

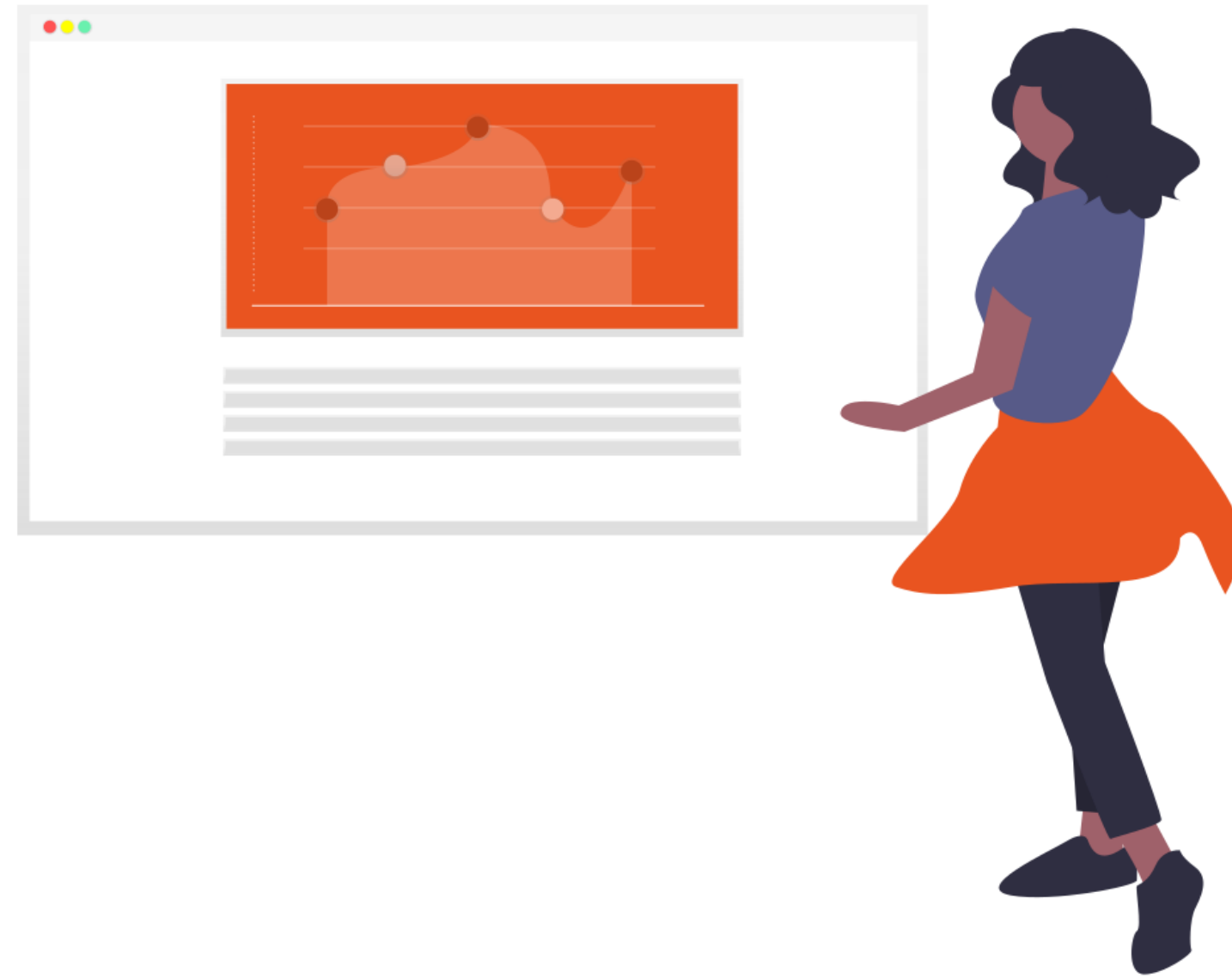


# Data Stewardship Wizard Workshop

# Outline

- Introduction
- DSW for Researchers + Demo
- DSW for Data Stewards + Demo
- How to Get Started
- Questions & Discussion

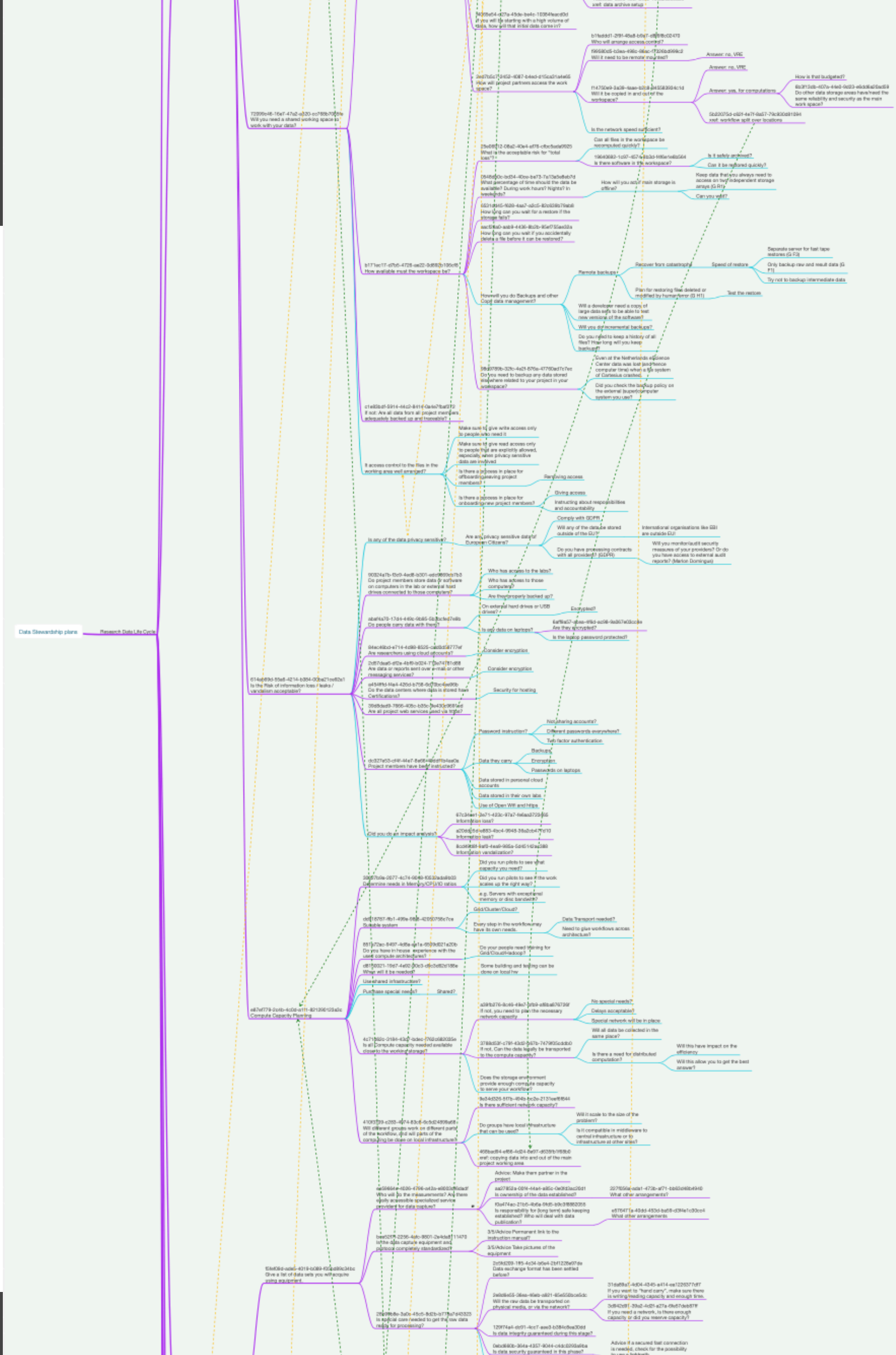




# Introduction

# Data Stewardship Mindmap

- A mind map of data stewardship topics
- Created by Rob Hooft
- <https://doi.org/10.5281/zenodo.2614820>



# Data Stewardship Wizard

- A tool for generating data management plans
- Using smart questionnaires to guide researchers
- Customizable questionnaire templates
- Can be used in other domains





