from GOFAIR 2. Will receive certification as FAIR data trainers 3. Support implementation of project-specific DMPs through provision of ongoing training & support

Figure 4: This figure describes the tasks, as currently defined, for the Data Champion, Data Liaison and FAIR Implementation Taskgroup (FIT).

The [GO FAIR Foundation](#) (GFF) has been subcontracted by PARC to provide training for the Data Champions, Data Liaisons and to establish the FAIR Implementation Task group. GFF have developed a [Three-point FAIRification Framework](#) that begins with a local data producer (e.g., university, hospital) deciding on a range of data policy issues and metadata descriptions needed to ensure FAIRness. These metadata are then rendered machine-actionable in (M4M) workshops. The reusable metadata schemata produced in the M4M compose part of the larger FAIR Implementation Profile (FIP), which in turn guides the configuration of the FAIR Data Point (see Figure 5).

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 20

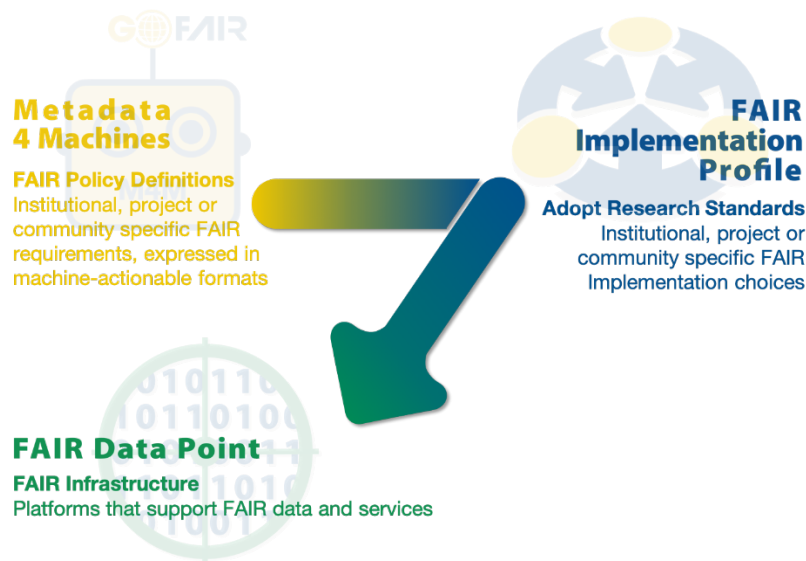


Figure 5: Schematic illustration of the Go FAIR Foundation 3-point implementation that will be applied in PARC through training of the FAIR Implementation Task group members (who can also be Data Liaisons or Data Champions) via the M4M workshops and through the development of FAIR Implementation Plans (FIPs) for each domain involved in PARC and for each of the WP7 use cases.

Data liaisons and data champions are key in making data FAIR. Data will be made FAIR at the source whenever possible, described by FAIR metadata, and associated with a persistent identifier. Specific use cases will enable this FAIRification process to be developed gradually and pragmatically. Not all use cases have been identified yet. The process of use case identification (Tasks 7.2.3 & 7.1.3) starts from the project planning phase by analysing and clustering the project initiation plans to the description of use cases. The data management sections in the project plans will point the attention to different data management aspects related to data generated in PARC as well as to use of data from external sources. The use cases will be formalised as PARC projects.

1.2 What types and formats of data will the project use?

The FAIR Principles require to provide both the data and its metadata. Ideally both should be rendered machine-actionable, but metadata should describe the format of the datasets in whatever form they take, whether they are machine-ready or not. Examples of most commonly used data/file formats can be found [here](https://en.wikipedia.org/wiki/List_of_open_file_formats) (https://en.wikipedia.org/wiki/List_of_open_file_formats) and here: [data formats in bioinformatics \(Lapatas et al, 2015\)](#). During the PARC project, data champions, data liaisons and data stewards should try to identify/determine in which specific domains standards are lacking and where there is a need for convergence to standards. The various forms or categories of data to be generated and collected include: raw or experimental data, derived (processed, computed, or computational) data and data associated with formal publications. Data types can be: cohort data, molecular data, occurrence data (including suspect and non-targeted screening data), occupational exposure data, hazard effect data ((eco-)toxicological endpoints for risk assessment),

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 21

modelling data (data to develop and validate in silico models, model output data), biological measurement data (including omics data), clinical data, health outcome data, data from questionnaires, pathway data, additional data (e.g., from grey literature, research paper).

