

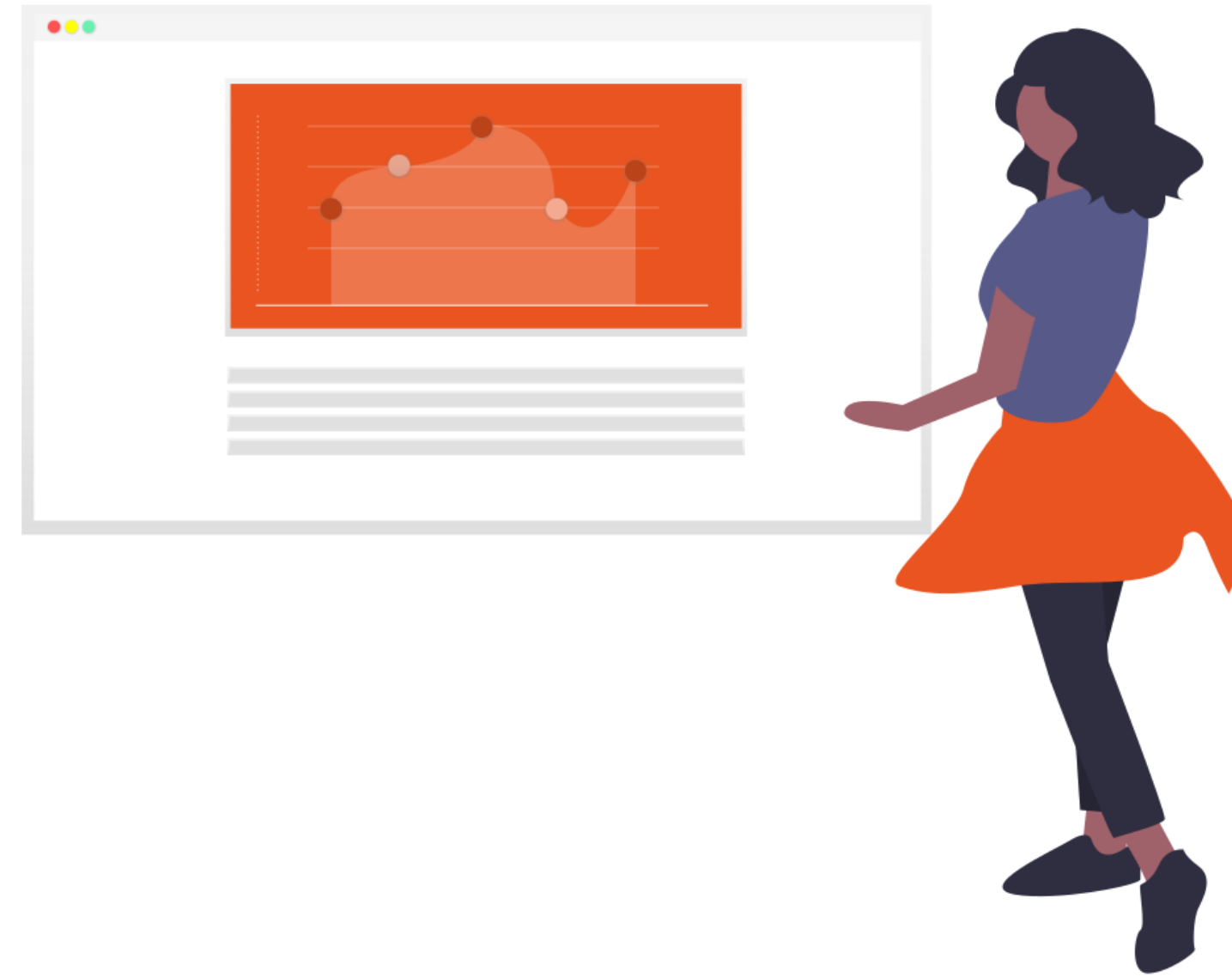


A Gentle Introduction to DSW for Convergencers

Outline

- **DSW Introduction**
 - DSW for Researchers
 - DSW for Data Stewards
- **Demo**
 - Setting up an integration question
 - Project and document templates
 - Configuring a submission service
- **Questions & Discussion**





Introduction

Data Stewardship Wizard

- A tool for generating data management plans in various formats
- Using smart questionnaires to guide researchers
- Customizable questionnaire templates and document export formats
- Online collaboration





