

# A goal-oriented method for FAIRification planning

César H. Bernabé<sup>1</sup>, Tiago Prince Sales<sup>2</sup>, Erik Schultes<sup>3,4</sup>, Niek van Ulzen<sup>5</sup>,  
Annika Jacobsen<sup>1</sup>, Luiz Olavo Bonino da Silva<sup>1,2</sup>, Barend Mons<sup>1,3</sup> and Marco Roos<sup>1</sup>

<sup>1</sup>Leiden University Medical Centre, Leiden, the Netherlands

<sup>2</sup>Semantics, Cybersecurity & Services, University of Twente, Enschede, the Netherlands

<sup>3</sup>GO FAIR Foundation, Leiden, the Netherlands

<sup>4</sup>The Leiden Academic Center for Drug Research, Leiden, The Netherlands

<sup>5</sup>R&D Observations and Data technology, Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands

## Abstract

The FAIR Principles provide guidance on how to improve the Findability, Accessibility, Interoperability, and Reusability of digital resources. Since the publication of the principles in 2016, several workflows have been proposed to support the process of making data FAIR (FAIRification). However, to respect the uniqueness of different communities, both the principles and the available workflows have been deliberately designed to remain agnostic in terms of standards, tools, and related implementation choices. Consequently, FAIRification needs to be properly planned in advance, and implementation details must be discussed with stakeholders and aligned with FAIRification objectives. To support this, this paper describes a method for identifying and refining FAIRification objectives. Leveraging on best practices and techniques from requirements and ontology engineering, the method aims at incrementally elaborating the most obvious aspects of the domain (e.g. the initial set of elements to be collected) into complex and comprehensive objectives. The definition of clear objectives enables stakeholders to communicate effectively and make informed implementation decisions, such as defining achievement criteria for distinct principles and identifying relevant metadata to be collected.

## Keywords

FAIR, FAIRification, FAIRification objectives

## 1. Introduction

The vast amount of data generated every day is only valuable if it can be properly interpreted and reused. However, it is humanly unfeasible to manually merge and make sense of all the information currently available, therefore the support of machines is required. Although machines can automatically analyse and interpret data to efficiently find useful information, they still require time-consuming human support to prepare and merge data [1]. To address this,

---

*ER2023: Companion Proceedings of the 42nd International Conference on Conceptual Modeling: ER Forum, 7th SCME, Project Exhibitions, Posters and Demos, and Doctoral Consortium, November 06-09, 2023, Lisbon, Portugal*

\*Corresponding author.

✉ c.h.bernabe@lumc.nl (C. H. Bernabé); t.princesales@utwente.nl (T. P. Sales); eriks@gofair.foundation (E. Schultes); niek.van.ulzen@knmi.nl (N. v. Ulzen); a.jacobsen@lumc.nl (A. Jacobsen); l.o.boninodasilvasantos@utwente.nl (L. O. B. d. Silva); b.mons@lumc.nl (B. Mons); m.roos@lumc.nl (M. Roos)  
ORCID 0000-0003-1795-5930 (C. H. Bernabé); 0000-0002-5385-5761 (T. P. Sales); 0000-0001-8888-635X (E. Schultes); 0000-0003-0771-3516 (N. v. Ulzen); 0000-0003-4818-2360 (A. Jacobsen); 0000-0002-1164-1351 (L. O. B. d. Silva); 0000-0003-3934-0072 (B. Mons); 0000-0002-8691-772X (M. Roos)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

the FAIR principles have been proposed to guide the transformation and production of resources that are **F**indable, **A**ccessible, **I**nteroperable and **R**eusable by humans and machines [2]. FAIR resources can be easily managed by machines with minimal human intervention, thus reducing human workload.

The four letters of FAIR are further decomposed into 15 principles [2]. Findability is enforced by using globally unique and persistent identifiers to refer to data and metadata (F1), describing data with rich metadata (F2), explicitly associating metadata with data (F3), and indexing metadata in searchable resources (F4). Accessibility is achieved by using standardised, open communication protocols for data exchange (A1, A1.1) that allow access authorisation procedures (A1.2) while ensuring the longevity of metadata (A2). Interoperability is enhanced by publishing metadata and data in broadly applicable knowledge representation languages (I1), reusing vocabularies that also follow the FAIR principles (I2), and including qualified references to other metadata and data (I3). Finally, reusability is facilitated by describing metadata and data with accurate and relevant attributes (R1), including usage licences (R1.1), detailed provenance (R1.2) and using domain-relevant community standards (R1.3).

Data that is made FAIR (FAIRified data) has significant value in many areas. One such area is rare diseases, where projects such as the European Joint Programme on Rare Diseases (EJPRD) <sup>1</sup> interoperate FAIR data and metadata from different institutions for the benefit of rare disease research. Without FAIR, this inherently siloed and dispersed knowledge would be of reduced value, as it would not be large enough to answer research questions on its own.

The process of making data FAIR ('FAIRification') is organised in steps by FAIRification workflows (e.g., [3, 4, 5]). Nonetheless, neither the FAIR principles nor the FAIRification workflows mandate the use of any specific standard, format or software. This is because FAIR and FAIRification have been made agnostic to respect the unique requirements and needs that different communities face when managing and sharing data. Therefore, FAIR can be implemented in different manners and at different levels. However, this flexibility requires careful guidance throughout the FAIRification process to ensure that the implementation decisions (e.g., standards, metadata) align with the FAIRification objectives. In fact, the identification of FAIRification objectives is the initial and crucial step of several FAIRification workflows [6].