In terms of size and heterogeneity of the data generated, PARC is extensive. It is expected that various biomonitoring and environmental and occupational exposure data (WP4), hazard data (WP5) (*in vitro*, *in vivo*, *in silico*) and data emerging from novel risk assessment models (WP6) will be produced. Right now, it is not possible to fully describe all data in full detail, as PARC projects, in which these data will be generated, are just being conceived, but we are implementing a mechanism in WP7 to achieve the required level of detail on the types of data being used. For illustration, exemplary data types and potential details, are listed in Table 1, including the reuse of information from existing DMPs.

Type of data	Give details on category of data (e.g., database, excel file, etc.), data format (e.g., csv, xls, etc.), and how the data will be generated/collected.
<i>Data of measurements in biological matrices</i>	<i>Raw data (measurement data) output files from instruments will be MS Excel compatible file formats (CSV, txt and proprietary software files (for suspect screening data)) (HBM4EU DMP)</i>
<i>Sensor based data for chemical occurrence</i>	<i>Expected to be mainly CSV compatible. (EPHOR DMP, NORMAN DCT)</i>
<i>Data from in vitro assay experiments where a biological sample (e.g., cell line) is exposed to a chemical</i>	<i>Omics data in proprietary format as well as CSV compatible (RiskHunt3r DMP).</i>

Table 1: Examples of data types and data formats, the method by which data are generated. ^[OBJ] More details will follow from the projects DMPs.

1.2 Which data(sets) will the project generate or collect?

Please specify here which datasets you are going to use, either based on existing data or those that are being created in PARC.

Type of data	Specific datasets	Description of data
<i>Data of measurements in biological matrices</i>	<i><not known yet></i>	<i>Measurements in urine, sputum, blood, hair from human subjects</i>

EUROPEAN PARTNERSHIP



Co-funded by
the European Union

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 22

<i>Sensor based data for chemical occurrence</i>	<i>WFD surface water Monitoring data - 2017 (France)</i>	<i>Surface water measurements</i>
<i>Data from in vitro assay experiments where a biological sample (e.g., cell line) is exposed to a chemical</i>	<i>List of endpoints for PPP</i>	<i>Hazard assessment to fill in gaps</i>

Table 2: Example of datasets that the project will generate or collect.

1.3 Will you re-use any existing data and how? What is the origin of the data?

If applicable, state any constraint with reason on the re-use of existing data. During the PARC projects these details will emerge and will be used to populate the table.

Explain why re-use of existing data is not feasible and data needs to be generated in this project.

Type of data	Dataset (Origin of the data)	Give details on any constraints. If applicable, explain why re-use of data is not possible and the project will generate the data.
<i>Personal exposure data</i>		<i>GDPR</i>
<i>Data at ECHA uploaded by companies</i>		<i>Confidential data</i>
<i>Data from unfinished research</i>		<i>Data related to concept documents which are not yet published</i>
<i>Chemical occurrence data</i>	<i>WFD surface water Monitoring data - 2017 (France)</i>	<i>Publicly available</i>

Table 3: Examples of data types, their origin and description of the constraints associated with the data sets.

1.4 What is the expected size of the data?

Type of Data	Size estimation (if known)

EUROPEAN PARTNERSHIP



Co-funded by
the European Union

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 23

<i>Data of measurements in biological matrices</i>	<i>The expected size the crude data size will be about 2 GB per study</i>
<i>Sensor based data for chemical occurrence</i>	<i>An indication: for chemical concentrations from passive sampling wrist band: 100-500 GB</i>
<i>Data from in vitro assay experiments where a biological sample (e.g., cell line) is exposed to a chemical</i>	<i>About 1-1.5 GB</i>

Table 4: Examples of data types and the estimation of the size if this can be known in advance.