# DSW for Researchers

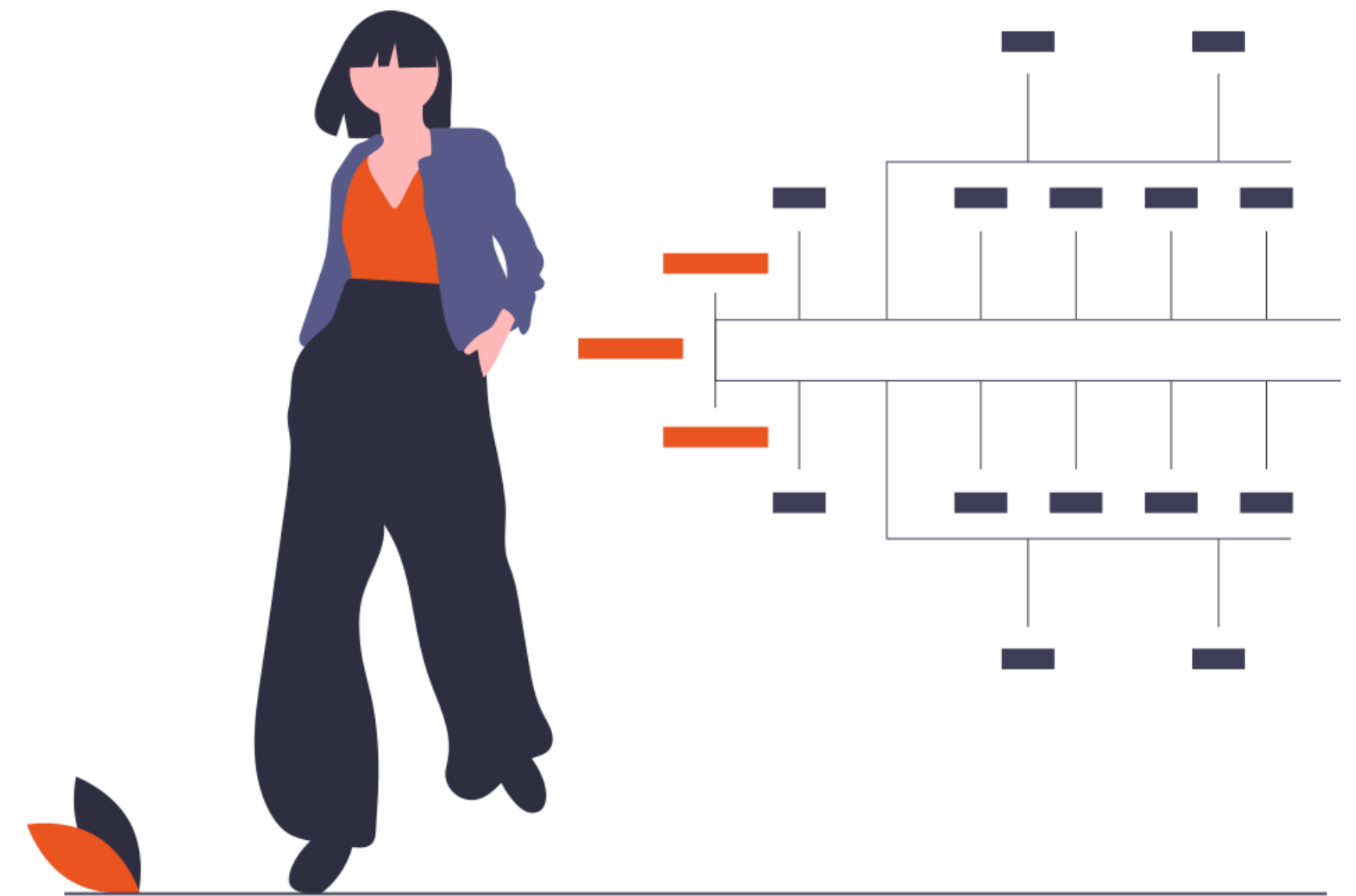
# Researchers Workflow



**Knowledge Model**

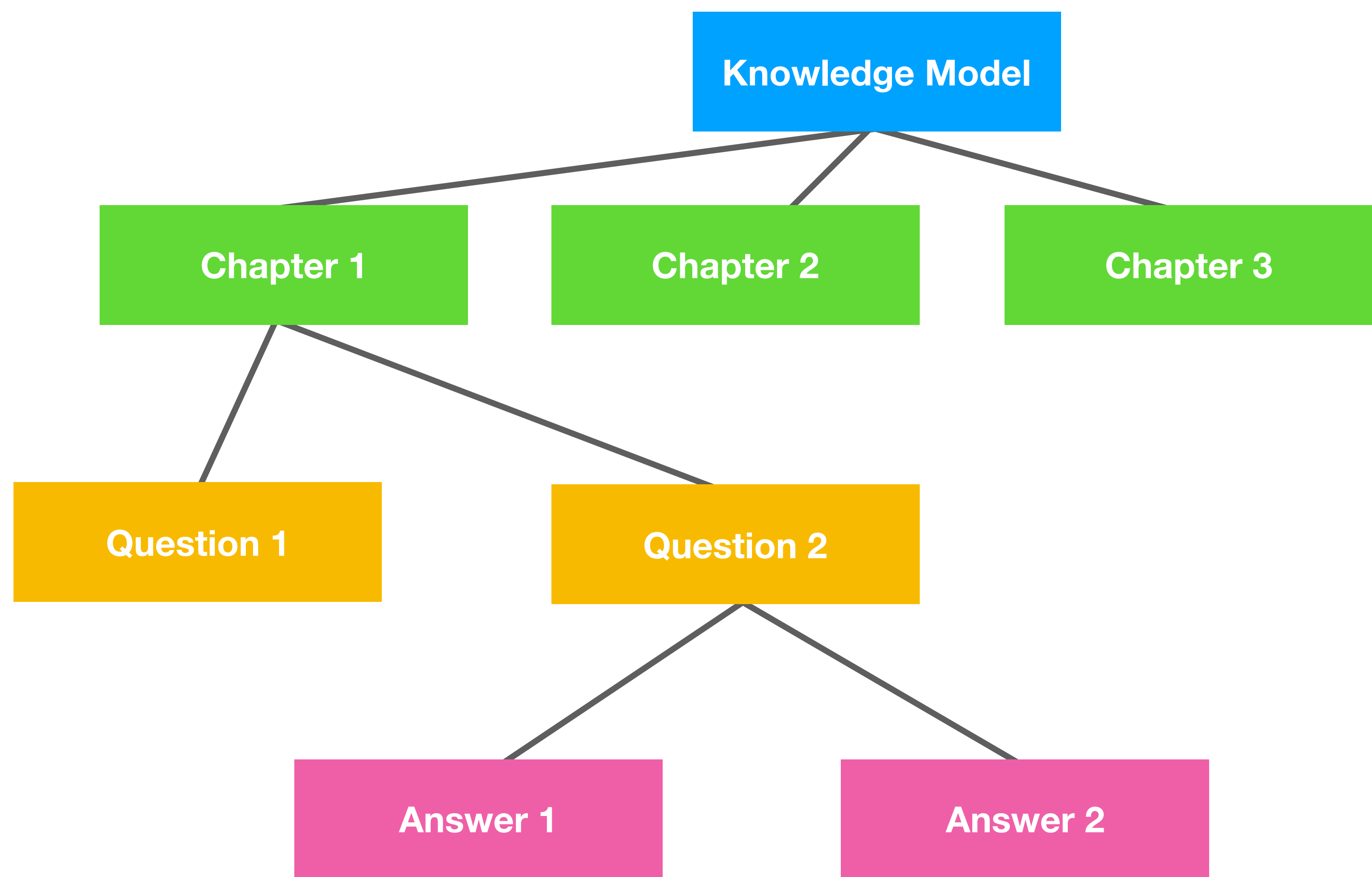
# Knowledge Model

- Contains the **knowledge** about what should be asked and how
- **Tree-like structure** of Chapters, Questions, Answer and References or Experts
- “**Template**” for the Questionnaire
- Provided by DSW or institution data stewards





# Knowledge Model



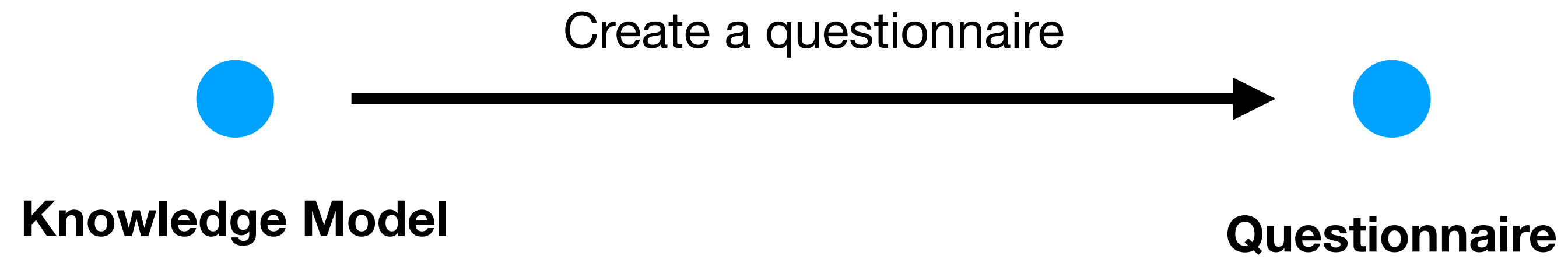
- ▼ Core DS Knowledge Model
  - ▼ Design of experiment
    - ▼ Is there any pre-existing data?
      - ☒ No
      - ▶ ☒ Yes
      - atq
    - ▶ Will reference data be created?
    - ▶ Will you be storing samples?
    - ▶ Will you be collecting experimental data?
  - ▼ Data design and planning
    - ▼ What data formats/types will you be using?
      - Data format/type:
        - ▶ Is this a standard data format used by others too?
      - ▼ Does this data format enable sharing and long term archiving?
        - ▶ ☒ No
        - ☒ Yes
      - njy
    - ▶ Will you be using new types of data?
    - ▶ How will you be storing metadata?
    - ▶ During the project, will you be archiving data (using so-called 'cold storage')?
    - ▶ Will you need a shared working space to work with your data?
    - ▶ Is the risk of information loss, leaks and vandalism acceptably low?
    - ▶ Do you need to do compute capacity planning?
  - ▶ Data Capture/Measurement
  - ▶ Data processing and curation

# Researchers Workflow



**Knowledge Model**

# Researchers Workflow



# Questionnaire

- Interactive form how to get the answers from users based on the Knowledge Model
- Contains report on the FAIR metrics
- Answers can be exported to a document (e.g., PDF)

Chapters

I. Design of experiment1

II. Data design and planning7

III. Data Capture/Measurement3

IV. Data processing and curation4

V. Data integration7

VI. Data interpretation3

VII. Information and insight11

More

[Summary Report](#)

2 Will reference data be created?

Will any of the data that you will be creating form a reference data set for future research (by others)?

Data Stewardship for Open Science: [rbz](#)

☒ a. No

☐ b. Yes

Clear answer

3 Will you be storing samples?

Data Stewardship for Open Science: [kuz](#)

☐ a. No

☒ b. Yes

Clear answer

4 Will you be collecting experimental data?

Data Stewardship for Open Science: [csx](#)

☐ a. No

☒ b. Yes

# Questionnaire Accessibility

- Define who and how can access the questionnaire



# Questionnaire Accessibility: Private

- Use when you don't want to share it with anyone
- You: View and Edit
- Other users: Nothing

**Accessibility**

- ☒ **Private**  
Questionnaire is visible only to you.
- ☐ **Public Read-Only**  
Questionnaire can be viewed by other users, but they cannot change it.
- ☐ **Public**  
Questionnaire can be accessed by all users.

# Questionnaire Accessibility: Public Read-Only

- Use when you want to share your answers with other users
- Can be used for template questionnaires
- You: View and Edit
- Other users: View

**Accessibility**

- ☐ **Private**  
Questionnaire is visible only to you.
- ☒ **Public Read-Only**  
Questionnaire can be viewed by other users, but they cannot change it.
- ☐ **Public**  
Questionnaire can be accessed by all users.

# Questionnaire Accessibility: Public

- Use when you want to collaborate with other users on the questionnaire
- You: View and Edit
- Other users: View and Edit

**Accessibility**

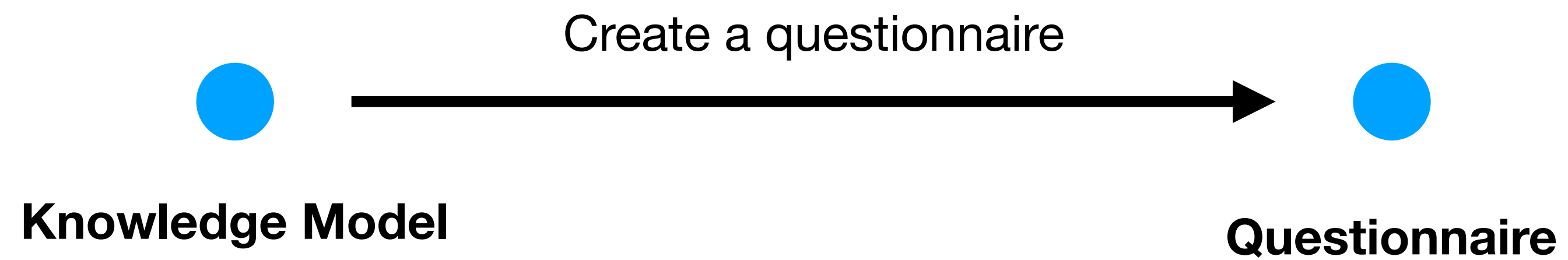
☐ **Private**  
Questionnaire is visible only to you.

☐ **Public Read-Only**  
Questionnaire can be viewed by other users, but they cannot change it.

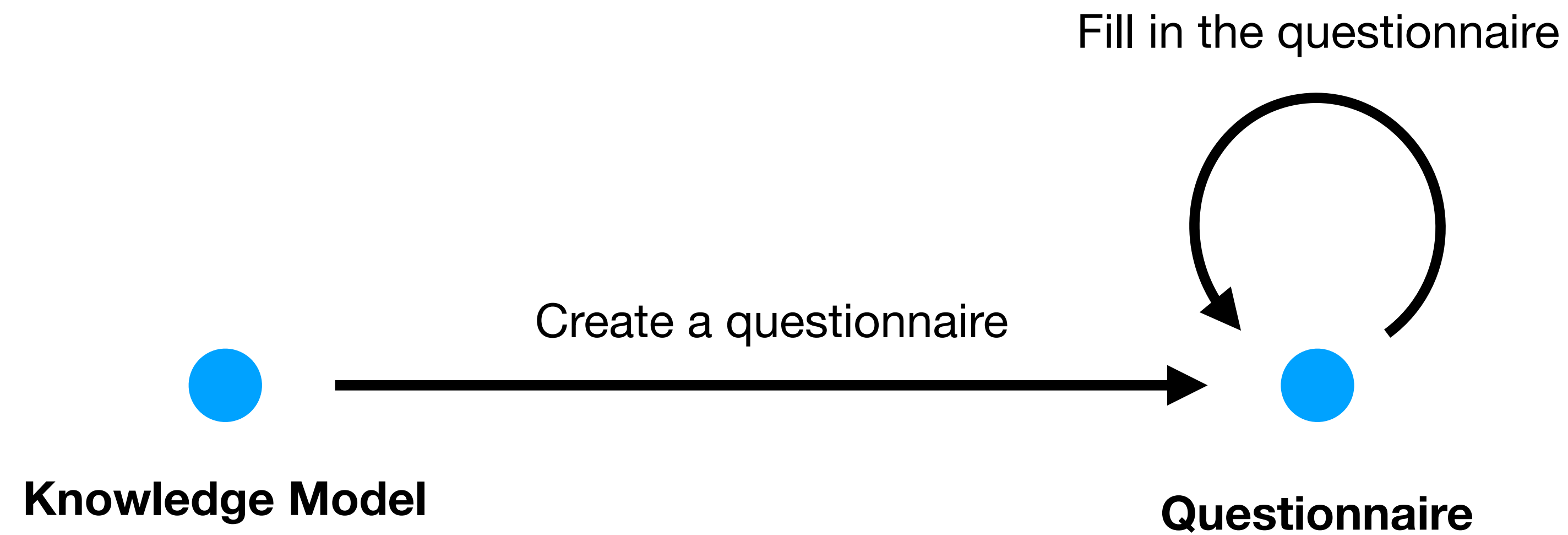
☒ **Public**  
Questionnaire can be accessed by all users.



# Researchers Workflow



# Researchers Workflow



# Filling the Questionnaire

- Free navigation through the Questionnaire (no given order of answering)
- Can be saved anytime

The screenshot displays the 'Data Stewardship for Open Science' questionnaire interface. On the left, a 'Chapters' sidebar lists seven sections: I. Design of experiment (5), II. Data design and planning (7), III. Data Capture/Measurement (3), IV. Data processing and curation (4), V. Data integration (7), VI. Data interpretation (3), and VII. Information and insight (11). Below this is a 'More' section with a 'Summary Report' link. The main content area shows the 'a. No' option selected for the question 'Did you research all the data that exists?'. Below this, question '1.b.2 Do you need to harmonize different sources of existing data?' is shown with 'b. Yes' selected. A sub-question '1.b.2.b.1 Will you be making your harmonization results available to others?' is partially visible at the bottom.

**Chapters**

- I. Design of experiment 5
- II. Data design and planning 7
- III. Data Capture/Measurement 3
- IV. Data processing and curation 4
- V. Data integration 7
- VI. Data interpretation 3
- VII. Information and insight 11

**More**

- [Summary Report](#)

**Data Stewardship for Open Science: *ezi***

☒ a. No

☐ b. Yes

[Clear answer](#)

Did you research all the data that exists? You may not be aware of all existing data that could be available. Although using and/or integrating existing data sets may pose a challenge, it will normally be cheaper than collecting everything yourself. Even if you decide not to use an existing data set, it is better to do this as a conscious decision.

**1.b.2 Do you need to harmonize different sources of existing data?**

If you are combining data from different sources, harmonization may be required. You may need to re-analyse some original data.

**Data Stewardship for Open Science: *whi***

☐ a. No

☒ b. Yes

[Clear answer](#)

**1.b.2.b.1 Will you be making your harmonization results available to others?**

By publishing either exactly what you did or (better) make sure that the harmonized data is available for reuse, you may save others the effort

# Filling the Questionnaire

- Only relevant questions are asked based on previous answers

1

**Will you be using a shared working space to work with your data?**

+

!

Will you be using a working space that is shared between all the people working on the data in the project? Sometimes such a system is called a *Virtual Research Environment*.

☒ Desirable: *Before Submitting the Proposal*

☐ a. No

☐ b. Yes

# Filling the Questionnaire

## 1 Will you be using a shared working space to work with your data? + !

Will you be using a working space that is shared between all the people working on the data in the project? Sometimes such a system is called a *Virtual Research Environment*.

☒ Desirable: *Before Submitting the Proposal*

☐ a. No ⋮

☒ b. Yes ⋮

 Clear answer

### 1.b.1 Will this work space be run by dedicated specialists? + !

If your work space is run and maintained by specialists, e.g. the ICT department of one of the institutes involved in the projects, this means that backup and restore as well as access management is properly addressed.