The task of properly identifying goals and requirements has been studied by the requirements engineering community from a software development perspective (e.g., [7, 8]). Literature in the area informs that the lack of proper planning and refinement of goals and requirements has a significant impact on the software development process. For instance, Pressman [7] points out that changing requirements after the software product has been delivered can cost up to 60 to 100 times more than changing a requirement during the software planning phase. We hypothesise that inadequate identification of FAIRification objectives may have a similar impact on planning and executing a FAIRification process. However, there is a lack of research on methods specifically focused on supporting FAIRification planning via the identification and refinement of FAIRification objectives. Furthermore, a recent study on the challenges of FAIRification concluded that clarifying goals prior to implementation is a key step in FAIRification, as it helps the team to make decisions that are consistent with its objectives [9].

To address the aforementioned gap we developed GO-Plan (Goal-Oriented FAIRification

---

<sup>1</sup><https://www.ejprarediseases.org/>

**Planning**), a method to plan FAIRification through a systematic identification and refinement of FAIRification objectives. The method reflects our understanding that distinct objectives can have different impacts on the planning and execution of FAIRification. Consequently, resources should be made FAIR at a level that aligns with the specific objectives of the FAIRification project. That is, resources should be made “FAIR enough” to fulfil the objectives of the involved collaborators<sup>2</sup>. Thus, the FAIRification planning should not only focus on the selection of suitable technologies or standards, but also on prioritising the effort required to raise the FAIR level of the targeted resources. Moreover, as FAIRification is a community-driven, aspirational and incremental process [2], these objectives must encompass the perspectives of collaborators directly participating in the project and also relevant external stakeholders (i.e., those who will eventually reuse the FAIRified resource). As such, each effort undertaken to make a resource FAIR (or more FAIR—FAIRer) for one’s own objectives will also make that resource FAIRer for others.

GO-Plan was designed based on good practices from requirements engineering (e.g., goal-oriented approaches [8], competency questions [11] and ontology engineering [12, 13]) while embedding our experiences from FAIRification projects, including training on FAIR [14], and conducting FAIRification within single [15] and among multiple institutions [9]. Additionally, the method has been optimised based on feedback obtained from a real-world application in developing a FAIR ontology catalogue [16].

The method hereby described is applicable to both *post-hoc* FAIRification [3], where existing resources are made FAIR, and *de novo* FAIRification [4], where resources are created FAIR (e.g., data made FAIR upon collection).

We discuss related works on Section 2. Then, we describe GO-Plan and illustrate it with a fictitious running example in Section 3. Finally, Section 4 discusses the strengths and weaknesses of our proposal, our impressions from its real-world application, and implications for future research. In the remainder of this paper, we use the spelling “(meta)data” to refer to both data and metadata. The words “goal” and “objectives” are used as synonyms. Note that the literature on FAIRification workflows usually uses the word “objective”, while the requirements engineering literature usually uses the word “goal”.

## 2. Related works

Several workflows and frameworks have been proposed to support FAIRification in different ways [6]. The generic [3] and the *de novo* [4] FAIRification workflows define the steps to be followed in the FAIRification of different types of FAIR resources, and both describe the identification of FAIRification objectives as the first step of FAIRification. Similarly, the FAIRplus FAIRification framework [5] defines steps to be followed during FAIRification and a work plan layout to support organising the FAIR implementation work. The first phase of this framework consists of setting “realistic and practical goals” [5], with focus on defining an acceptable “FAIR enough” state for the resource to be made FAIR. A valuable recommendation given by FAIRplus is to avoid “the word ‘FAIR’ and its derivatives in goals entirely as it is too general to impart

---

<sup>2</sup>When referring to collaborators, we align with the understanding of the similar term “stakeholder”, given by [10] as individuals, groups, or organizations that affect or are affected by a given project.

clear meaning” [5].

While these and other FAIRification workflows define a step for identifying FAIRification objectives [6], to the best of our knowledge, none of them have provided detailed guidance on defining FAIRification objectives or other FAIRification planning related aspects, such as distinguishing between the different types of stakeholders involved in FAIRification projects.

### 3. The goal-based FAIRification planning method

GO-Plan aims at supporting FAIRification planning by systematically defining mature FAIRification objectives through iterative steps. From our experience, we have found that starting with small steps and building on them is a more feasible approach than describing objectives from scratch. The method initially targets the most visible characteristics of the FAIRification project, such as the project domain, scope and available resources. It then leverages them to address more complex aspects such as relevant data concepts and competency questions. Finally, by following this structured and incremental approach, the method guides stakeholders towards the definition of comprehensive objectives that encompass all relevant aspects of FAIRification.

GO-Plan is organised in six phases, namely (i) FAIRification preparation, (ii) assessment of current FAIR supporting infrastructure and target resources, (iii) preparation of project collaborators, (iv) identification of domain scope and reuse stakeholders, (v) refinement of FAIRification goals and alignment to FAIR principles, and (vi) decision-making. These phases are refined in several steps and described in the sections that follow.

A distinction between two categories of stakeholders (collaborators) is made throughout the phases of the method: project stakeholders and reuse stakeholders. The former refers to those who are involved in the FAIRification project and have their own goals and requirements for it (e.g., data custodians, patient representative). The latter refers to those who will eventually reuse the FAIRified resource (e.g., researchers).

The method should be applied from the moment when the FAIRification project has already been idealised. For instance, when the organisation board members have already agreed on FAIRification for a certain need. At this stage, it is assumed that some aspects, such as the group of people that will be involved in the FAIRification project and the target resources, have already been defined. Moreover, GO-Plan is aimed at guiding people with varying levels of experience, from beginners to experts in FAIR and in goal-oriented elicitation of objectives. However, people with distinct levels of experience can use the method in different manners. For instance, a beginner would follow every step of the method to assure an effective identification of FAIRification objectives. In contrast, an expert leading a FAIRification project would use the method not only for identifying and refining the FAIRification objectives, but to also communicate the aspects of FAIRification with the rest of the team. Additionally, researchers, newcomers and educators can use the method as a knowledge source.