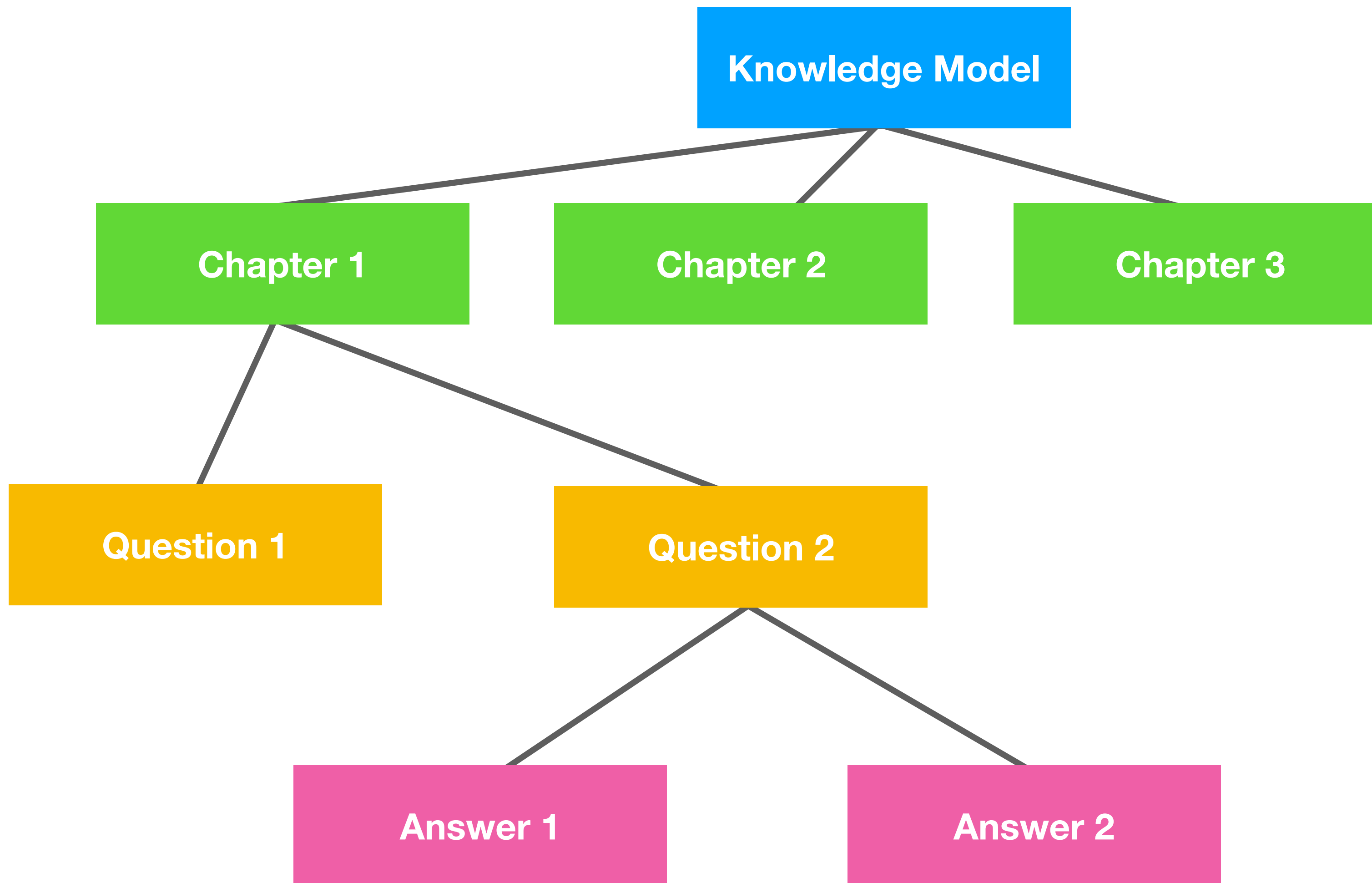
DSW for Researchers

Researchers Workflow



Knowledge Model

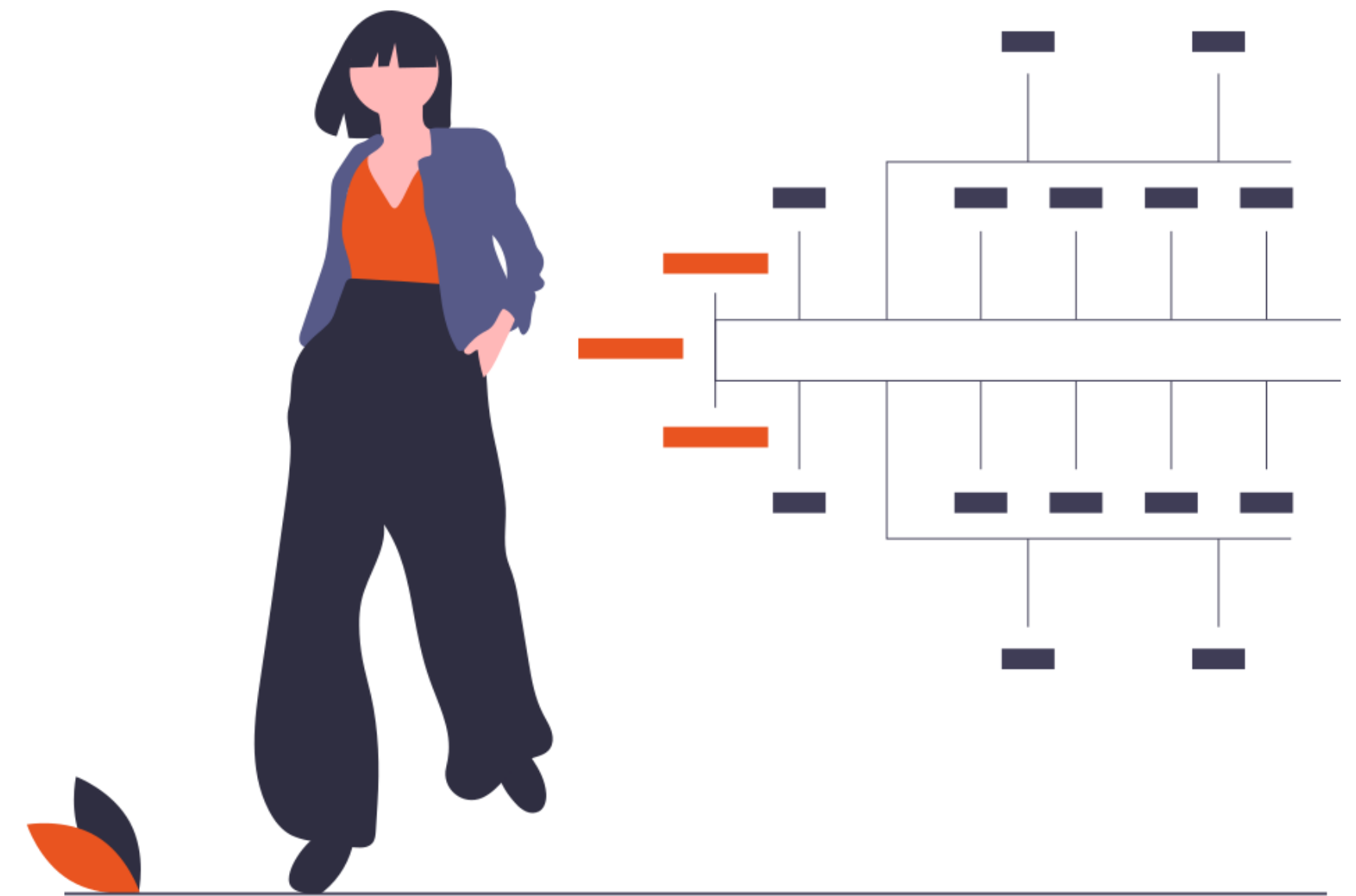
Knowledge Model



- ▼ 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

Knowledge Model

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

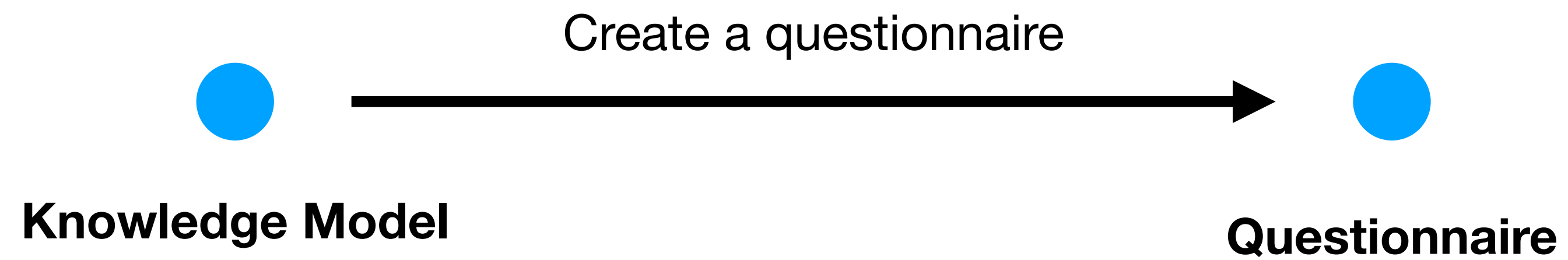


Researchers Workflow



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**
- Can be exported to a **human-readable** (e.g., PDF) or **machine-actionable** (e.g., RDF) document

The screenshot displays a questionnaire interface. On the left, a 'Chapters' table of contents lists seven sections: I. Design of experiment (1), 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 