☒ Desirable: *Before Submitting the DMP*

☐ a. No

Accessibility

☐ b. Yes

Accessibility

# Data Stewardship for Open Science

- By Barend Mons
- References from questions in the Common DSW Knowledge Model



## Data Stewardship for Open Science: Chapter 1.1

With kind permission of  
 **CRC Press**  
Taylor & Francis Group

## Is there pre-existing data?

### What's up?

For many decades if not centuries, virtually every experiment started with the collection or creation of 'observations' and in fact data. In social sciences and humanities the tendency to 'reuse' data that had been created earlier, in all kinds of surveys and increasingly of course from sources such as social media maybe already somewhat more established. However, in many of the hard experimental sciences, the generation of new data specifically generated to answer a hypothetical question is still so commonplace that careful thinking about the actual need to generate new data may just not be on the radar screen. Obviously, data creation will need to continue, but increasingly we have to ask the question whether such new data are absolutely necessary to answer the question we want to answer. With more and more data becoming available in reusable format, there may well be existing data collections 'Other People's' Data and associated Services (OPEDAS) that without or with some extra effort needed, can answer at least part of the question or least may be crucial for the interpretation of your own data.

### Do

- Search for data sets (OPEDAS) that may be re-usable and can help you to reduce the number of new data sets you may have to generate (and steward later on).
- Include annotated collections of data and curated databases in your search.
- Check the accessibility and license situation attached to the relevant data sets you found.
- Check their interoperability. They may be relevant but not interoperable with your analysis pipelines. In that case you may have to extract, transform and load (ETL) them or decide that -although relevant- they are not reusable for your purpose.
- Ensure that using OPEDAS will not restrict in any way the use of your results later on, including copyright and freedom to operate on the request of IPR.



# Questionnaire TODOs

- When you are unsure how to answer a question, you can add TODO and come back later
- List of TODOs

## 1 What data formats/types will you be using? TODO x

Have you identified types of data that you will use that are used by others too? Some types of data (e.g. genetic variants in the life sciences) are used by many different projects. For such data, often common standards exist that help to make these data reusable. Are you using such common data formats?

 Data Stewardship for Open Science: *njy*

 Add

# Feedback

- When a researcher find something unclear or wrong with a question
- Integrated to GitHub issues

### Feedback

If you found something wrong with the question, you can send us your feedback how to improve it.

**Title**

**Description**

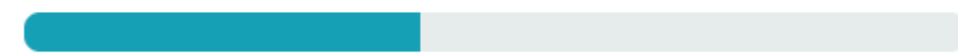


# Summary Report

- Information about FAIR metrics for each chapter
- Calculated as a weighted average of answers

## Information and insight

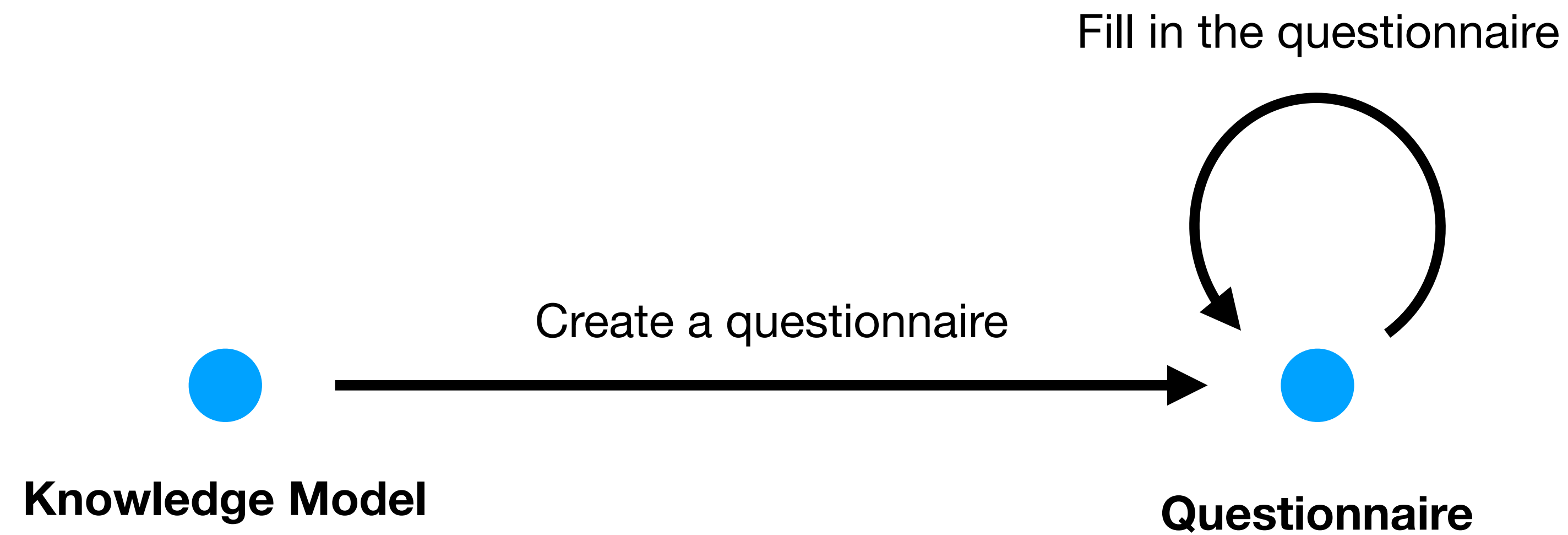
Answered: 13/31



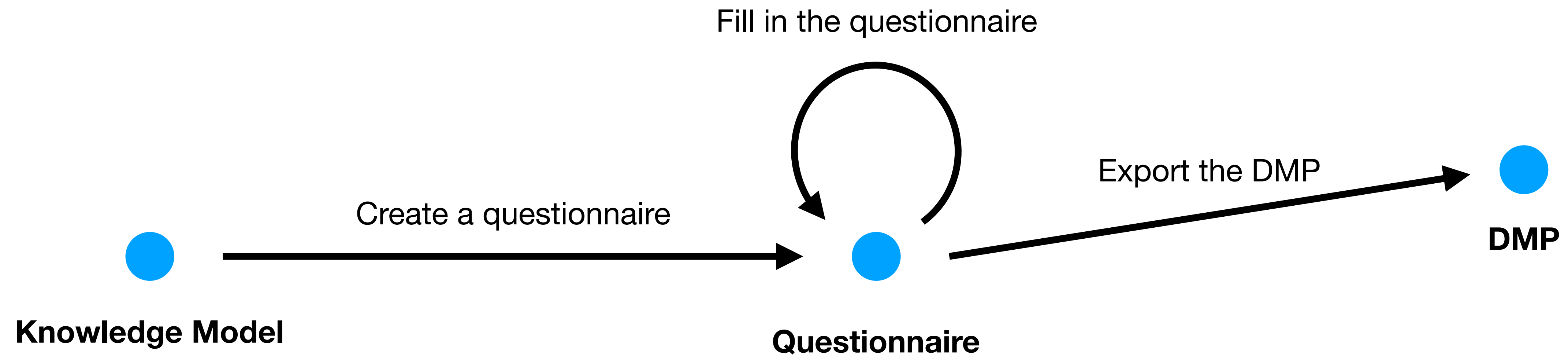
Metric	Measure	
Findability	1.00	<div></div>
Accessibility	0.67	<div></div>
Reusability	0.30	<div></div>



# Researchers Workflow



# Researchers Workflow



# Exporting the DMP

- Documents can be generated from a filled questionnaire
- Customizable document templates  
DSW provides **Science Europe template**
- Various document formats

### Export My Experiment

**Template**

Science Europe DCC Template

**Format**

☒ PDF Document

☐ LaTeX Document

☐ MS Word Document

☐ HTML Document

☐ JSON Data

☐ OpenDocument Text

☐ Markdown Document

Cancel

Download

# Exporting the DMP

## Data Management Plan

### HVSC: Hypothetical Vascular Study by a Chemist

Contact person: **Bob A** ([bob.a@example.com](mailto:bob.a@example.com),  9999-5559-9999-9999)  
Example Corporation

Based on: Life Sciences DSW Knowledge Model, 2.0.0 (*dsw:lifesciences:2.0.0*)

Generated on: 24. 02. 2020

Data Management Plan created in Data Stewardship Wizard <<https://ds-wizard.org>>

1

## Abstract

We will be deriving models of the "virtual" age of arterial walls by combining new data with existing biobanks and reference data, in order to create early warning systems for aging patients.

2

## Section A: Data Collection

### 1. What data will you collect or create?

#### Re-used datasets

We will use the following reference datasets:

- [Human Protein Atlas](#)

We will use version "19.1" of this dataset. If a new version becomes available during the project, new analyses will be done with the new version.

- [UniProt Knowledgebase](https://www.uniprot.org/uniprot/) (<https://www.uniprot.org/uniprot/>)

We will use version "2019\_09" of this dataset.

### 2. How will the data be collected or created?

There will be no instrument dataset in this project.

## Section B: Documentation and Meta-data

### 3. What documentation and meta-data will accompany the data?

List of data to be published is given in Section E, Question 9. This also includes information about catalogs where the data can be found. Information about data types used is given in Section A, Question 1.

## Section C: Ethics and Legal Compliance

### 4. How will you manage any ethical issues?

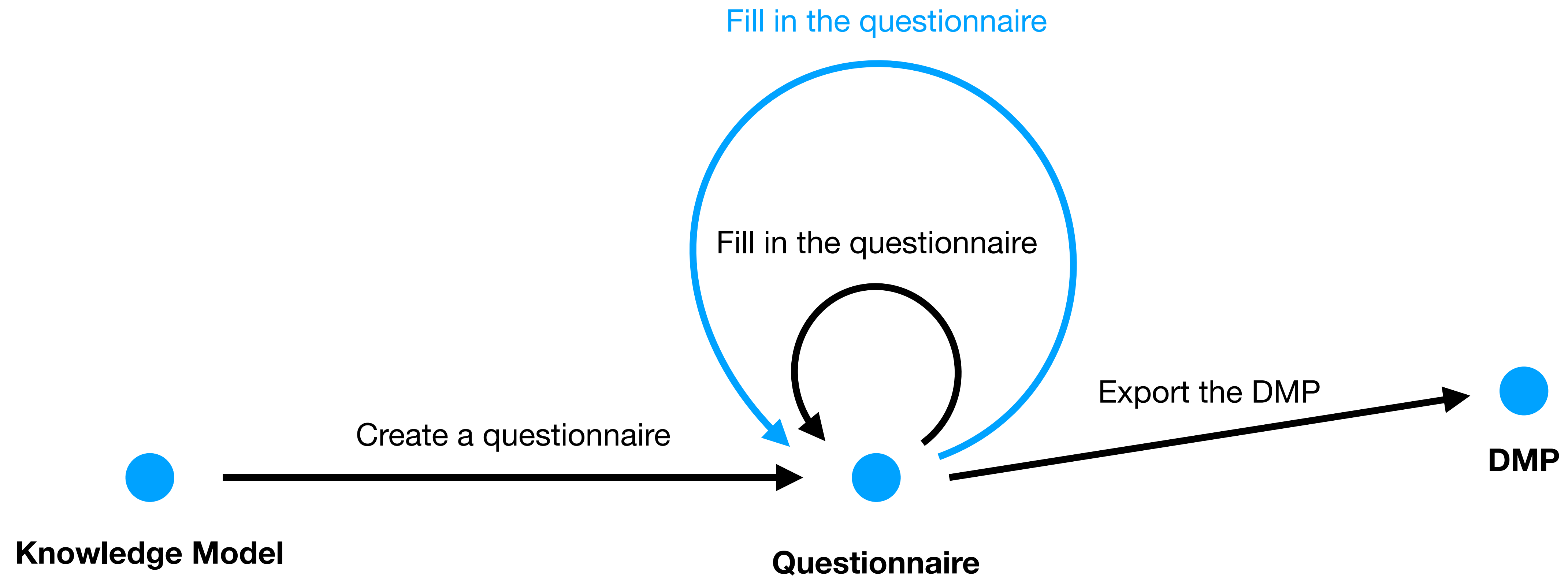
### 5. How will you manage copyright and Intellectual Property Rights (IPR) issues?