GO-Plan has been used to improve the FAIRness of a catalogue for ontology-driven conceptual modelling research, henceforth the OntoUML catalogue [17, 16], which contains a growing set of conceptual models defined using the OntoUML modelling language [18] or by extending the Unified Foundational Ontology (UFO) [19]. The OntoUML catalogue was initially built using an *ad hoc* FAIRification workflow, as reported in [17]. Later, the FAIR aspects of the

catalogue were reviewed using the method presented in this paper. The reader can refer to Sales *et al.* [16] for a detailed description of the method's application. The feedback received during the use of GO-Plan was utilised to adjust aspects related to the phases and steps phrasing (e.g., clarification), ordering (e.g., removing unnecessary steps) and artefacts produced (e.g., making the list of metadata concepts explicit). Our impressions on the application of the method are discussed in Section 4.

The following subsections describe GO-Plan using a running example of a research organisation that collects data about patients with rare diseases. This organisation has two aims: (i) to make legacy data FAIR (i.e. *post-hoc* FAIRification), and to implement an Electronic Data Capture System (EDC) that already creates FAIR data at the point of collection (i.e. *de novo* FAIRification). In addition to budget and deadline, the most important requirement for this project is the protection of patient privacy through controlled access to the data. The organisation wants to publish non-sensitive data and metadata to foster research on rare diseases.

### 3.1. Phase 1: FAIRification preparation

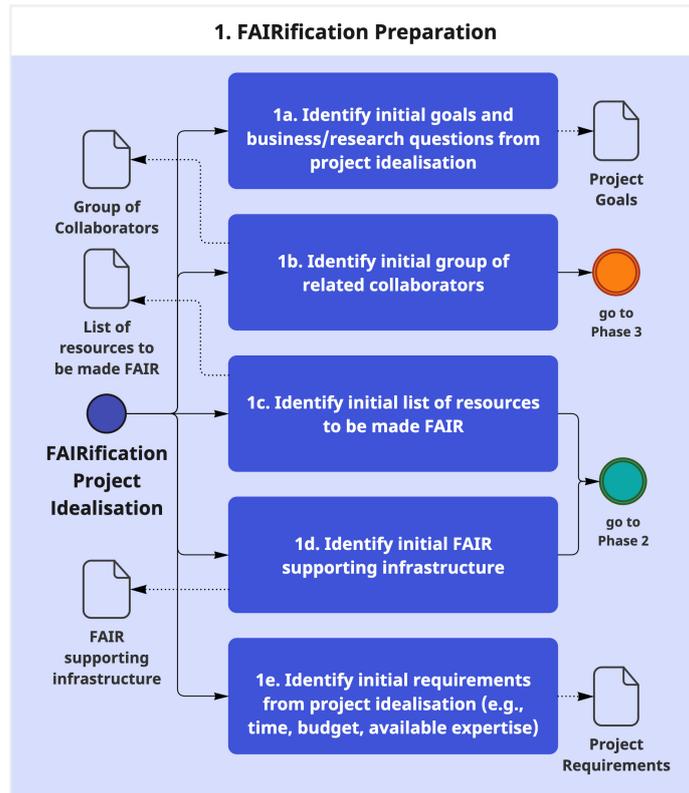
As shown in Figure 1, the method initiates with preparation tasks that entail examining the FAIRification project idealisation documents (e.g., grant proposals, kick-off slides, meeting minutes) and/or holding meetings with related stakeholders (e.g., managers, IT personnel) to identify artefacts that will support subsequent phases. The artefacts produced in all phases are described and exemplified in Table 1.

To illustrate, an analysis of the grant application for the rare diseases registry project is conducted to identify relevant stakeholders (step *1b*) and to determine the goals and requirements of the project (steps *1a* and *1e*), as exemplified in Table 1. In addition, conducting interviews with project leaders, patient representatives, and researchers can help to identify additional goals and requirements, as well as to identify what resources need to be made FAIR (i.e., legacy patient data and the EDC system) (*1c*). The organisation's information technology (IT) team, together with a FAIR expert, can assist in understanding the existing infrastructure (e.g., storage server for data and metadata, long term longevity plan for metadata) (*1d*) and determining the necessary adaptations required to accommodate the resource to be made FAIR (e.g., changes on the data storage format of the EDC system).

### 3.2. Phase 2: Assessment of FAIRification infrastructure

This phase addresses the resources to be made FAIR and the organisation's currently available FAIR supporting infrastructure. As shown in Figure 2, the resources to be made FAIR are assessed (step *2a*) to check if they can be retrieved (e.g., are they in a SQL server hosted locally? In a USB stick at the researcher's home office? Can the current EDC system be modified to generate ontologised data?), understood (e.g., are the headers of CSV files documented? Are the data elements collected by the current EDC system clear enough?) and if there are legal constraints in place (e.g., limited access due privacy-sensitive data).

Similarly, the current infrastructure that will accommodate the FAIR resource needs to be reached and assessed (*2c*) to check if it can be used, if it needs to be adapted and/or if additional infrastructure needs to be arranged. The type of infrastructure may vary depending on the type



**Figure 1:** Phase 1 of the method. Steps are described inside the phases using the phase number and step letter (e.g., 1a, 1b) as identifiers. The figure follows a BPMN notation [20].

of FAIR resource it is intended to support. For example, to make data FAIR, the infrastructure may include storage servers for data and metadata, and data capturing systems (that might have to be adapted). In the case of privacy-sensitive data, an access control system must be incorporated. Similarly, to make an ontology FAIR, the infrastructure may involve an ontology repository and a metadata server. In the case of software, it can include a software code repository and a version control system.

The primary aim of these steps is to ensure that both the resources to be made FAIR and the current infrastructure intended to accommodate the FAIR resource do not pose any obstacles to FAIRification. This involves verifying, for instance, the availability and capability of storage servers to handle the data volume associated with FAIRification, among other considerations. If any issues are identified in this phase, they must be addressed before continuing to the next phase (steps 2b and 2d).