three questions, each with a 'Data Stewardship for Open Science' reference and radio button options for 'a. No' and 'b. Yes'. Question 2 asks 'Will reference data be created?' with reference 'rbz' and 'a. No' selected. Question 3 asks 'Will you be storing samples?' with reference 'kuz' and 'b. Yes' selected. Question 4 asks 'Will you be collecting experimental data?' with reference 'csx' and 'b. Yes' selected. Each question also includes a 'Clear answer' button.

Chapters	
I. Design of experiment	1
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](#)

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

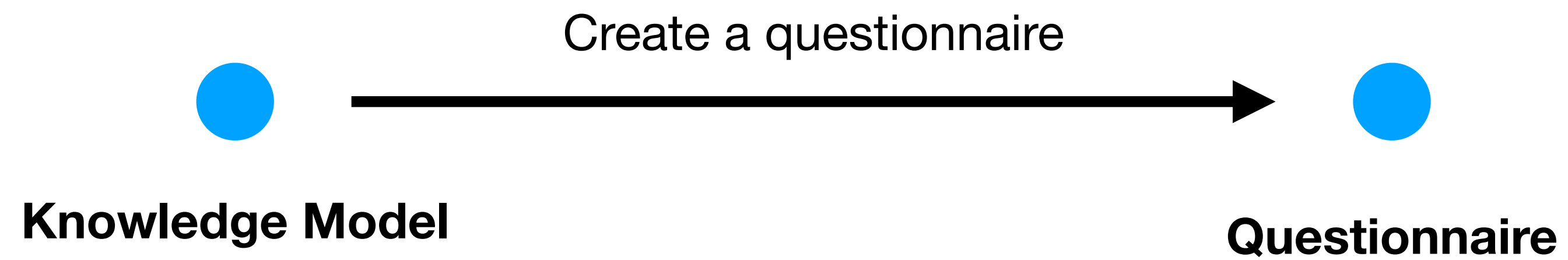
4 Will you be collecting experimental data? +