For the reference and non-reference data sets that we reuse, conditions are as follows:

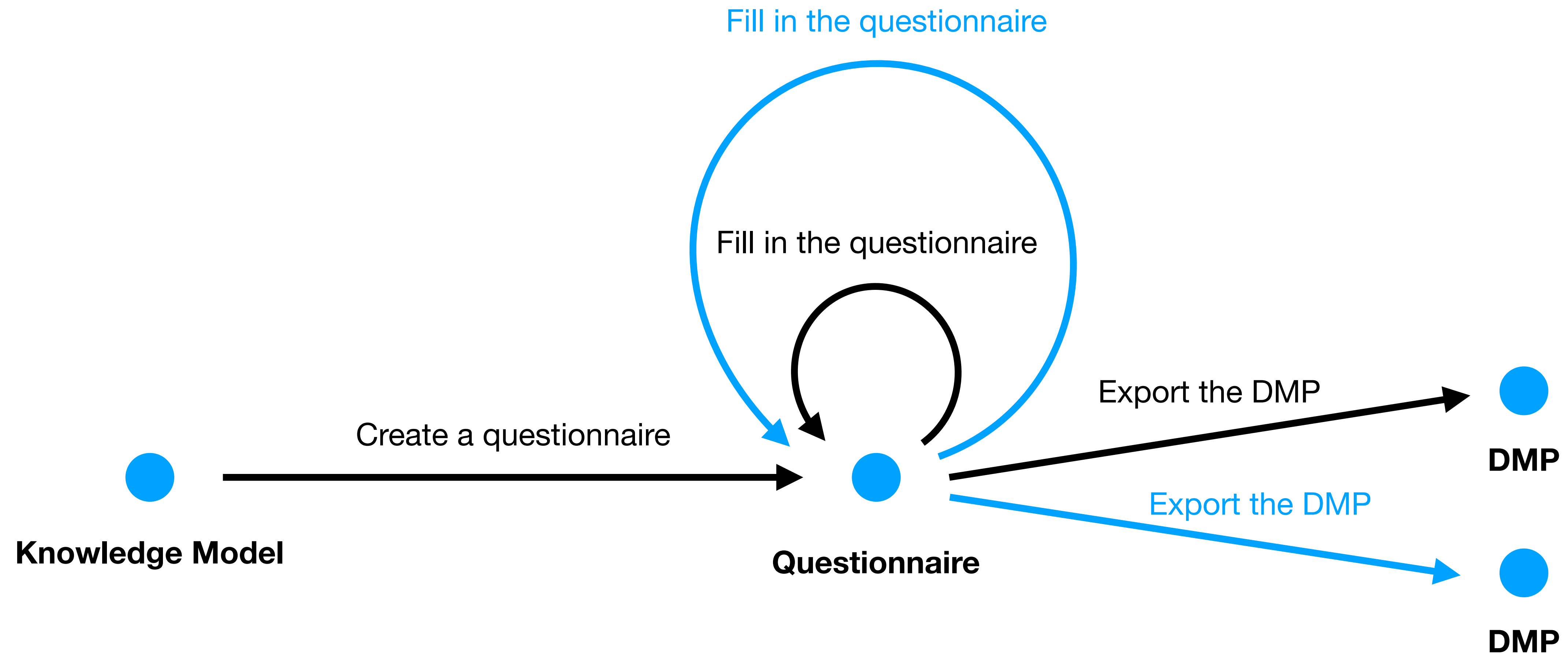
- [Human Protein Atlas](#) – available under specific restrictions, which we will

3

# Researchers Workflow



# Researchers Workflow



# Additional Research Workflows

- Questionnaire cloning
- Questionnaire migration





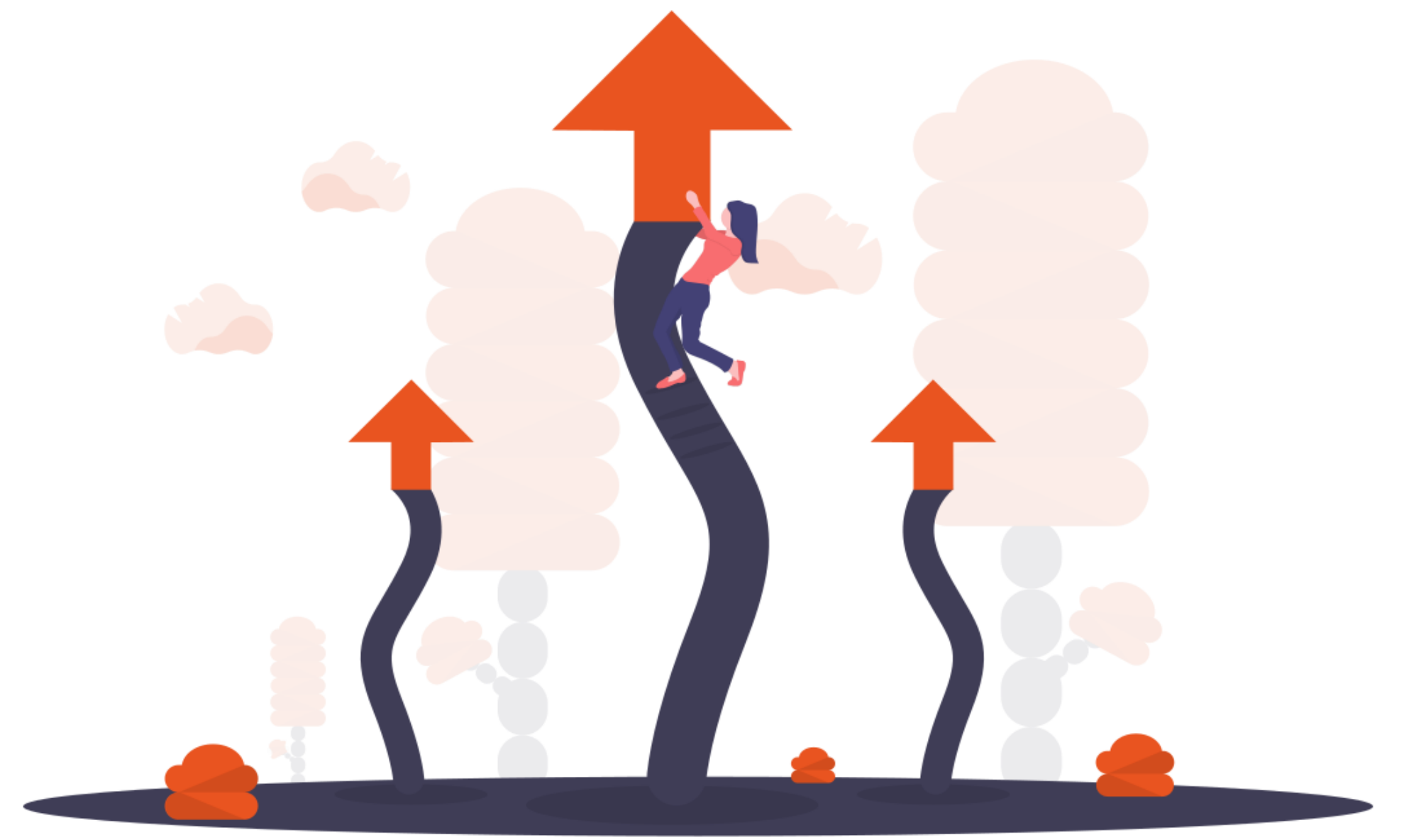
# Questionnaire Cloning

- Create a **copy of existing questionnaire**
- Can be used to create partially filled questionnaire templates
  - Public Read-Only questionnaire as a template
  - Researcher can clone it
  - And fill in the specific details

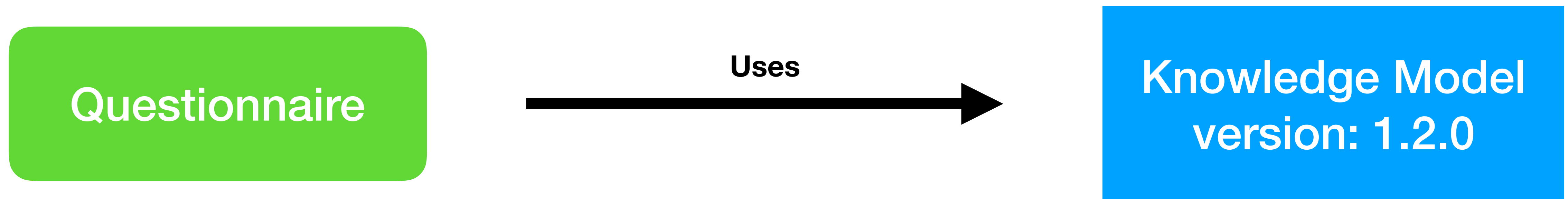


# Questionnaire Migration

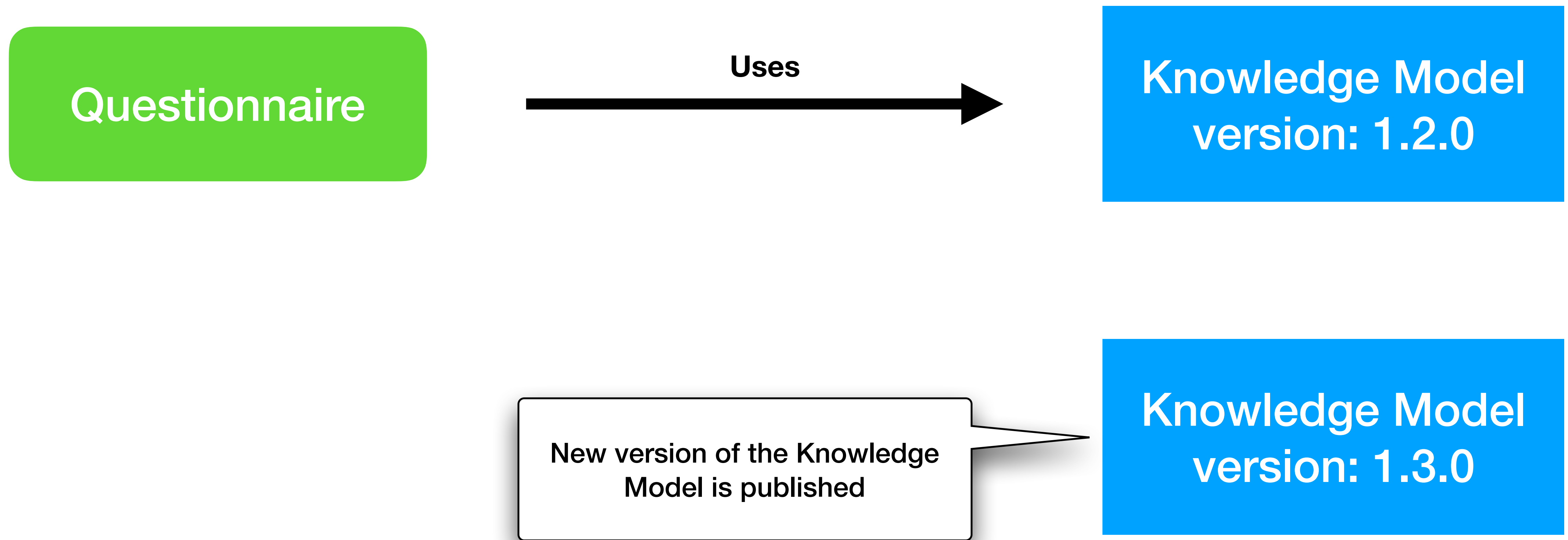
- Process of upgrading existing Questionnaires with their answers to new versions of Knowledge Models