Finally, at this stage, the team must have enough information to decide whether a retrospective or and *de novo* FAIRification must be planned for the resources identified. For instance, if a patient registry needs to make existing data FAIR, but also needs to start generating FAIR data as it is collected, then both retrospective and *de novo* FAIRification will need to be planned.

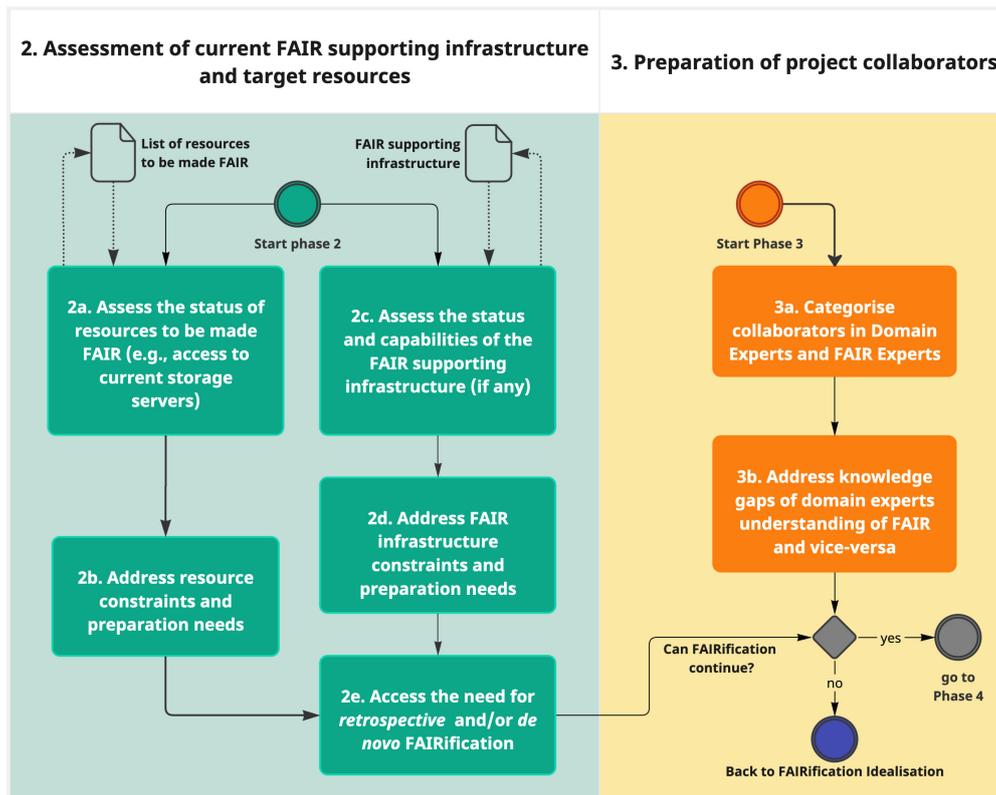
**Table 1**

List of artefacts produced during the method.

| Artefact   | Description  | Created (phase) | Used (phase) | Examples  |
|--|--|-----------------|--------------|---|
| <b>Project collaborators</b>                                 | List of people actively contributing to the FAIRification project  | 1               | 3, 6         | Domain expert (clinician, researcher), FAIR expert (semantic modeller, developer)   |
| <b>Project related goals and business/research questions</b> | Goals and business/research questions that can impact or be affected by the FAIRification, extracted from a larger project that includes FAIRification | 1               | 4, 6         | “Collect patient data to foster research with Rare Diseases”  |
| <b>Project requirements</b>                                  | Project requirements that will constrain the FAIRification   | 1               | 4, 6         | Budget, deadline, data privacy constraints, interoperability requirements (e.g., “must interoperate with the EJP RD Metadata Model [21]”) |
| <b>Resources to be made FAIR</b>                             | Pre-existing resources that will be made FAIR or modified to generate FAIR data during the FAIRification project                                       | 1               | 2            | Legacy data, data collection systems, software, ontology catalogues   |
| <b>FAIR supporting infrastructure</b>                        | Pre-existing infrastructure that has been allocated to accommodate the FAIR resource   | 1               | 2, 6         | Current storage servers, access control systems, longevity plans  |
| <b>Domain description</b>                                    | Description of the domain of the resource to be made FAIR  | 4               | 4            | Textual description of the domain, in the format of an abstract for a scientific journal  |
| <b>Semantic types</b>  | Groups of related concepts of similar meaning  | 4               | 5            | “Pain: discomfort, ache, soreness.”   |
| <b>Reuse stakeholders</b>                                    | Eventual reusers of the FAIRified resource   | 4               | 5            | Researchers, data publishers, patient representatives   |
| <b>FAIRification Competency questions</b>                    | Questions that must be answered by the FAIRified resource  | 5               | 5            | see Table 2   |
| <b>Metadata concepts</b>                                     | Metadata terms to be collected for the FAIR resource   | 6               | -            | License, provenance   |

### 3.3. Phase 3: Preparation of FAIRification stakeholders

The third phase of the method focuses on identifying and preparing the people who will be involved in the FAIRification project. For this, the list of the initial project collaborators is used. The main aim of this task is to bridge the knowledge gap between domain and FAIR experts to prepare them for subsequent phases. The motivation for this comes from the work of Neuhaus & Hastings [12], who suggests techniques to involve stakeholders in the ontology development process. By engaging the project collaborators into each other’s domain, we reuse the authors’ proposed techniques of “creating micro-level consensus” (micro-level: project scope), which is



**Figure 2:** Phases 2 to 3 of the method.

expected to establish a more inclusive participatory environment for the discussion of objectives.