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

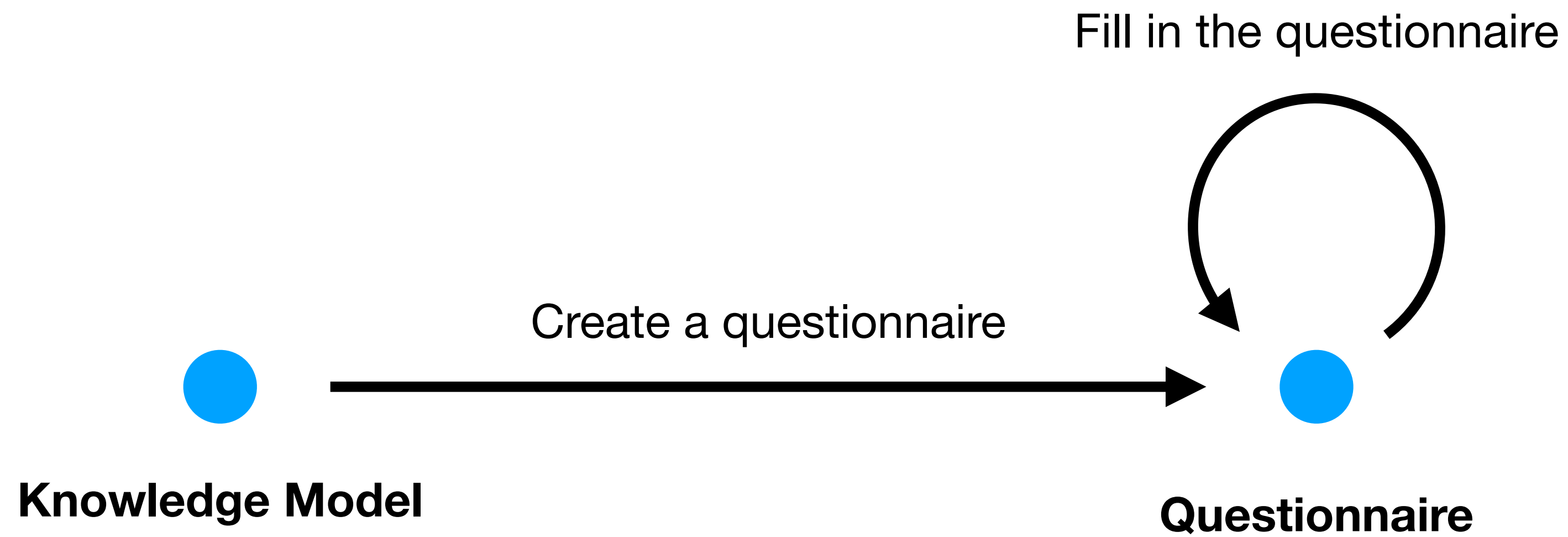
a. No

b. Yes

Researchers Workflow



Researchers Workflow



Filling the Questionnaire

- Free navigation through the Questionnaire (no given order of answering)
- Invite other people to collaborate online

The screenshot displays a web-based questionnaire interface. On the left, a sidebar titled 'Chapters' 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 two questions from the 'Data Stewardship for Open Science' survey. The first question, '1.b.2 Do you need to harmonize different sources of existing data?', is highlighted in green and includes a text box with explanatory text. The second question, '1.b.2.b.1 Will you be making your harmonization results available to others?', is highlighted in grey and also includes explanatory text. Both questions feature radio button options for 'a. No' and 'b. Yes', with 'b. Yes' selected in both. A 'Clear answer' link is provided for each question.