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

Generic: Scientists will be enabled to use and enhance the data, methods and models in innovative RA research. Regulators (e.g. ECHA, EMA, EEA, EFSA), policy makers and risk assessors within the EU will use the data for the development of evidence-based policies and risk advice.

Given PARC's remit to support regulatory risk assessment of chemicals, key stakeholders or end-users for the PARC data and models are the EU regulatory agencies (ECHA, EFSA etc.) and the member state regulatory organisation, including those involved in the PARC project.

Scientists (toxicologists, ecotoxicologists, modellers, risk assessors etc.) within and beyond PARC will also be major users of the datasets generated in, and harmonised by PARC.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 24

2. FAIR data

Points to be addressed

In general terms, your research data should be 'FAIR', i.e. that it is findable, accessible, interoperable and re-usable by humans and also by machines. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution. Note that, in general, the FAIR elements in "F", "A", "R" provide benefits beyond the project scope, whereas the elements in "I" can also directly enhance data integration benefits across PARC projects.

2. 1. Making data findable, including provisions for metadata

Metadata are values or texts that provide a description of data of interest. As one use, metadata can describe a complete dataset (e.g. datatype, author, period, subject). As another use, metadata can describe individual data points or samples for data selection and interpretation.

Points to be addressed.

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

2.1.1 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)?

Generic: Data sets are stored with metadata and DOIs (via available PID services, e.g. [ePIC](#), or via trusted data repositories, publishers), which will be available after publication of the paper. If possible, metadata are published preceding the publication of the actual data. The data related to the publications are findable by users via the repository website via a metadata search and/or as outlined in a Data Availability Statement. Entities, such as chemical names and gene names, will be used with their standard global identifier. In case chemicals or metabolites are used which do not have a global identifier, the [European Registry of Materials](#), which was introduced by the NanoCommons project, may be a solution to assign a unique registry number.

2.1.2 What naming conventions for files do you follow?

Generic: The general advise is: files should be named consistently per project. File names should be short but descriptive (<25 characters). Use capitals and underscores instead of periods or spaces or slashes, no special characters or spaces. Use date format: YYYYMMDD as suffix. A version number or the term "FINAL" in the suffix will be included. Build filenames from general elements to specific.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 25

A possible format can be: *PARC_WPn_Short_name_Version_YYYYMMDD*. The file names containing the key towards pseudonymisation should not contain personal information.

Concerning the hierarchical file folder structure (as in Windows), it is advised not to extend these towards too deeply nested structures as this may impair the retrieval of files with too long path names. If possible, semantic file systems are used for information persistence.

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

Generic: Keywords will be used as a standard. Aside free text, researchers will be encouraged to use as much as possible keyword terms listed within controlled vocabularies and metadata. Researchers should also know that increasingly research outputs in repositories are indexed automatically by assigning keywords.

2.1.4 Do you provide explicit machine-readable version numbers?

Explain how the versioning is done, depending on the research environment (e.g. SharePoint, MySQL, iRODS\YODA, other such as GitHub).

Generic: The following versions of the data will be recoded as a new version:

1. The raw data as collected
2. The pre-processed data as the basis for analyses (clean data). There can be more versions here.
3. The processed data used for a publication (for example peer-reviewed or report)

2.1.5 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.

Per PARC domain specific choices will be made. Projects are suggested the following. The [Digital Curation Centre](#) and [RDA](#) maintain an overview of metadata standards covering a wide range of research domains and general purposes. Here below a list with generic and PARC specific examples which will be updated during the PARC project. The inventory of relevant metadata schema's will also evolve from the interaction with the PARC projects. In addition to currently established metadata standards, there are initiatives towards developing standards for specific purposes such as [nanomaterials](#).