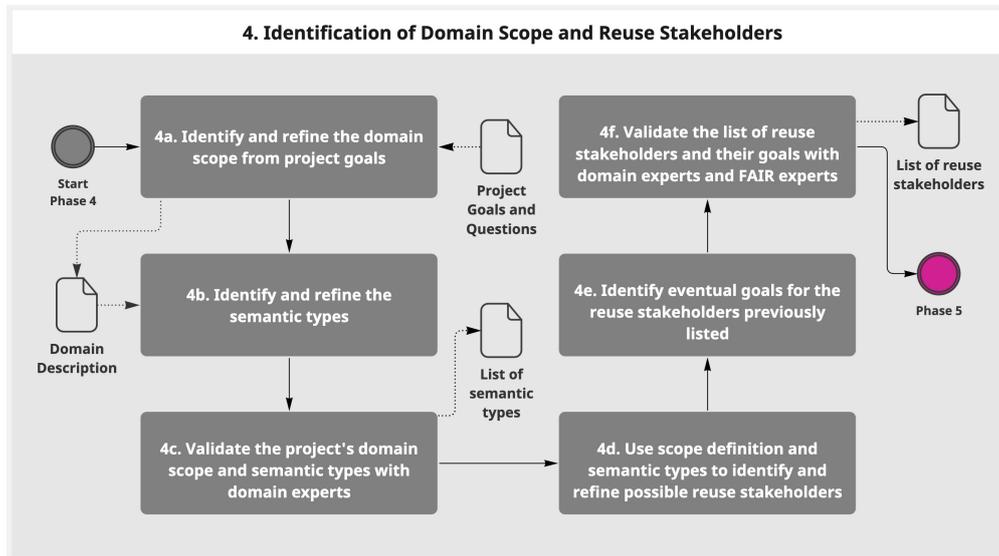
In this phase, the group of project collaborators is categorised into FAIR experts and domain experts (3a). Then, relevant knowledge gaps between them are assessed to an extent that allows for sufficient understanding of each other’s expertise (3b). This will create a common “ground language” for stakeholders to communicate their own objectives.

To exemplify, FAIR experts involved in our example project (i.e., rare disease registry FAIRification) could have a question-and-answer session with domain experts about common data elements for rare disease registration [22]. Meanwhile, domain experts get a short lecture on the basics about the FAIR principles and what can be expected and done with FAIR data. We outline that, for the sake of expectation management, it is important to inform domain experts about what is possible with FAIR and what should not be expected as output from a FAIRification project. For instance, while FAIR data may facilitate it, a data visualisation dashboard is an unusual output of FAIRification.

### 3.4. Phase 4: Identification of domain scope and groups of reuse stakeholders

Phase 4 relies on the premise that reuse is the ultimate aim of FAIR, and therefore the FAIRification objectives must consider eventual reuse case scenarios. As shown in Figure 3, the list of

project goals and research/business questions are input in this phase to identify and describe the domain scope (4a). For instance, rare diseases are the domain of the rare disease registry FAIRification project, while the scope refers to a subset of the domain that considers only the terms of interest for the FAIRification project (e.g., information from patients with rare diseases including treatment procedures may be within the scope, while other medical information unrelated to the rare disease might be out of the scope).



**Figure 3:** Phase 4 of the method.

This phase also consists of identifying semantic types pertaining to the scope (4b). We refer to semantic types as groups of concepts of similar meaning (e.g., *pain* is a semantic type group that covers similar concepts such as *discomfort*, *ache*, and *soreness*). In our running example, semantic types would include *patient*, *treatment*, *diagnosis* and *genetic information*. These would also be useful in later stages of FAIRification (i.e., conceptual modelling of (meta)data). Next, on step 4c, the semantic types and their definitions are discussed and agreed upon by the group of domain experts. During the agreement process, they may identify additional semantic types to be added to the list.

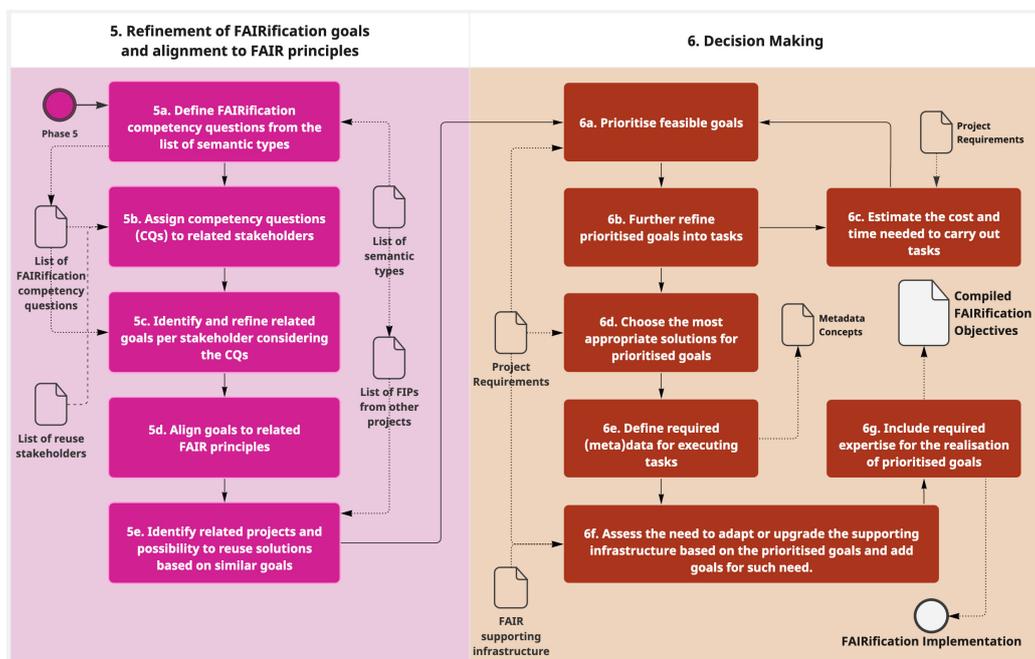
In step 4d, the description of the domain and semantic types is used to identify reuse stakeholders. To illustrate, a researcher and a healthcare provider are examples of stakeholders who will reuse patient, diagnosis and treatment data from the rare disease patient registry. Next, the expected goals of the reuse stakeholders when reusing the FAIR resource are predicted by the FAIR project stakeholders (4e). For instance, using the data to “identify cohorts for clinical trials” may be a goal of the researcher towards the rare disease patient registry. Other examples of reuse stakeholders can be patient representatives, clinicians and healthcare providers. The list of reuse stakeholders and their goals should also be validated with domain experts (4f).