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*

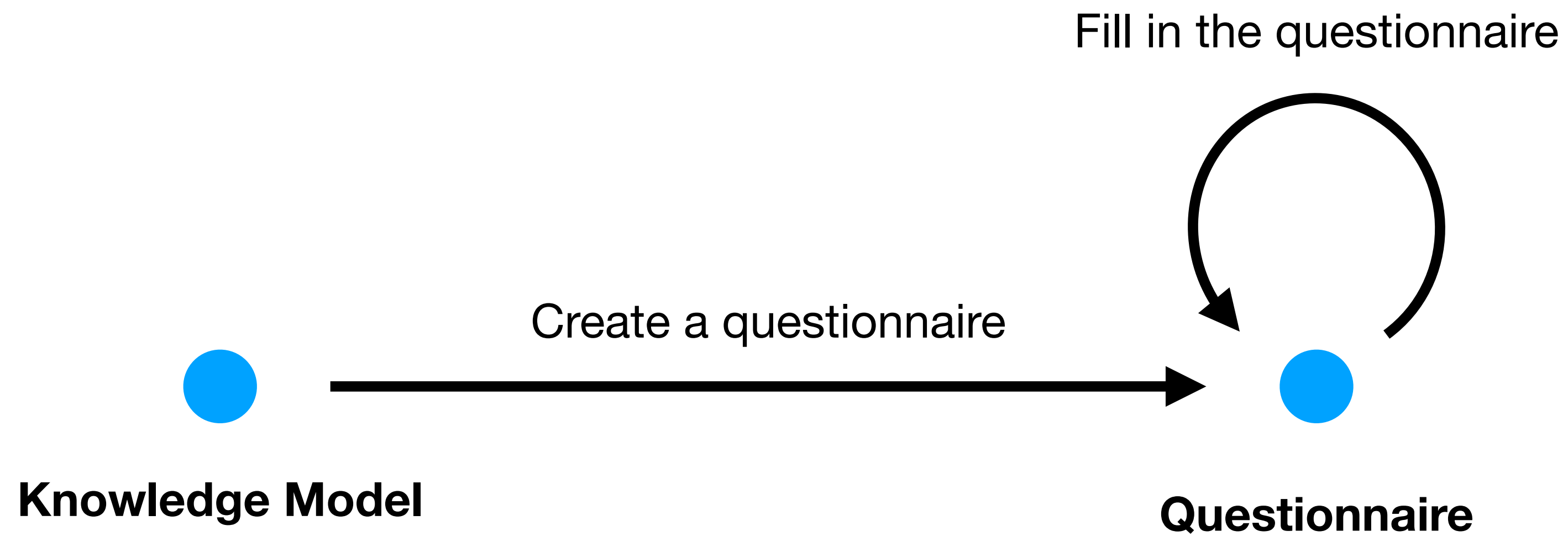
a. No

Accessibility

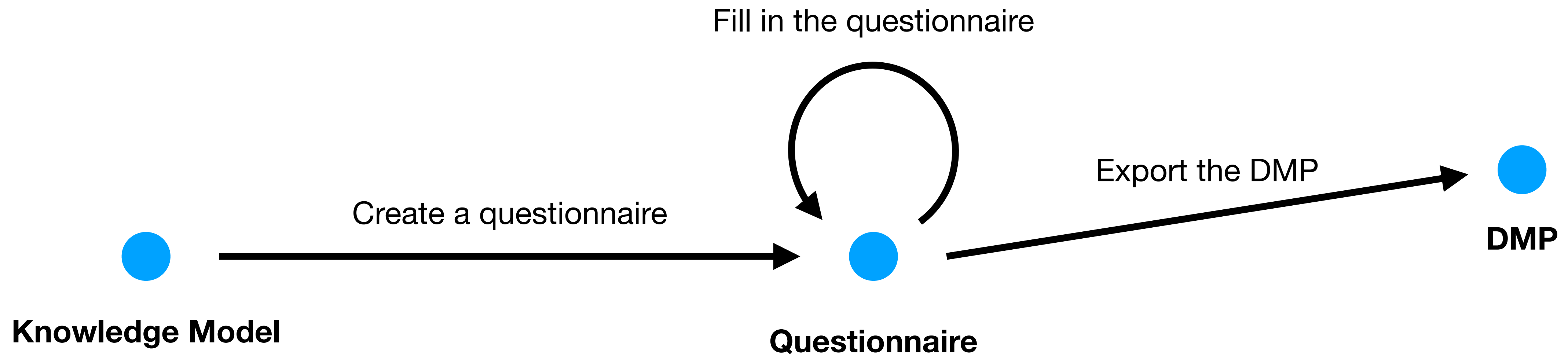
b. Yes

Accessibility

Researchers Workflow



Researchers Workflow



Exporting the DMP

- Documents can be generated from a filled questionnaire
- Customizable document templates
- Various document formats
- DSW provided templates
 - **Science Europe**
 - **Horizon 2020**
 - **maDMP**

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,  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?

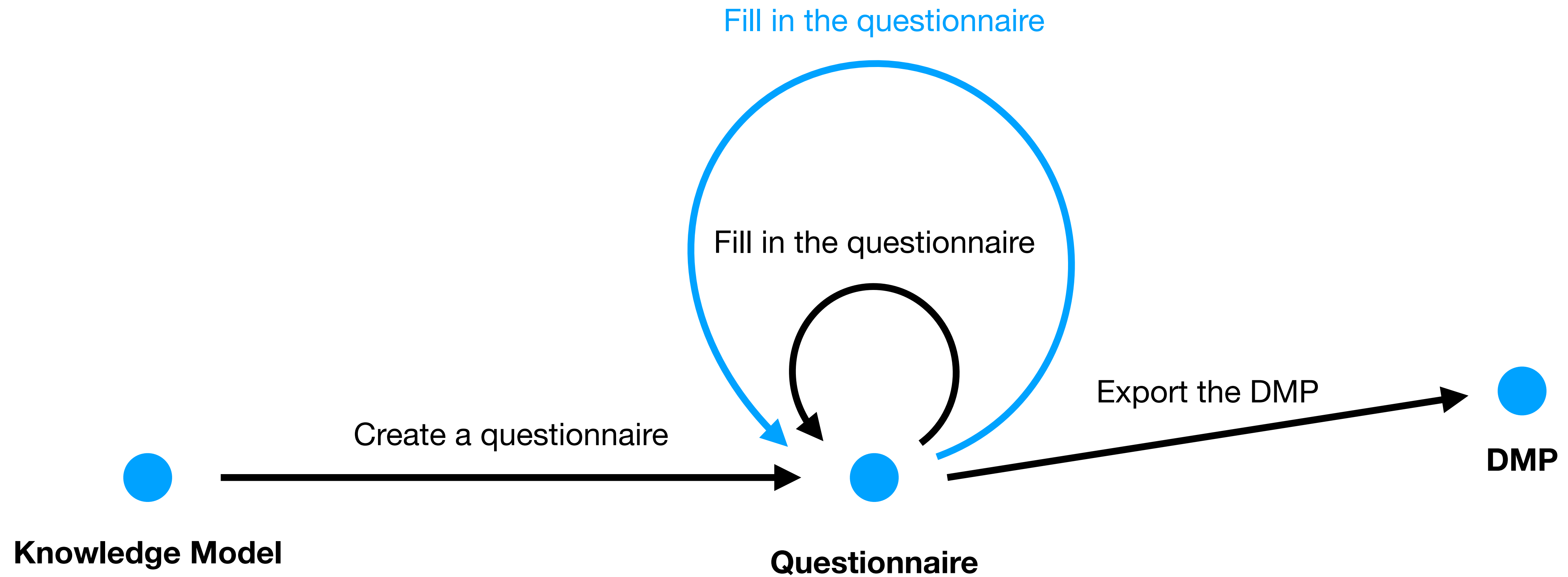
5. How will you manage copyright and Intellectual Property Rights (IPR) issues?