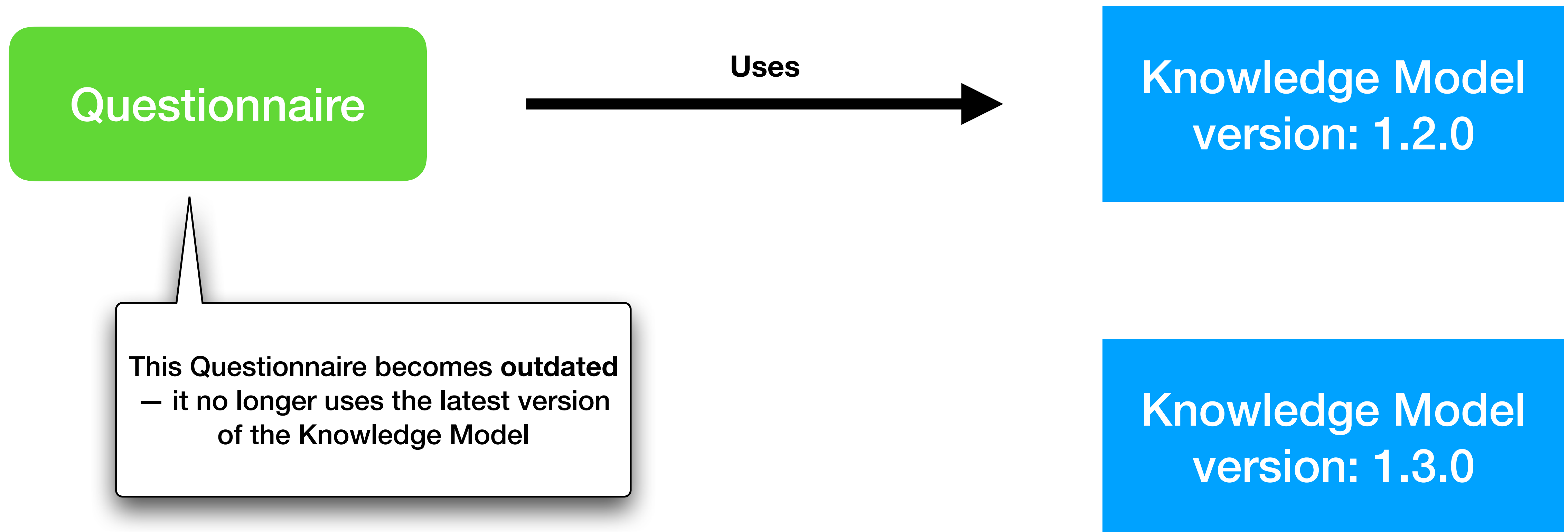
# Questionnaire Migration



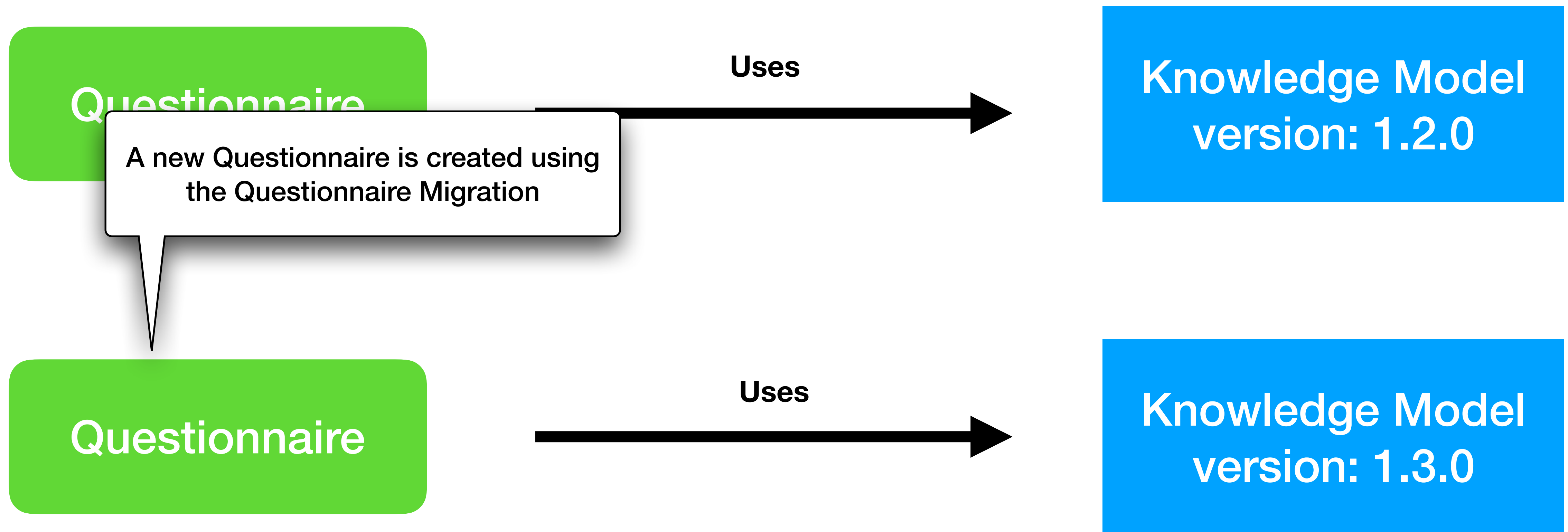
# Questionnaire Migration



# Questionnaire Migration



# Questionnaire Migration



# Researchers Workflow Demo





# DSW for Data Stewards

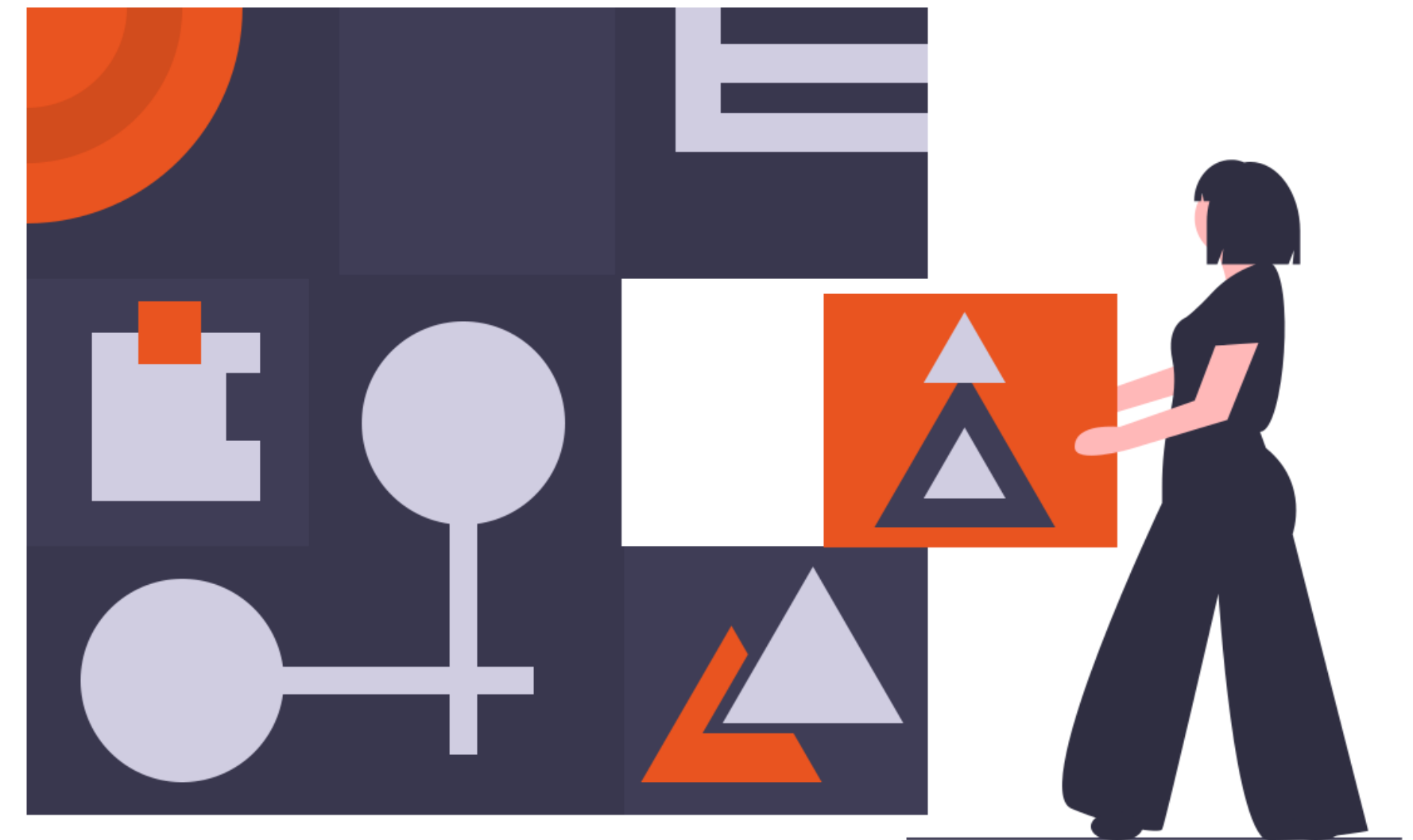


# Data Steward Responsibilities in DSW

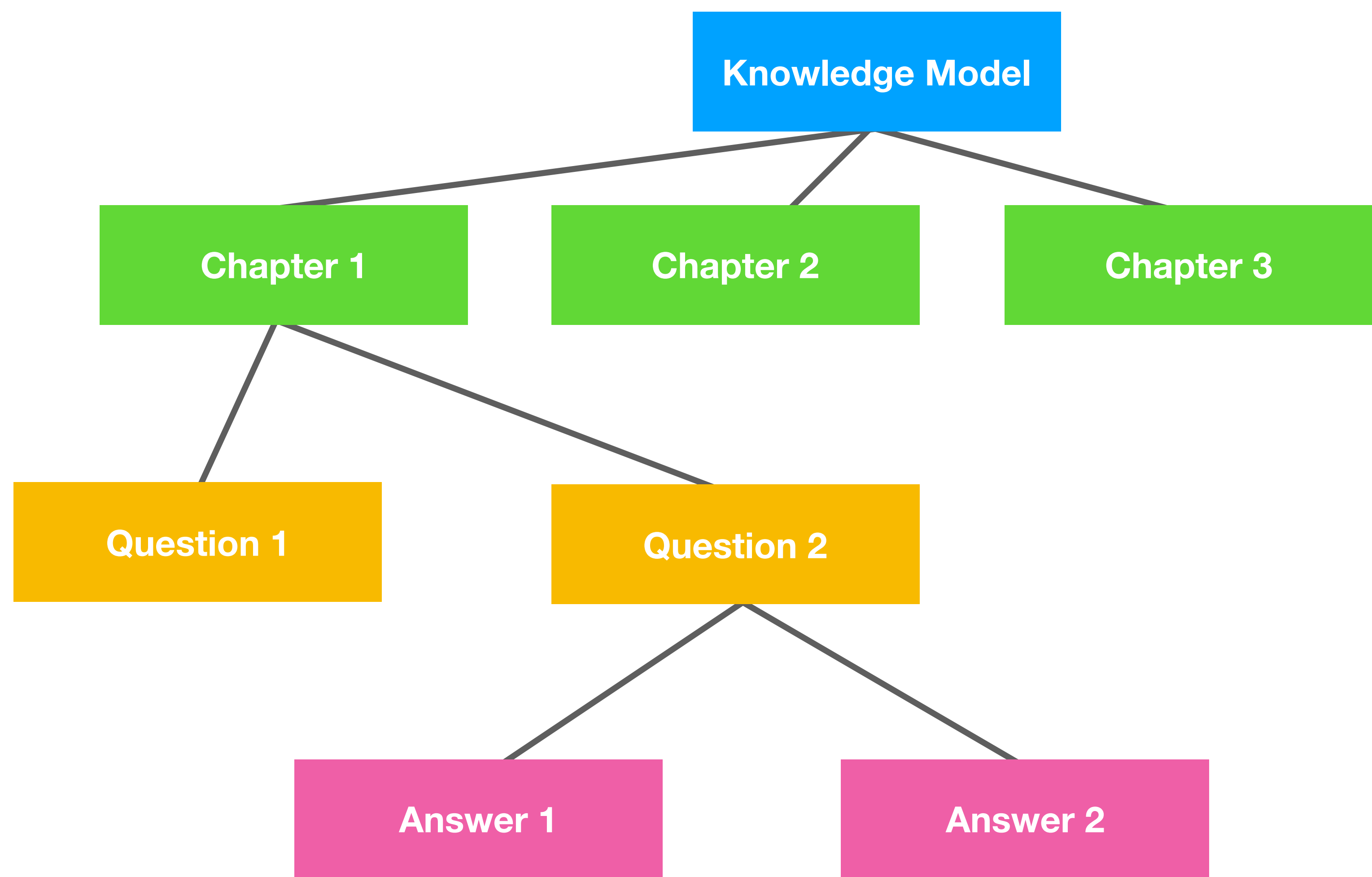
- Manage existing Knowledge Models
- Build & publish new Knowledge Models
- Import Knowledge Models from other DSW instances or Registry

# Building a Knowledge Model

- From scratch
- Extending existing Knowledge Model



# Knowledge Model



- ▼ Core DS Knowledge Model
  - ▼ Design of experiment
    - ▼ Is there any pre-existing data?
      - ☒ No
      - ▶ ☒ Yes
        - atq
    - ▶ Will reference data be created?
    - ▶ Will you be storing samples?
    - ▶ Will you be collecting experimental data?
  - ▼ Data design and planning
    - ▼ What data formats/types will you be using?
      - Data format/type:
        - ▶ Is this a standard data format used by others too?
      - ▼ Does this data format enable sharing and long term archiving?
        - ▶ ☒ No
        - ☒ Yes
    - njy
    - ▶ Will you be using new types of data?
    - ▶ How will you be storing metadata?
    - ▶ During the project, will you be archiving data (using so-called 'cold storage')?
    - ▶ Will you need a shared working space to work with your data?
    - ▶ Is the risk of information loss, leaks and vandalism acceptably low?
    - ▶ Do you need to do compute capacity planning?
  - ▶ Data Capture/Measurement
  - ▶ Data processing and curation

# Chapter

- At the top level of each Knowledge Model
- Contains **short introduction** and **questions**

## Chapters

I. Design of experiment	4
II. Data design and planning	7
III. Data Capture/Measurement	3
IV. Data processing and curation	4
V. Data integration	7
VI. Data interpretation	3
VII. Information and insight	11

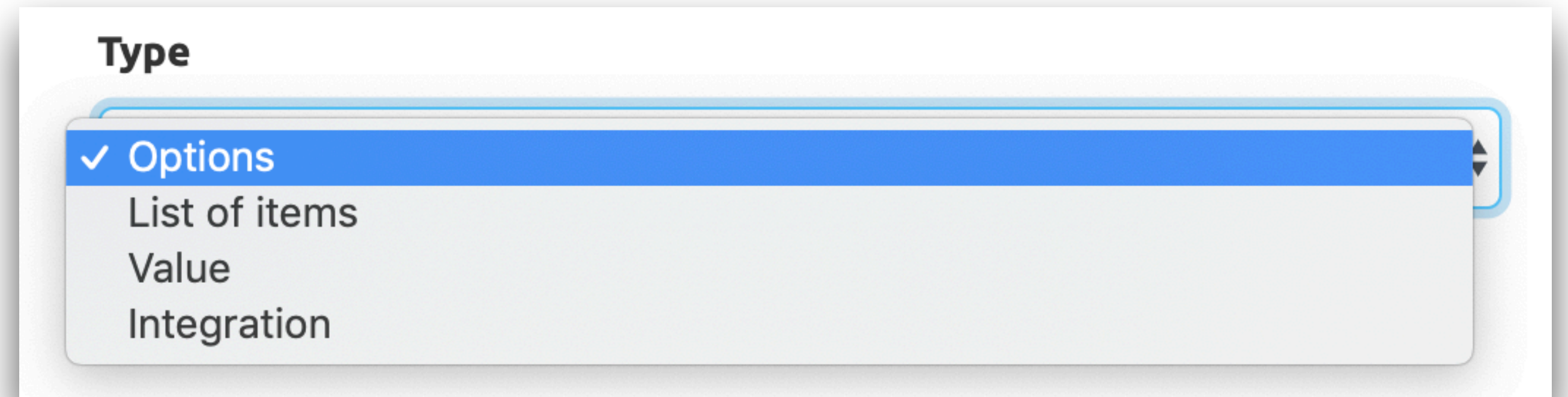


# Question

- Core entity of the Knowledge Model
- Each question has a **Title** and a description **Text**
- Can be of different types
- Can have **References** and **Experts** assigned

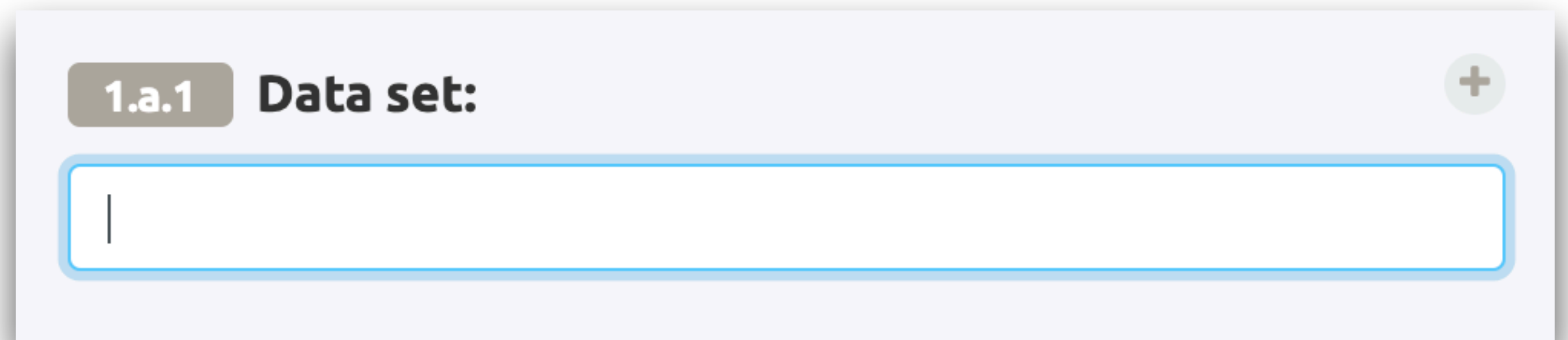
# Question Type

- Value
- Integration
- Options
- List of items



# Question Type: Value

- Simple value that users type
- Value types
  - Number
  - Date
  - String
  - Text



1.a.1 Data set:

# Question Type: Integration

- Similar to Value type
- The answers can be taken **from external resource**
- More complex to set up

1 What database will you use?

database

CHD7 Database  
Telomerase Database  
MetaBase - The wiki-database of biological databases  
Gene Disruption Project Database  
Ebola and Hemorrhagic Fever Virus Database  
International Ocean Discovery Program Database




# Question Type: Integration

- Similar to Value type
- The answers can be taken **from external resource**
- More complex to set up

## 1.a.1 Data format/type:

RDF/XML Syntax Specification

 <https://fairsharing.org/bsg-s001261>

# Question Type: Options


- Closed list of **answers**
- Users can pick one
- Answers can have **follow-up questions**

1


Is there any pre-existing data?