Note that, in step 4d, it should not be expected a fully comprehensive list of stakeholders, as it would be very difficult to predict all eventual reuse cases. However, the FAIRification planning team should strive for creating a list that considers relevant expected cases. We also

point out that later project extensions to incorporate more reuse cases should be technically feasible given the flexibility of FAIR resources.

### 3.5. Phase 5: FAIRification goals refinement and alignment to FAIR principles

As depicted in Figure 4, the fifth phase of the method starts by reusing the list of semantic types defined in the previous phase to identify competency questions (CQs) [11] that should be answered by the FAIR resource (5a), including the metadata of the resource. In the context of a FAIRification project, a CQ should be a question that cannot be answered without the FAIR resource, or that can be answered in a significantly easier manner with the FAIR resource. We suggest that CQs elicited in this step should be complex enough to connect and explore the relationship between different semantic types. Table 2 shows some examples of CQs that can be defined for the semantic types exemplified in Section 3.4. In step 5b, the CQs are assigned to related stakeholders (i.e., reuse stakeholders and relevant project stakeholders) and further refined as objectives (5c). These objectives can be identified by asking **why** a certain CQ needs to be answered and **how** it can be answered. Some objectives are also exemplified in Table 2.



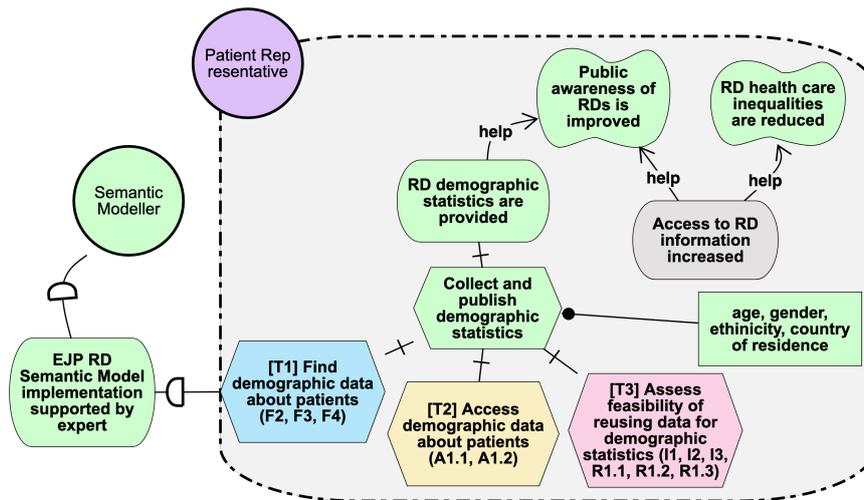
**Figure 4:** Phases 5 and 6. The final set of compiled FAIRification objectives is depicted as the white coloured artefact.

The objectives identified from the CQs are then aligned with related principles (5d). For this step, it should be identified which and how a FAIR principle will support achieving a specific objective. For instance, the objective “public awareness of rare diseases is improved” (Figure 5), which is further refined until it can be realised by the task “collect and publish demographic statistics”, may be supported by F2 (rich metadata to make the patient registry findable) and R1.1 (data licence to allow reuse of the data for demographic statistics). Meanwhile, other principles

**Table 2**

Examples of competency questions (CQs), related stakeholders and their objectives.

| CQ  | Related Stakeholder    | Objectives  |
|---|------------------------|---|
| What is the age range and gender distribution of patients with a particular rare disease in Europe? | Patient Representative | “Public awareness of Rare Diseases is improved”                                     |
|   | Health Care Provider   | “Patient management is improved”  |
| What previous diagnoses and treatments have been tried for patients with a particular rare disease? | Researcher             | “Cohorts for clinical trials are identified”,<br>“Disease progression is predicted” |



**Figure 5:** Modeling the objectives of a patient representative using *iStar*. Rounded shapes represent goals (objectives), cloud-like shapes represent qualities, hexagons represent tasks and rectangles represent resources reused by tasks. Elements in dark grey are not prioritised. The dashed line defines the boundaries of the stakeholder intentions.

(e.g., F1) may not be prioritised for this specific objective.

To facilitate the management of objectives, we suggest the use of goal-modelling techniques such as *iStar* [23], which helps to capture the stakeholders intentions and their relationships in a structured way. Models created with *iStar* include concepts such as actors, goals, tasks, resources, and relationships such as decomposition and contribution links. The reader is referred to [23] for further information on *iStar*.

The final step of this phase consists of using the list of semantic types to identify related FAIRification projects (5e) through, for instance, the use of FAIR Implementation Profiles (FIPs) [24] or catalogues such as FAIRSharing [25]. FIPs are specifications of implementation solutions for realising the FAIR principles in a specific context or domain, and their use is intended to foster convergence on FAIR implementation decisions [24]. Another example of a knowledge source for implementation solutions includes the Smart Guidance RD Wizard, a questionnaire-based tool to guide data stewards in making rare disease patient registries

FAIR [26]. In the context of GO-Plan, related projects can support collecting implementation solutions that can be reused in the FAIRification project. The EJP RD project [21] is such a project to our running example.