Metadata schema	Purpose	Examples of use
DataCite Metadata	Generic metadata schema. A set of mandatory metadata that must be registered with the DataCite Metadata Store when minting a DOI persistent identifier for a dataset. For citation and retrieval purposes.	OpenAire , Zenodo , iRODS

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 26

Metadata schema	Purpose	Examples of use
DCAT	Generic metadata schema. To describe datasets and to facilitate interoperability between data catalogs.	FAIR Data Points architecture.
DCAT-AP	Generic metadata schema. To describe datasets and to facilitate interoperability between data catalogs	European Data Portal
EEA's Meta-data reported by countries	Metadata schema used for measurements or modelling of air pollutant concentrations.	EEA
EMA's List of metadata for Real World Data catalogues	Metadata schema. The EMA produced a list of metadata for describing real world data (RWD) sources and studies. The metadata will be included in a catalogue of data sources.	EMA
GEO DataSet descriptors	Metadata schema. GEO is a public repository that archives genomics data submitted by the research community.	GEO (NCBI)
INSPIRE Directive metadata	Metadata schema for spatial data sets. Member States must ensure that metadata correspond to the themes listed.	EU
IpCheM metadata	Defined by the HBM4EU project to describe human biomonitoring metadata. The template is an extension of the IPCheM metadata schema which also covers environmental data.	IPCHEM
ISA-TAB	ISA is a metadata framework, containing the ISA-Tab file format, software, curation, for 'omics-based' experiments. It has extensions for nano, toxicogenomics.	ISA-TAB use cases. The ISA-TAB software facilitates metadata submission to data repositories (e.g. EMBL-EBI ArrayExpress, European Nucleotide Archive, MetaboLights).
Nanosafety	Metadata schema for nanosafety	In development stage, in research

Table 5: Examples of metadata schemas intended to provide preferred options for project data management plans. The table contains name of the schema, description and example of use.

EUROPEAN PARTNERSHIP



Co-funded by
the European Union

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 27

2.2. Making data openly accessible

Points to be addressed

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access or visit the data? 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)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

2.2.1 Which data produced and/or used in the project will be made openly available as the default?

Projects are suggested the following. Please note the principle “as open as possible, as closed as necessary”. For human data follow the GDPR. If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

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.

Data	Comment on whether or not be made openly available
<i>Data of measurements in biological matrices</i>	<i>GDPR and consent of data subjects apply to individual human biomonitoring data; storing and access by other parties than those that collected the data, are only possible following signing of data controller and/or processor agreements</i>
<i>Sensor based data for chemical occurrence</i>	<i>Will become open available in repositories</i>
<i>Data from in vitro assay experiments where a biological sample (e.g., cell line) is exposed to a chemical</i>	<i>Will become open available in repositories</i>

Table 6: Examples of data types and comments whether these data types will be made open or not.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 28

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

Per PARC domain specific choices will be made. Projects are suggested the following. Researchers should first determine if a domain repository exists for their research data and if that repository is FAIR. On [FAIRsharing.org](https://www.fairsharing.org) and re3data.org one can find an overview of domain specific repositories. If a domain specific repository is not possible or is not FAIR, then a general or the institutional repository must be used. The institutional information specialist can help to assess the FAIRness of the repository by checking the [FAIRsFAIR requirements](#) (page 11-12). Science Europe has published [criteria for the selection of trustworthy repositories](#); trustworthy does not necessarily imply FAIR. Regarding GDPR, when it concerns aggregated human biomonitoring data, storing and access may be possible, as example, via IPCHEM. The [BioStudies](#) (EMBL-EBI) database holds descriptions of biological studies (e.g., multi-omics), long-term storage data repository and simple data access.

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

Generic: Access to code (developed within the project), related documentation and related test data can be achieved after the project by, for example: GitHub, Docker, GitLab. Details will be provided in the next updates of the DMP.

FAIR data points (FDPs) and application Programming Interfaces (APIs) are two approaches to access data. APIs are needed for each data warehouse to allow access to the data – REST or SOAP are the preferred ones to support interoperability. It is unlikely to be possible to ensure that all PARC-endorsed databases / data warehouses utilise a common API.

[adapted from [FAIR Data Point](#)] FAIR Data Points can be used to describe data sets in a FAIR way, using standard metadata and make them available through simple internet protocols. A FAIR data point takes care of a lot of the issues that need to be taken care of to make data FAIR; especially with the metadata needed for Findability and Reusability, and a uniform open way of Accessing the data. The FAIR data point addresses the Interoperability of the metadata it stores, but it leaves the Interoperability aspects for the data itself to the data provider. Yet, FDP's are not foreseen to be set up in PARC.