+

Are there any data sets available in the world that are relevant to your planned research?

 Data Stewardship for Open Science: [atq](#)

☐ a. No

☒ b. Yes 

 Clear answer

# Question Type: List of Items

- Multiple **items** of the same type
- Each item has the same set of questions

1 Please specify what data sets you will acquire using measurement equipment

You can use any name for the data set, make sure that it identifies the data set to yourself.

+ Add



# Question Type: List of Items

- Multiple **items** of the same type
- Each item has the same set of questions


## 1 Please specify what data sets you will acquire using measurement equipment

You can use any name for the data set, make sure that it identifies the data set to yourself.

### 1.a.1 Data set:

### 1.a.2 Who will do the measurements? And where?

Are there easily accessible specialized service providers for data capture?

- ☐ a. Experts in the project, with our own equipment
- ☐ b. Experts in the project, at a specialized infrastructure
- ☐ c. External party 

# Answer

- Used in **Options Question**
- Can have **advice**, **follow-up questions** and **FAIR metrics** assessment

# Answer Advice


- When an answer is selected, advice can be provided
- Guide users to achieve better results

1


Is there any pre-existing data?


+

Are there any data sets available in the world that are relevant to your planned research?

 Data Stewardship for Open Science: [atq](#)

☒ a. No

☐ b. Yes 

 Clear answer

You know that this is very unlikely? This question is not only about data sets that are similar to what you want to determine yourself, but also reference data or data that should be mined from the existing literature. Further, it is very likely that you will refer to related data, e.g. other databases where you usually "quickly look something up", but that could maybe be properly integrated, especially if you need to do such lookups multiple times.

# Follow-up Questions


- When an answer is selected, more questions can be unfolded
- Used for questions that are relevant only until certain conditions

1


Is there any pre-existing data?


+

Are there any data sets available in the world that are relevant to your planned research?

 Data Stewardship for Open Science: [atq](#)

☐ a. No

☒ b. Yes 


 Clear answer

1.b.1


Will you be using any pre-existing data (including other people's data)?

+

Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?

 Data Stewardship for Open Science: [ezi](#)

☐ a. No

☐ b. Yes 

1.b.2

Do you need to harmonize different sources of existing data?

+

If you are combining data from different sources, harmonization may be required. You

# FAIR Metrics

- Each answer can affect different **FAIR metrics**
- Result is calculated as a weighted average of selected answers

2

Will you be using common or exchangeable units?

+

☐ a. No

Interoperability

☒ b. Yes

Interoperability

↶

Clear answer



# FAIR Metrics

- Each answer can affect different **FAIR metrics**
- Result is calculated as a weighted average of selected answers

**Metrics**  
☐ Findability  
☐ Accessibility  
☒ Interoperability  
☐ Reusability

Weight

Measure

# FAIR Metrics

- **Weight** — how important the answer is

Not important 0  1 Very Important

- **Measure** — how does it affect the result

Bad 0  1 Good

# Reference


- Can be used to provide links to **external resources**
- User can open links when filling the Questionnaire

**1 Is there any pre-existing data?**

Are there any data sets available in the world that are relevant to your planned research?

☒ External Links: [Example Reference](#)

☐ a. No

☐ b. Yes 

# Expert


- Can be used to provide contacts to experts
- Users can see the contact by the question

## 1 Is there any pre-existing data?

Are there any data sets available in the world that are relevant to your planned research?

 Experts: *John Brown* ([john.brown@example.com](mailto:john.brown@example.com))

☐ a. No

☐ b. Yes 

# Using Markdown

- Markdown is a lightweight syntax for styling texts
- You can use basic features
  - bold or italic text
  - links
  - bullet or numbered lists
- More examples:  
<https://guides.github.com/features/mastering-markdown/>

Editor

Preview

It's very easy to make some words **bold** and other words *italic* with Markdown. You can even [link to Google!](http://google.com)

Sometimes you want numbered lists:

1. One
2. Two
3. Three

Sometimes you want bullet points:

- \* Start a line with a star
- \* Profit!

*You can use Markdown and see the result in the Preview tab.*

# Using Markdown

- Markdown is a lightweight syntax for styling texts
- You can use basic features
  - bold or italic text
  - links
  - bullet or numbered lists
- More examples:  
<https://guides.github.com/features/mastering-markdown/>

Editor

Preview

It's very easy to make some words **bold** and other words *italic* with Markdown. You can even [link to Google!](#)

Sometimes you want numbered lists:

1. One
2. Two
3. Three

Sometimes you want bullet points:

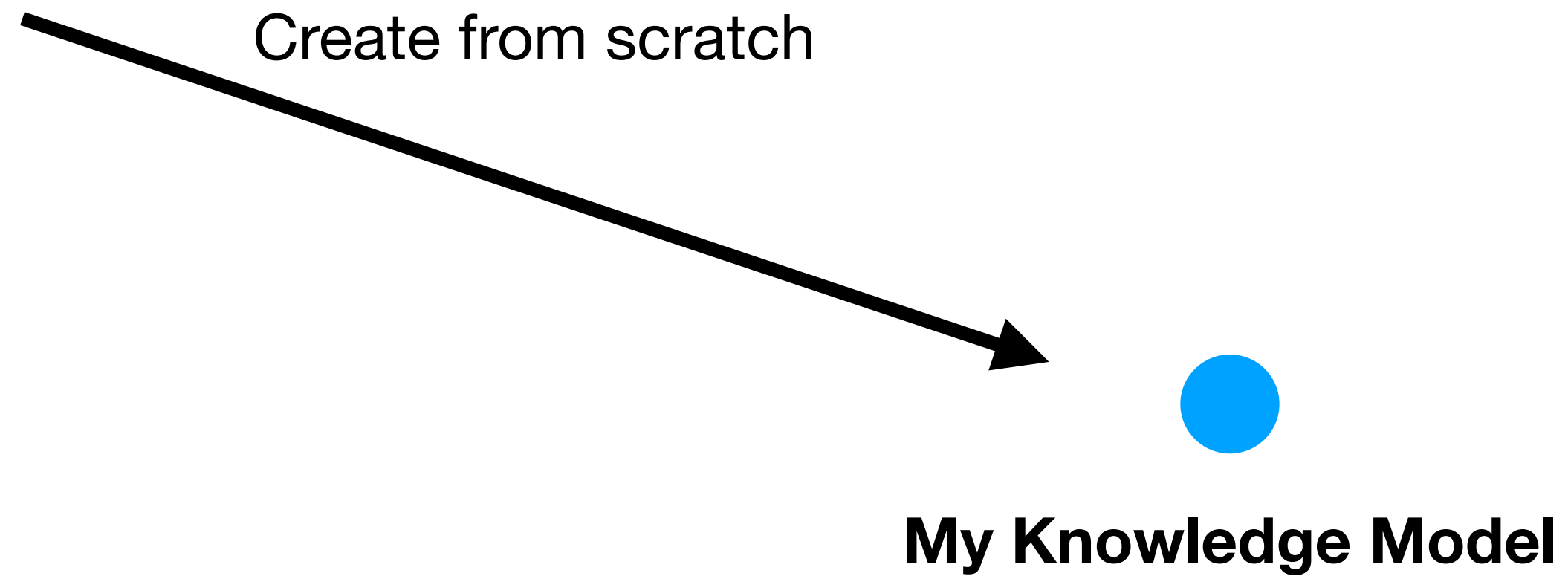
- Start a line with a star
- Profit!

*You can use Markdown and see the result in the Preview tab.*

# Publishing the Knowledge Model

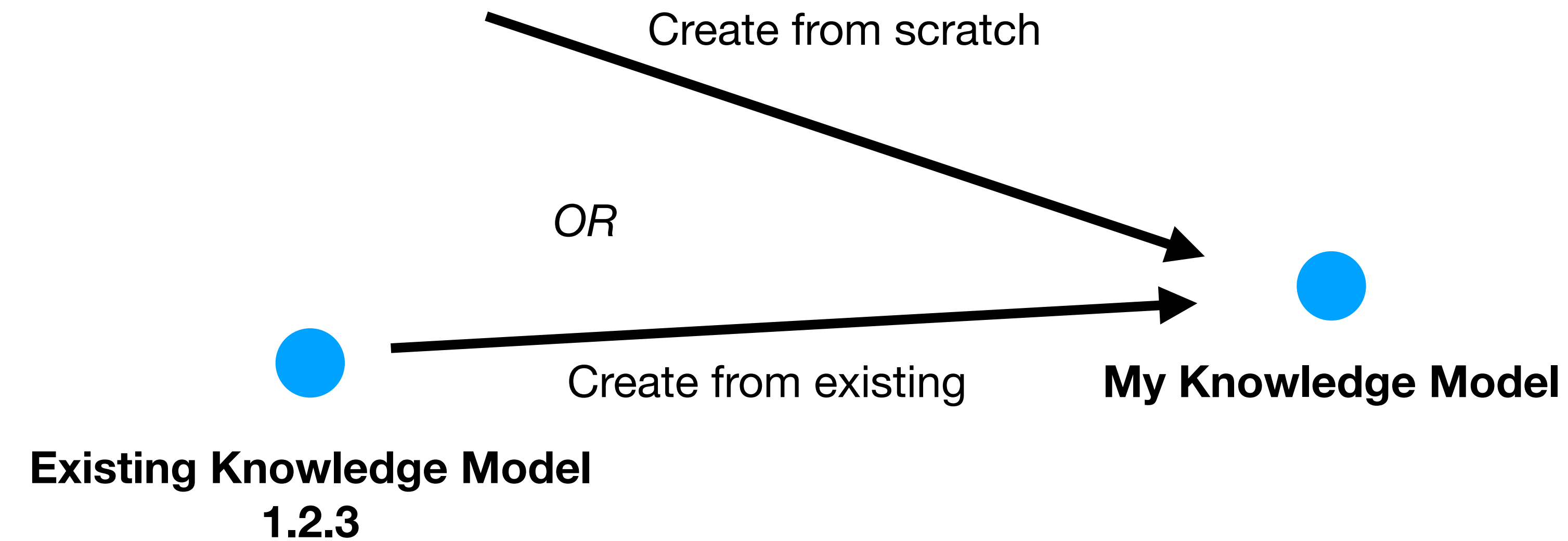
- Before a Knowledge Model can be used, it has to be **published**
- **Version, description** and **readme** is assigned
- Once the version is published **it cannot be changed**
- New changes as a new version
- Published Knowledge Models can be used to create a Questionnaire