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

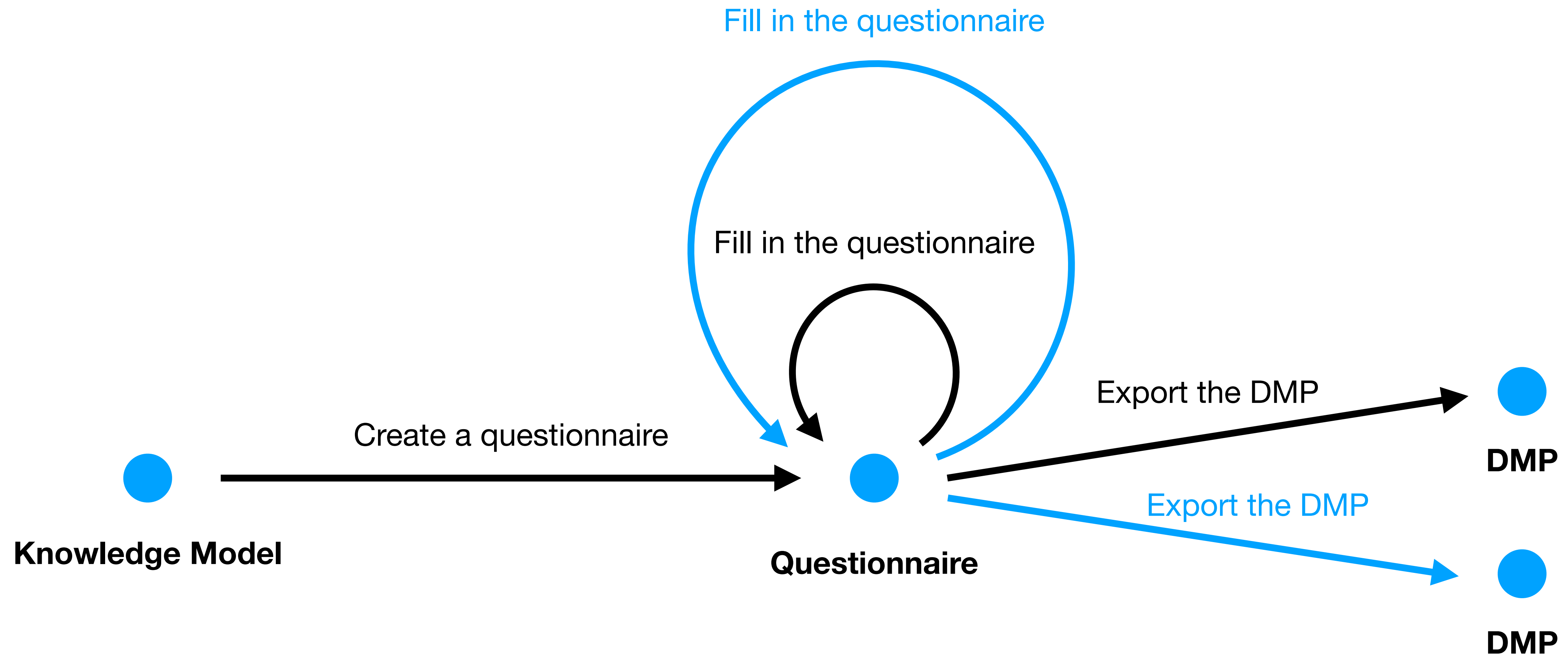
- [Human Protein Atlas](#) – available under specific restrictions, which we will