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

Generic: Access to documentation of the code (developed within the project) can be achieved after the project by, for example: GitHub, Docker, GitLab.

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

Generic: Access to code, if developed within the project, can be achieved after the project by, for example: GitHub, Docker, GitLab.

2.2.6 Where will the data and associated metadata, documentation and code be deposited?

Per PARC domain specific choices will be made. Projects are suggested the following examples.

Generic: Access to code (developed within the project), related documentation and related test data can be achieved after the project by, for example: GitHub, Docker, GitLab. Publications, related data and metadata will be deposited in a public trustworthy repository (e.g. Zenodo). The institutional

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 29

information specialist can help to assess the FAIRness of the repository by checking the [FAIRsFAIR requirements](#) (page 11-12).

2.2.6 Have you explored appropriate arrangements with the identified repository?

Generic: No such arrangements are yet identified.

For projects the following example can be informative. Regarding human biomonitoring data, arrangements can be made with IPCHEM (European Commission's Information Platform for Chemical Monitoring). This has been done in HBM4EU. One of the objectives of HBM4EU is integration of HBM data into IPCHEM, which is managed by JRC. In this context, a collaboration between HBM4EU and JRC has been established. The HBM4EU repository is hosted at JRC as one of the components of the IPCHEM architecture.

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

Generic: Restrictions on use will apply to human data and are subjected to GDPR. Data controllers/data owners shall implement appropriate technical and organisational measures and can authorise or deny access to data. The data provider grants relevant authorisation on a case-by-case basis regarding data sets and data users following confidentiality, authentication and integrity.

2.2.8 Is there a need for a data access committee?

Generic: A data access committee will be formed. The terms of reference of this data access committee will be decided upon during the running of the project and will be discussed and agreed upon in the general assembly.

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

Generic: The license URL is part of the metadata, and as such machine readable. The corresponding license text will be part of the data package. Access to restricted data is only possible for authenticated users and explicit (and perhaps PARC standardized) access protocols will be created.

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

Generic: The data will be accessible upon open access principles. The access to the data will be without any login or user account, unless such information will be useful to measure use and reuse of the data.



D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 30

2.3. Making data interoperable

Points to be addressed

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability?

2.3.1 Are the data produced in the project interoperable?

Specifically, allowing data exchange in an interoperable way 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)? Examples of most common data/file formats can be found [here](#) and here: [data formats in bioinformatics](#).

Generic: Data formats will be provided in open formats such as csv. Metadata descriptors and keywords grounded to controlled vocabularies and ontologies will be provided as much as possible. Specific choices made will be included in the update of the DMP.

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

Per PARC domain specific choices will be made. Projects are suggested the following. Below a list of preferred data standards is presented. Deviation is possible, but it then needs to be justified. These standards can be used for reporting data. If standards are missing, they can be added or if standards are lacking in a PARC domain, such gaps are identified by, for example, the data liaison.

Data standard	Domain	Examples of use
EFSA's Data Collection Properties	Data standard for reporting data from monitoring programmes or analytical results. EFSA needs the results of these activities in its databases.	EFSA's data collections
OECD Harmonised Templates (OHTs)	Data standard for reporting data used for the risk assessment of chemicals.	IUCLID 6 (ECHA)
Approved Data Standards at the EPA	Data standards that define EPA's data collection and exchange.	EPA

Table 7: Examples of standards for reporting data. The table contains the name of the standard, description and use.

EUROPEAN PARTNERSHIP



Co-funded by
the European Union

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 31

Below a short list of well-known standard controlled vocabularies is given for the PARC knowledge domain. In case for a specific knowledge domain for which no standard controlled vocabulary exists, one can stand on the shoulders of others who uploaded their vocabulary in e.g. the [NCBO BioPortal](#) (repository for bio-ontologies), [OBO Foundry](#) (biological and biomedical ontologies), [FairSharing](#) (includes a registry of semantic artefacts). In addition to established semantic artifacts, vocabularies are being developed and published in science. These ones are also included in the table below.