### 3.6. Phase 6: Decision making

The sixth and last phase of the method starts by prioritising feasible objectives (6a) given the project requirements (e.g., data privacy) and constraints (e.g., budget, deadline, available expertise). At this point, prioritisation also includes removing objectives that are not feasible, may not be supported by FAIR principles or are not related to FAIRification. Then, the prioritised objectives are further refined (6b) and tasks required to realise them are elicited. Here it is recommended that the team estimates the cost and time associated with the elicited tasks to assist in further prioritisation of goals given the project requirements (6c).

Next, the most appropriate solutions for prioritised objectives are identified and selected considering the project goals, requirements, expertise and the limitations of available supporting infrastructure (6d). This step can be supported by reusing solutions from the similar projects identified in step 5e, by consulting experts on FAIR or by querying resources such as FAIR-Sharing and the Smart Guidance RD Wizard. Next, the necessary (meta)data for achieving the identified tasks are listed (6e) and described in the goal diagrams as resources, as exemplified in Figure 5. Subsequently, the team needs to assess whether there is a need to adapt the supporting infrastructure for the prioritised goals and, if so, add goals to address this need (6e). Finally, the expertise required for the implementation of the selected solutions (6f) is defined.

To illustrate, the reuse of the EJP RD Metadata Model is a possible implementation choice for the objectives depicted in Figure 5 (in the context of F2 – “Find demographic data about patients”) given the project requirements, and a semantic modelling expert would be a required expertise to support reusing this solution.

At this point, the goal diagram should contain enough information to inform and guide FAIRification. The FAIRification objectives, tasks and chosen implementation solutions can now be seen as actions to be taken towards realising FAIRification. It is upon the experts conducting the FAIR project to prioritise tasks and define implementation cycles and evaluation activities. We suggest using a FAIRification workflow to organise the FAIRification process that follows.

## 4. Final remarks

The method presented in this paper is defined with sequential phases and steps. However, we have observed that real-world applications, such as the one described in Sales *et al.* [16], may benefit from an agile approach. In this case, the method can be fitted into one iteration and executed several times, or have its phases broken down into different cycles that can be executed iteratively until the outputs of those phases are satisfactory. For instance, in a first iteration, the process of creating the competency question can raise the need to include more semantic types, which can be addressed during the method’s re-execution in a second cycle, or in a re-execution of phase 5. It is up to the FAIRification team to decide how many iterations should be performed considering the project constraints (especially budget and time).

Additionally, distinct FAIRification iterations can be tailored to address the specific needs and considerations of different stakeholders, thereby defining different levels of FAIR and related aspects for them. That is particularly valuable, for instance, when dealing with sensitive data (e.g. some types of users have access to different portions of data) or with FAIRification projects involving non-public data (e.g. from private companies), where certain reuse stakeholders might have limited access to the (meta)data.

We acknowledge the need for a more detailed evaluation of the expected benefits of the method when compared to *ad hoc* FAIRification. We are currently working on evaluating GO-Plan from a usability perspective, where we will study the perception of users when using the method (i.e., “is it easier and more efficient to define FAIRification objectives using GO-Plan compared to *ad hoc* FAIRification planning?”). In addition, we emphasise that the method is based on techniques from software engineering that have already been evaluated and used in several real-world applications (e.g., [27, 28]).

When applying the method to a real-world use case [16], we observed a significant influence of the definition of reuse stakeholders on the results of FAIRification, particularly in identifying which (meta)data concepts should be collected and published, as well as considerations regarding licensing and provenance. We attribute this impact to the fundamental emphasis of FAIR on facilitating reusability and assert that optimising the resource for reuse cases is key to effective FAIRification. Furthermore, we also observed that using goal model diagrams has facilitated the communication among collaborators.

When comparing the real-world use case with [16] and without [17] the use of our method, we noticed that our approach led to more informed and clearer decision-making and evaluation of the FAIRness of the catalogue. The stakeholders were able to prioritise solutions based on a comprehensive understanding of the relationship between objectives and the FAIR principles. To illustrate, the use of our method resulted in a re-definition of metadata concepts to be collected, a reprioritisation of the principles (e.g., more attention was given to R1), and the inclusion of FAIR supporting infrastructure such as the FDP. Finally, we observed that the objectives helped stakeholders in establishing achievement criteria for principles that lacked sufficient precision. For instance, the team was able to define a metadata set that would satisfy the “data are described with rich metadata” (F2) principle by ensuring that it supported all prioritised goals from the reuse stakeholders.

The main aim of the work presented in this paper is to help all FAIR enthusiasts to better define clear FAIRification objectives and plans that can lead to successful FAIRification. Nonetheless, we argue that communities should actively endeavour to share their FAIRification planning artefacts (e.g., goal diagrams, implementation decisions, FIPs) in order to accelerate standards convergence, disseminate solutions to implementation challenges, and share experiences so that others can prepare and execute FAIRification faster and more seamlessly. To support this, we propose that FAIRification plans, including goals and mappings to related principles, should also be made FAIR. In addition to that, we emphasise the publication of FAIR implementation decisions (i.e. FIPs) as an effective means to gradually diminish the work for subsequent projects and (re)users. This will also allow future work to focus on creating a catalogue of FAIRification plans and associated concrete tasks that can lead to improved automation.

## Acknowledgments

We thank the LUMC Biosemantics and the EJP RD FAIRification Stewards groups for constant feedback on this research. This initiative has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°825575 and the Trusted World of Corona (TWOC; LSH Health Holland).