3

Researchers Workflow



Researchers Workflow





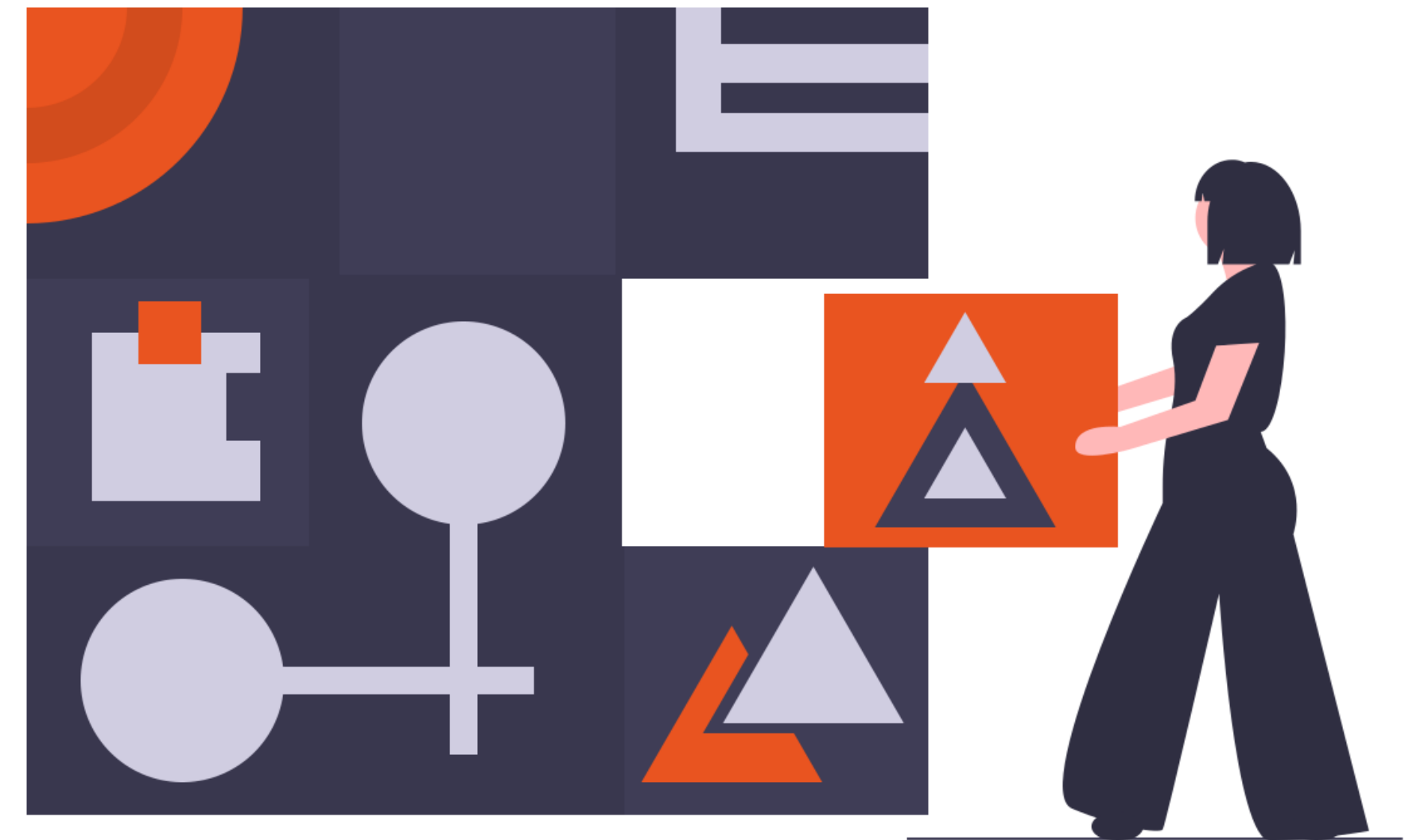
DSW for Data Stewards

Data Steward Responsibilities in DSW

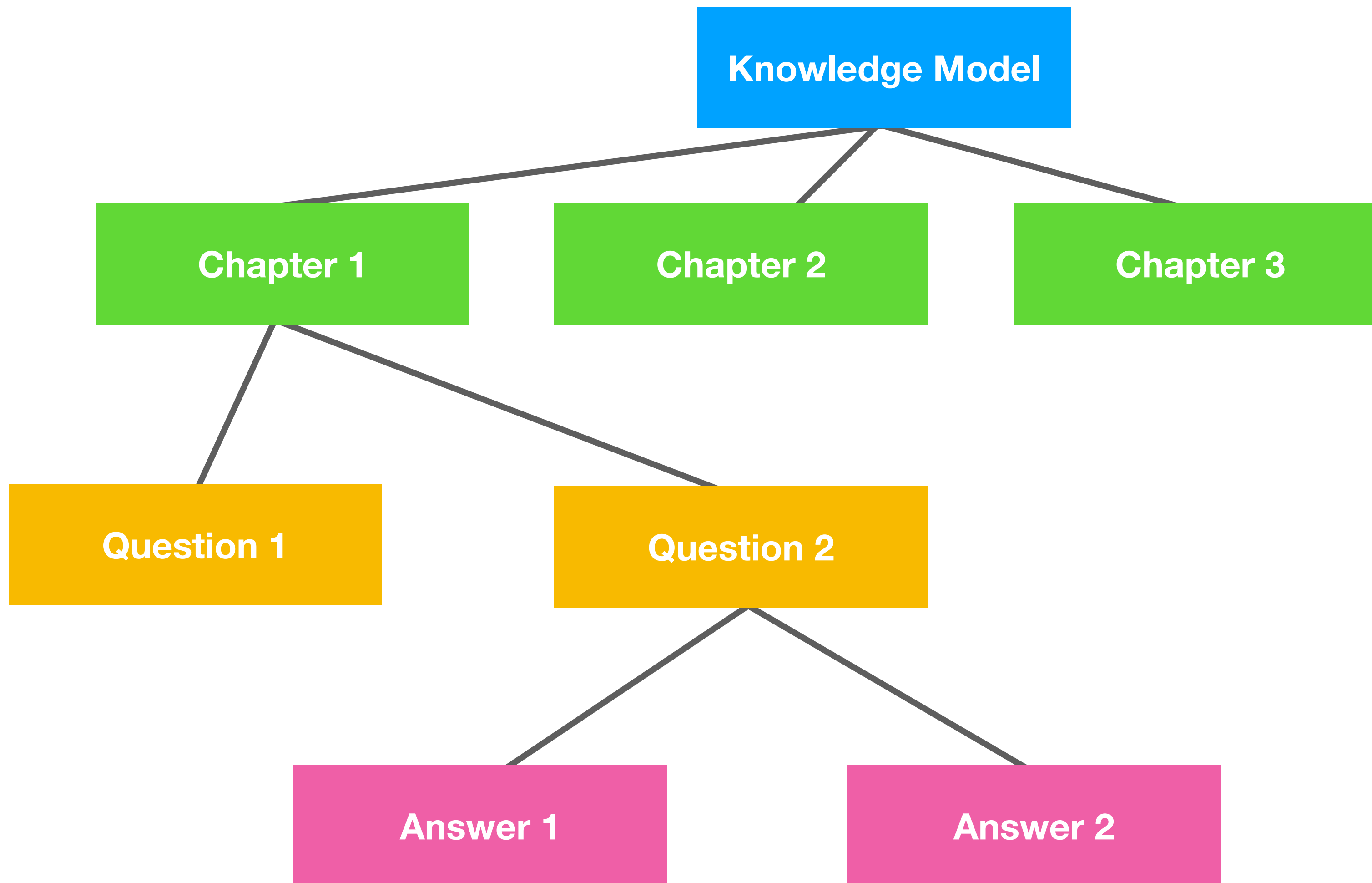
- Manage existing Knowledge Models
- Build & publish new Knowledge Models
- Import Knowledge Models from DSW Registry or other DSW instances