Controlled vocabulary	Domain	Examples of use
ChEBI	Molecular entities focused on 'small' chemical compounds	Endorsed by IUPAC and NC-IUBMB
eNanoMapper	Nanomaterials and nanosafety	ENanoMapper ontology
EPA-AOP-DB	Adverse Outcome Pathway	EPA AOP-DB , AOP Wiki
Gene Ontology	Bioinformatics	Gene Ontology Resource
IUPAC Gold Book	Chemical terminology	Chemists use this book
MeSH	Biomedical literature	PubMed (nih.gov)
No specific vocabularies exist	Human biomonitoring	
PubChem	Chemistry	About PubChem (nih.gov)
UniProt Controlled Vocabulary	Keywords to retrieve subsets of protein entries	UniProt

Table 8: Examples of controlled vocabularies, their domain of use and example of implementation.

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 32

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

Generic: The project will use standard controlled vocabularies when available (e.g., Gene ontology, MESH, International Classification of Disease, Exposure Science Ontology ExO). At present, no specific data and metadata vocabularies are available for the field of Human Biomonitoring.

2.3.5 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?

Generic: As for now, the project will not develop own ontologies or vocabularies. In case it is unavoidable to use uncommon or generate specific ontologies or vocabularies, mappings to more commonly used ontologies will be explored for feasibility.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 33

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

Points to be addressed

- Specify how the data will be licensed to permit the widest re-use possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- Describe data quality assurance processes
- Describe the data provenance or data lineage in the metadata or publication/documentation
- Specify the length of time for which the data will remain re-usable

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

The [EUDAT B2SHARE, Choose a license](#) and [Creative Commons License Chooser](#) tools facilitate the selection of an adequate license for research data. [Openaire](#) and [Elixir](#) guide researchers on how to license research data.

Generic: The Creative Commons license will be applied to all data. Ideally the CC-BY0 or CC-BY-SA - so that it can be re-used with the minimum limitations yet ensuring citation. All data will be made Open Access at the point of publication of the manuscripts or earlier where possible. Any open data and open-source tools and models will be complemented and compatible with the licenses agreed with the respective owners, and be indexed in OpenAIRE data. Similarly, any third-party data or tools will be reused with complete respect to their licensing requirements, the conditions specified by the respective owners and with appropriate attribution.

2.4.2 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.

Generic: Data will be made full available with a clear re-use license as soon as possible after generation. Alternatively, data will be made full available at the latest at the time of publication of the research. Data will be made available when no human identifying characteristics are ensured to be present in the disclosed data.

2.4.3 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.

Generic: The project data, methods and models will be made available to be used by others. In case of human data, restrictions for data use will be put in place.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 34

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

Generic: The data is intended to be reusable for at least 10 years upon upload to the repository.

2.4.5 Are data quality assurance processes described?

Per PARC domain specific choices will be made. Projects are informed on the examples of standards and recommendations to ensure data quality. If standards are lacking data stewards and data liaisons will provide input for the generic DMP.

Standard or recommendation for data quality	Domain	Examples of use
Good Laboratory Practice (GLP)	General	Testing of chemicals (OECD)
Good Cell Culture Practice (GCCP)	Starting of cell and tissue cultures	Testing of chemicals (OECD)
Good In Vitro Method Practices (GIVIMP)	Developing and implementing cell- and tissue-based methods	Evaluation of chemical safety (OECD)

Table 9: Examples of standards and recommendations that ensure data quality, their domains of application and examples of their use.

Generic: Data quality assurance processes and standards will be described in the protocols of the studies. The partners involved in this will follow the procedures and will also follow partner-specific guidance and procedures. Reproducibility of the data quality will be maintained by registering both the raw data and the curated data. The scripts used and other data handling steps will be described and documented comprehensively.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 35

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long-term preservation

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

Generic: The costs for making the data FAIR are covered within the PARC project budget as allocated in the grant agreement. See guidance on cost estimation, for projects within PARC, of these Universities: [TUDelft](#), [Utrecht University](#).

3.2 How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Generic: Per project, the project leaders and principal scientists are responsible for collection, managing, analysing the various data types. Each partner institute is responsible for storage, back-up, data archiving and sharing.