## References

- [1] A. K. Thomer, D. Akmon, J. J. York, A. R. Tyler, F. Polasek, S. Lafia, L. Hemphill, E. Yakel, The craft and coordination of data curation: Complicating workflow views of data science, *Proceedings of the ACM on Human-Computer Interaction* 6 (2022) 1–29.
- [2] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., The FAIR guiding principles for scientific data management and stewardship, *Scientific data* (2016).
- [3] A. Jacobsen, R. Kaliyaperumal, L. O. Bonino da Silva Santos, B. Mons, E. Schultes, M. Roos, M. Thompson, A generic workflow for the data FAIRification process, *Data Intelligence* (2020).
- [4] K. H. Groenen, A. Jacobsen, M. G. Kersloot, B. dos Santos Vieira, E. van Enckevort, R. Kaliyaperumal, D. L. Arts, P. A. t Hoen, R. Cornet, M. Roos, et al., The de novo FAIRification process of a registry for vascular anomalies, *Orphanet Journal of Rare Diseases* (2021).
- [5] D. Welter, N. Juty, P. Rocca-Serra, F. Xu, D. Henderson, W. Gu, J. Strubel, R. T. Giessmann, I. Emam, Y. Gadiya, et al., Fair in action—a flexible framework to guide fairification, *Scientific Data* 10 (2023) 291.
- [6] B. dos Santos Vieira, C. H. Bernabé, I. Henriques, S. Zhang, A. B. Camara, J. A. R. García, J. van der Velde, P. van Damme, P. A. Moreno, N. Benis, J. Strubel, F. Schoots, P. L'Henaff, P. 't Hoen, M. Roos, A. Jacobsen, R. Cornet, M. D. Wilkinson, F. Schaefer, M. Swertz, M. Jetten, Critical steps towards large-scale implementation of the FAIR data principles, 2023. URL: <https://doi.org/10.5281/zenodo.7867293>.
- [7] R. S. Pressman, *Software engineering: A practitioner's approach*, 7th ed., McGraw-Hill, 2010.
- [8] A. Van Lamsweerde, Goal-oriented requirements engineering: A guided tour, in: *Proceedings fifth ieee international symposium on requirements engineering*, IEEE, 2001, pp. 249–262.
- [9] B. dos Santos Vieira, C. H. Bernabé, S. Zhang, et al., Towards FAIRification of sensitive and fragmented rare disease patient data: Challenges and solutions in european reference network registries, *Orphanet Journal of Rare Diseases* 17 (2022) 436.
- [10] R. E. Freeman, *Strategic management: A stockholder approach*, Pitman, 1984.
- [11] M. Grüniger, M. S. Fox, *The role of competency questions in enterprise engineering, Benchmarking—Theory and practice* (1995).
- [12] F. Neuhaus, J. Hastings, *Ontology development is consensus creation, not (merely) representation*, *Applied Ontology* (2022). Preprint.

- [13] R. de Almeida Falbo, Sabio: Systematic approach for building ontologies., *Onto. Com/odise@ Fois 1301* (2014).
- [14] C. H. Bernabé, L. Thielemans, C. Carta, et al., Building expertise on FAIR through evolving Bring Your Own Data (BYOD) workshops: Describing the data, software, and management focused approaches and their evolution, 2023. Manuscript in preparation.
- [15] N. Queralt-Rosinach, R. Kaliyaperumal, C. H. Bernabé, et al., Applying the FAIR principles to data in a hospital: challenges and opportunities in a pandemic, *Journal of Biomedical Semantics* (2022).
- [16] T. P. Sales, P. P. F. Barcelos, C. M. Fonseca, et al., A FAIR catalog of ontology-driven conceptual models, 2023. Manuscript submitted to *Data & Knowledge Engineering*.
- [17] P. P. F. Barcelos, T. P. Sales, M. Fumagalli, et al., A FAIR model catalog for ontology-driven conceptual modeling research, in: *Conceptual Modeling. ER 2022*, volume 13607, Springer, 2022, p. 3–17.
- [18] G. Guizzardi, C. M. Fonseca, A. B. Benevides, et al., Endurant types in ontology-driven conceptual modeling: Towards OntoUML 2.0, in: *Conceptual Modeling. ER 2018*, volume 11157, Springer, 2018, p. 136–150.
- [19] G. Guizzardi, A. Botti Benevides, C. M. Fonseca, et al., UFO: Unified Foundational Ontology, *Applied Ontology* 17 (2022) 167–210.
- [20] OMG, Business Process Model and Notation (BPMN), Version 2.0, 2011. URL: <http://www.omg.org/spec/BPMN/2.0>.
- [21] European Joint Programme for Rare Diseases, EJP-RD VP Resource Metadata Schema, <https://github.com/ejp-rd-vp/resource-metadata-schema>, 2021. Accessed on April 24, 2023.
- [22] EU RD Platform, Set of common data elements, [https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements\\_en](https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en), accessed 2023.
- [23] F. Dalpiaz, X. Franch, J. Horkoff, iStar 2.0 language guide, arXiv preprint arXiv:1605.07767 (2016).
- [24] E. Schultes, B. Magagna, K. M. Hettne, et al., Reusable FAIR implementation profiles as accelerators of FAIR convergence, in: *Advances in Conceptual Modeling. ER 2020*, volume 12584, Springer, 2020.
- [25] S.-A. Sansone, P. McQuilton, P. Rocca-Serra, et al., FAIRsharing as a community approach to standards, repositories and policies, *Nature Biotechnology* (2019).
- [26] P. van Damme, P. Alarcón Moreno, A. Cámara Ballesteros, C. H. Bernabé, C. M. A. Le Cornec, B. Dos Santos Vieira, K. J. van der Velde, S. Zhang, C. Carta, R. Cornet, P. A. 't Hoen, A. Jacobsen, M. A. Swertz, M. Roos, N. Benis, A resource for guiding data stewards to make european rare disease patient registries fair, *Data Science Journal* (2023). Manuscript submitted for publication.
- [27] C. Pacheco, I. García, M. Reyes, Requirements elicitation techniques: A systematic literature review based on the maturity of the techniques, *IET Software* (2018).
- [28] J. Horkoff, F. B. Aydemir, E. Cardoso, et al., Goal-oriented requirements engineering: An extended systematic mapping study, *Requirements engineering* 24 (2019) 133–160.