# Knowledge Model Workflow

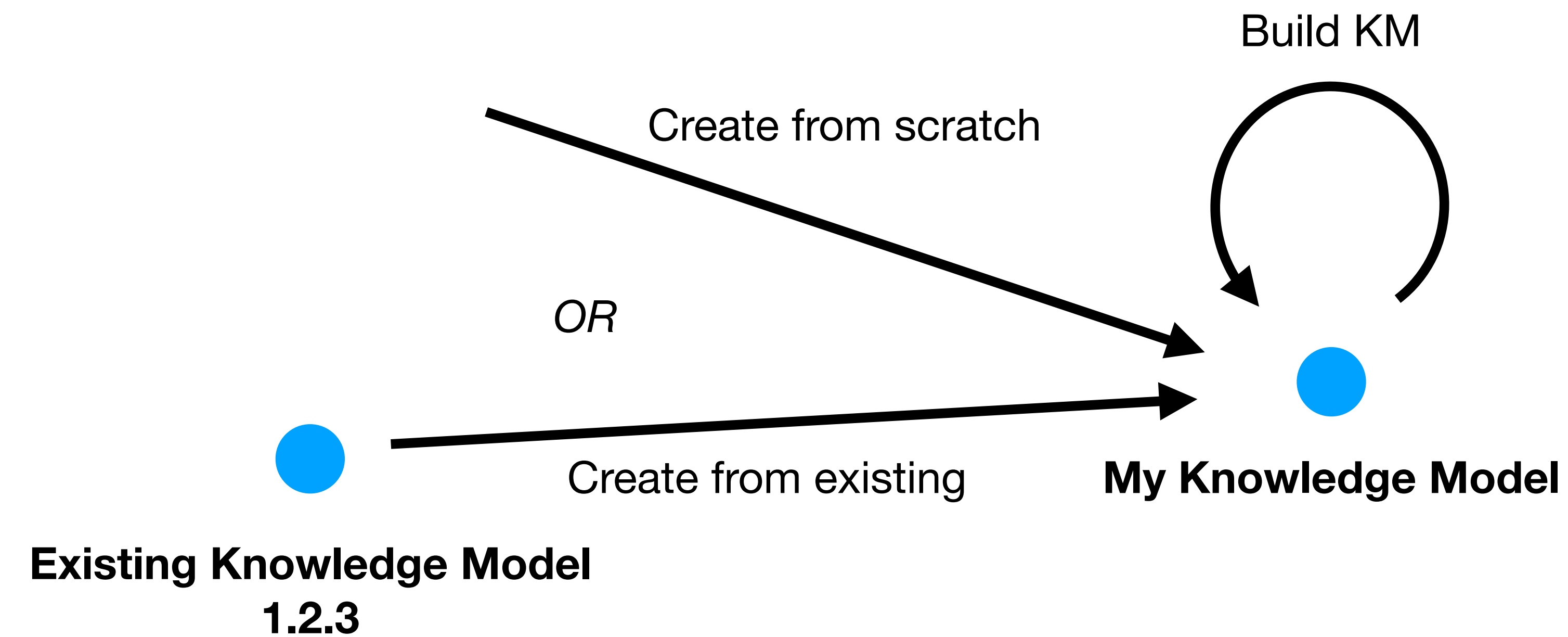




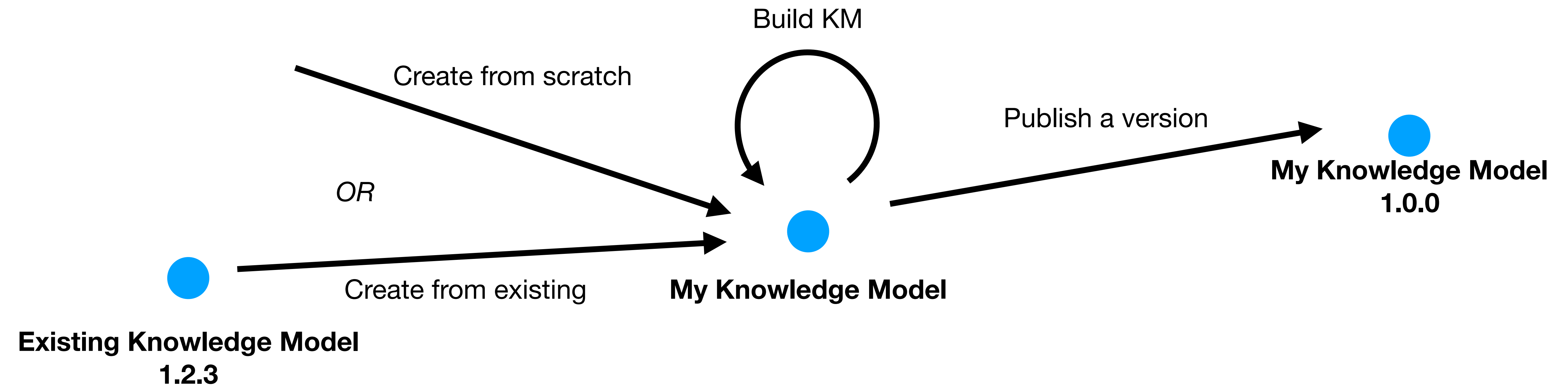
# Knowledge Model Workflow



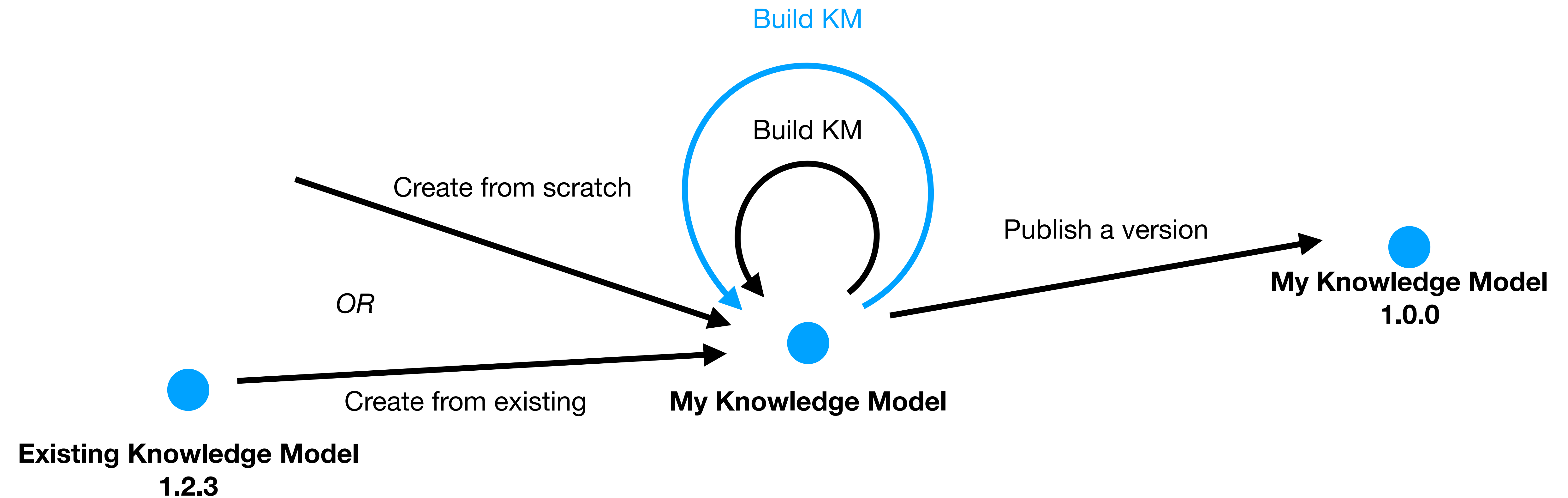
# Knowledge Model Workflow



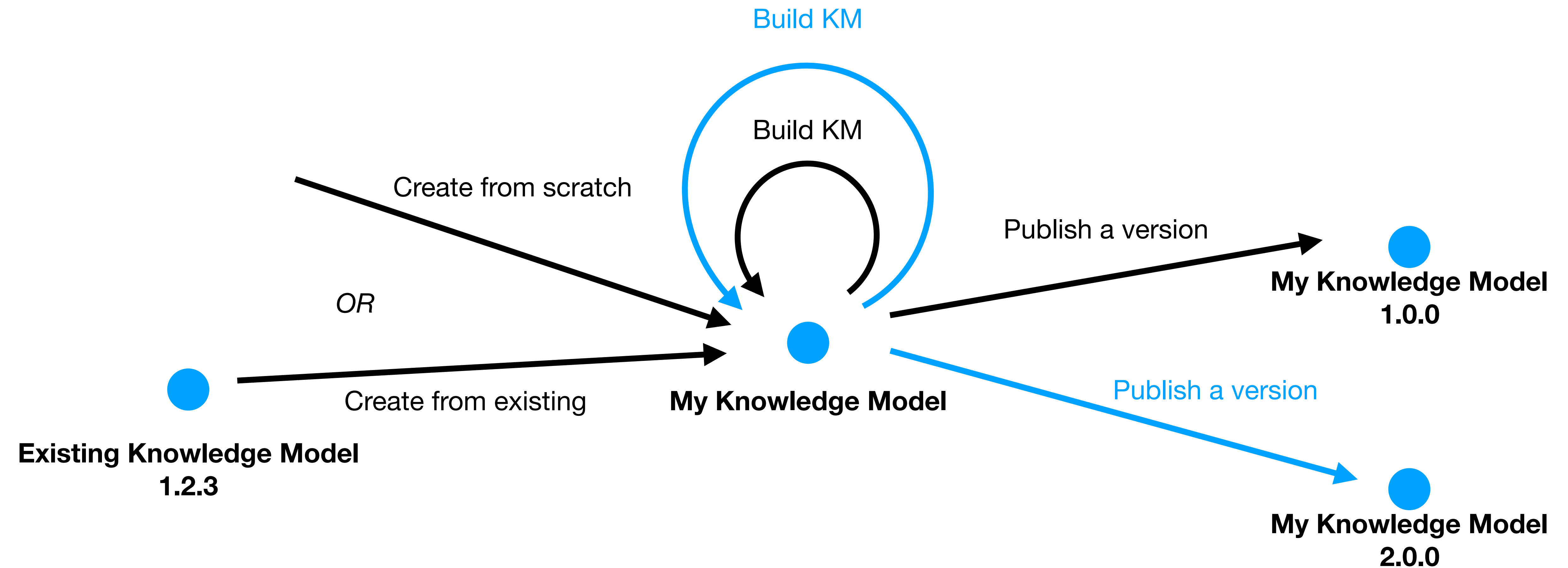
# Knowledge Model Workflow



# Knowledge Model Workflow



# Knowledge Model Workflow



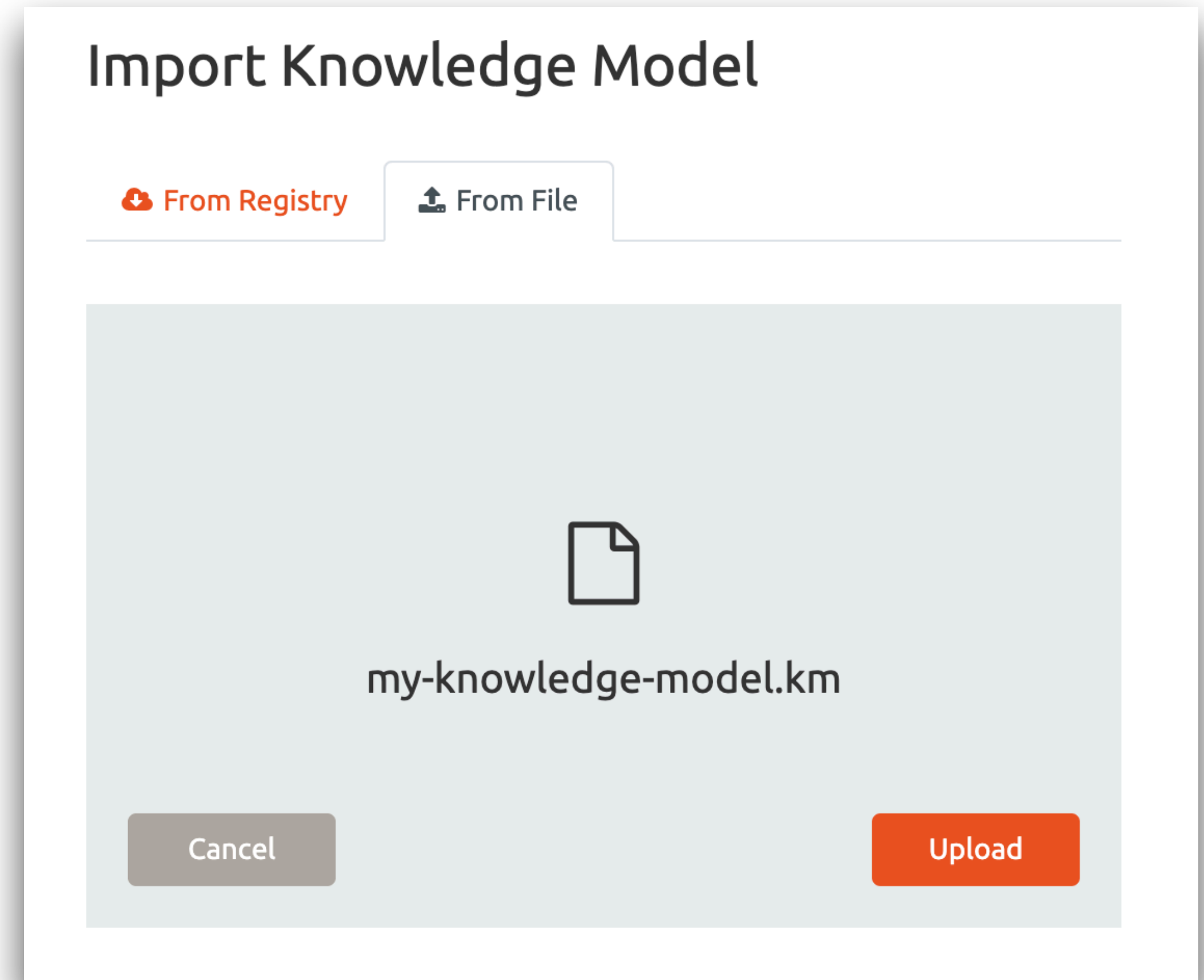
# Knowledge Model Import & Export

- From file
- From Registry



# Import & Export with Files


- When a Knowledge Model was published it can be **exported**
- A *.km* file is downloaded from the DSW Instance
- It can be **imported** to a different DSW Instance




# Import from Registry

- DSW Knowledge Models are publicly available in [registry.ds-wizard.org](https://registry.ds-wizard.org)
- Easy import into DSW instances using **Knowledge Model ID**