3.3 Who will be responsible for data management in your project?

Generic: Each of partner institutes for each study is responsible for each of the following: data collection, data pre-processing, data publication, data analysis, data sharing, data interoperability.

3.4 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)?

Generic: Agreement amongst the partners is needed on long term preservation and the costs associated with this. This will be discussed during the course of the project.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 36

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

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

Research performing organisations will have in place data secure research environments. Projects are advised to seek confirmation and describe the institutional data security provisions.

When you need technical guidance regarding personal data, here are some links. Statistical packages may have tooling for anonymisation (e.g. aggregation), or consider to use [ARX](#). Guidance to anonymisation and pseudonymisation techniques and best practices can be found at the [EU ENISA](#), [UK data service](#) and [CESSDA ERIC](#). This link, to [EU ENISA](#), gives guidance to privacy-preserving computation, transfer and storage.

You may describe here the digital/virtual research environment that you are using (for example: MS Sharepoint, iRODS). Microsoft Sharepoint works according to GDPR regulations.

Describe where the data, including the key to pseudonymisation, at rest will be stored and backed up and also describe the backup schedule during research activities. It is recommended to store data in least at two separate geographic locations.

Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Explain how the data will be recovered in the event of an incident.

Explain which institutional data protection policies (DPP) are in place. An institutional DPP is mostly an internal document. You may here provide a summary per institution.

Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships.

If your data is sensitive for example containing personal data or otherwise sensitive: describe the main risks and how these will be managed.

Describe how data in transit (data transfer) is secured by current secure file transfer protocols like the [TLS 1.2](#) standard and/or [SFTP](#). Ask your local security officer for advice. Also consider data visiting versus data sharing under FAIR, as ways to data visiting are being developed. As example: [SPHN - Swiss Personalized Health Network \(SPHN\)](#).

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 37

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

Per PARC domain specific choices for repositories will be made. For PARC projects it is advised to check the retention period of items at the repository and document this information in the project DMP. For example, the retention period of items in [Zenodo](#) is for the next 20 years at least.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 38

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former.

A PARC ethics and data protection framework will be developed, within Task 1.4, in parallel and in coordination with the Data Management Plan under WP7 (Task 7.1 Data policy) to ensure that procedures fully respect data management and privacy legislation, and measures are in place to prevent malevolent use of research findings. Future versions of the DMP will refer to the ethics framework.

Concerns human data, in vivo data and biological (cell) samples/models. The storage and transfer of data on human subjects is only considered when:

- informed consents,
- in-country ethical approval of the study and – when applicable –
- approval by local data protection authorities

cover the purpose that the data are envisaged to be used within PARC and thus allow storage and transfer of individual or aggregated data.

All data that are transferred within PARC shall be either pseudonymised or completely anonymised (concerns direct and indirect personal identifiers). Pseudonymous data are considered by the GDPR as personal data while anonymous data are not. The supplying data controller is responsible for the anonymisation or pseudonymisation process and for ensuring that identifiable variables are not transferred to the receiving data controller.

For project management purposes and to fulfil the ethics requirements of the project, ANSES or a dedicated PARC partner (to be confirmed) will keep a registry of the data exchanges and of data use.

5.1 Are there any ethical or legal issues that can have an impact on data sharing or data visiting?

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) or a policy paper which address legal and ethical aspects.

Generic: All ethical issues related to the project are described in the Description of Action.

In summary, the personal data of any donors will be converted into anonymous research data files. It is therefore not possible to correlate any cells or any experimental data, including genetic information, to the original donor.

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

Generic: Informed consents will cover data sharing and long-term preservation.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 39

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**

This partnership has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101057014

D7.1: PARC Data Management Plan V1.0	Dissemination level: PU
WP7: FAIR Data	Version: 1.0
Main Authors: Fred van de Brug (TNO), Rob Stierum (TNO)	Page: 40

6. Other issues

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any).

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

The procedures for data sharing, and requesting access to use the data are described in the PARC FAIR data policy, which will be developed during the project.

EUROPEAN PARTNERSHIP



**Co-funded by
the European Union**