## Import Knowledge Model

 From Registry

 From File

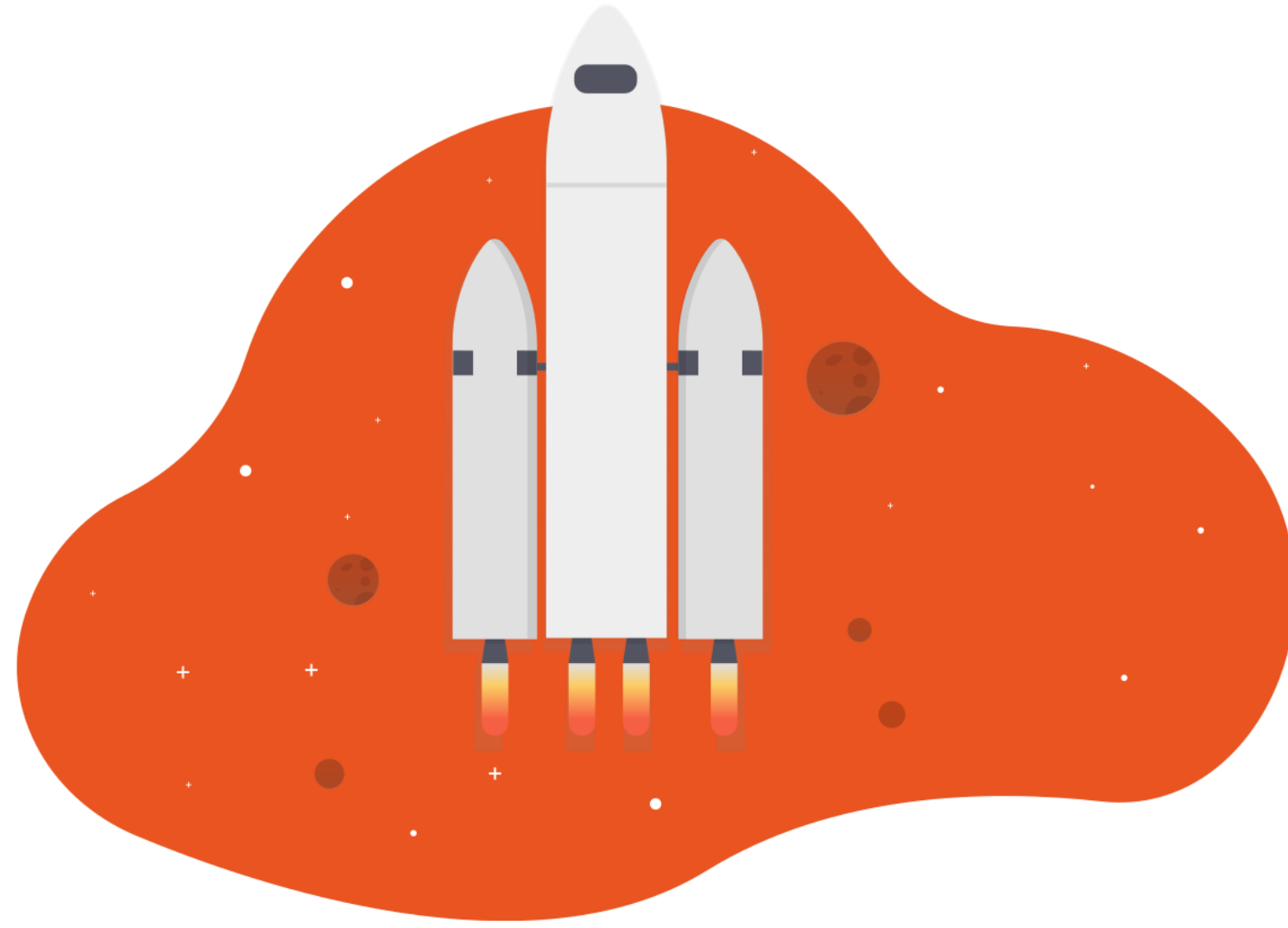
Import

You can find knowledge models in the [registry](#).



# Data Stewards Demo

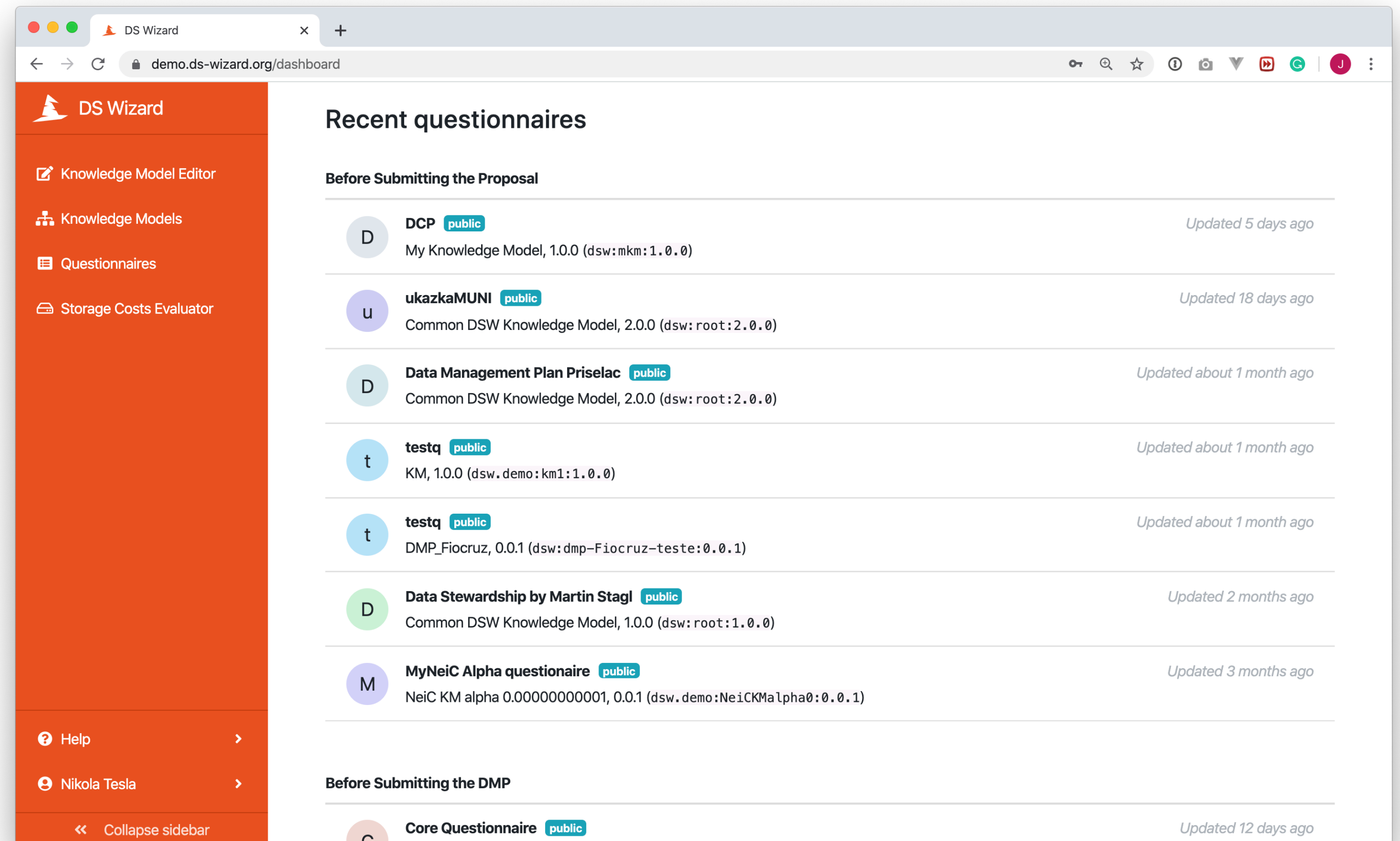




# How to Get Started

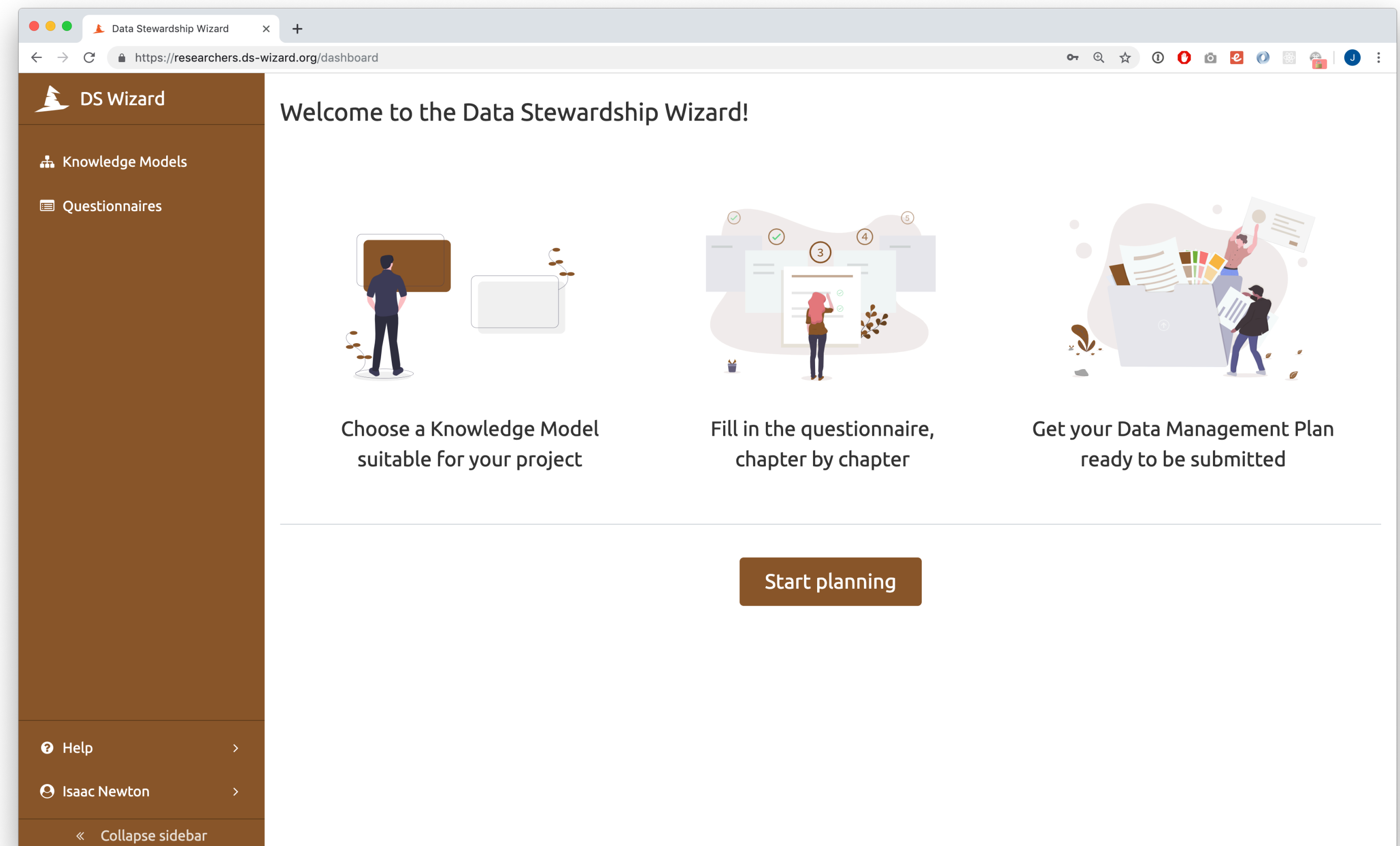
# Demo

- [demo.ds-wizard.org](https://demo.ds-wizard.org)
- Playground
- Share instance with other users

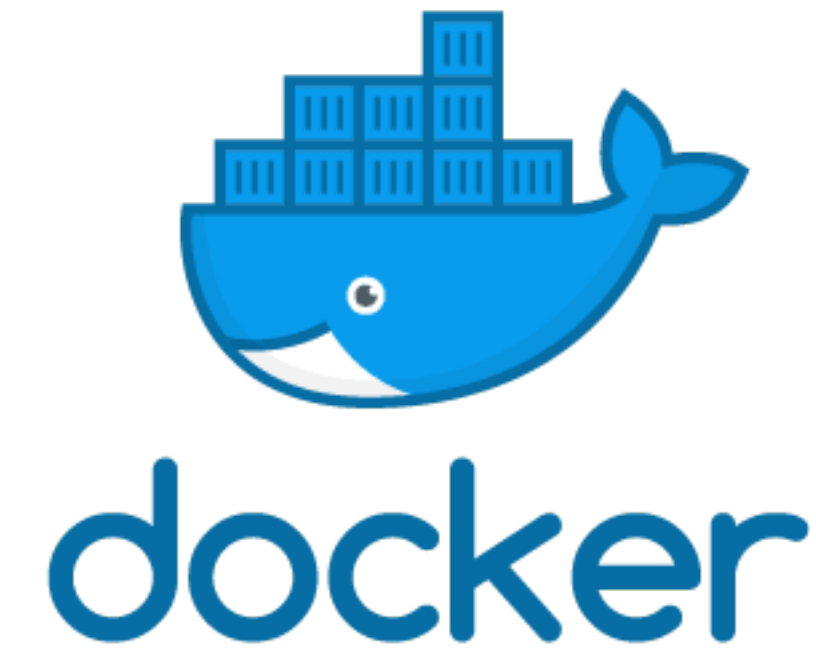


# Researchers

- [researchers.ds-wizard.org](https://researchers.ds-wizard.org)
- For individual researchers
- Easy to sign up and use
- Ready to use Knowledge Model



# Self-Managed



- DS Wizard is distributed via Docker
- Documentation: <https://docs.ds-wizard.org/>
- Docker registry: <https://hub.docker.com/u/datastewardshipwizard>

# DSW Cloud

- We offer managing the DS Wizard instance for interesting projects that want to use it seriously but don't want to run it by themselves



# Overview

	Demo	Researchers	Self-Managed	DSW Cloud
Production Ready		✓	✓	✓
Questionnaires & DMPs	✓	✓	✓	✓
Knowledge Models Management	✓		✓	✓
Custom Export Document Templates			✓	✓
User Management & Organization Settings			✓	✓
Features configuration & Visual style customizations			✓	✓





# Questions & Discussion



# Try DS Wizard yourself!



- [ds-wizard.org](https://ds-wizard.org)
- [info@ds-wizard.org](mailto:info@ds-wizard.org)
- [https://twitter.com/dswizard\\_org](https://twitter.com/dswizard_org)