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/M 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

The screenshot shows a form with the following structure:

- 1 Please specify what data sets you will acquire using measurement equipment** (with a plus icon)
 - You can use any name for the data set, make sure that it is meaningful to yourself.
 - 1.a.1 Data set:** (with a plus icon)
 - Input field: |
 - 1.a.2 Who will do the measurements? And where?** (with a plus icon)
 - 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

Two pop-up windows are overlaid on the form:

- 1.a.1 Data set:** (with a plus icon)
 - Input field: |
- 1 Is there any pre-existing data?** (with a plus icon)
 - 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: Integration

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

1 What database will you use? !

- 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:

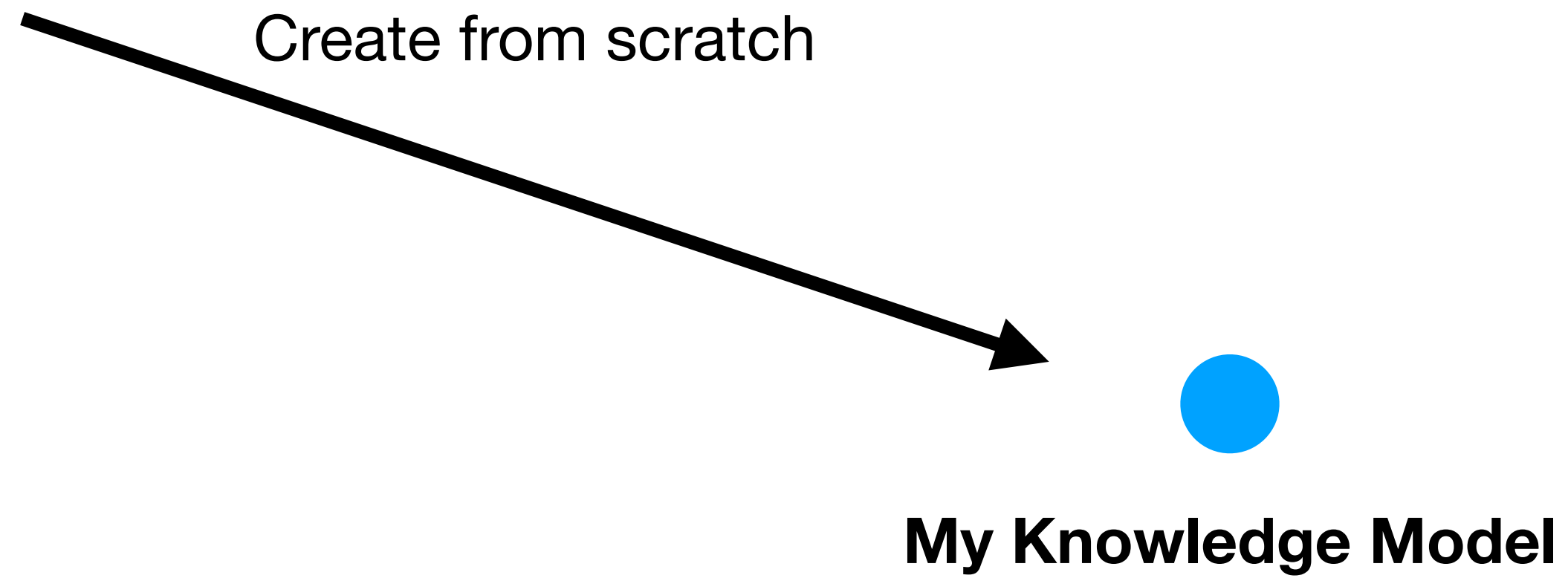
RDF/XML Syntax Specification

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

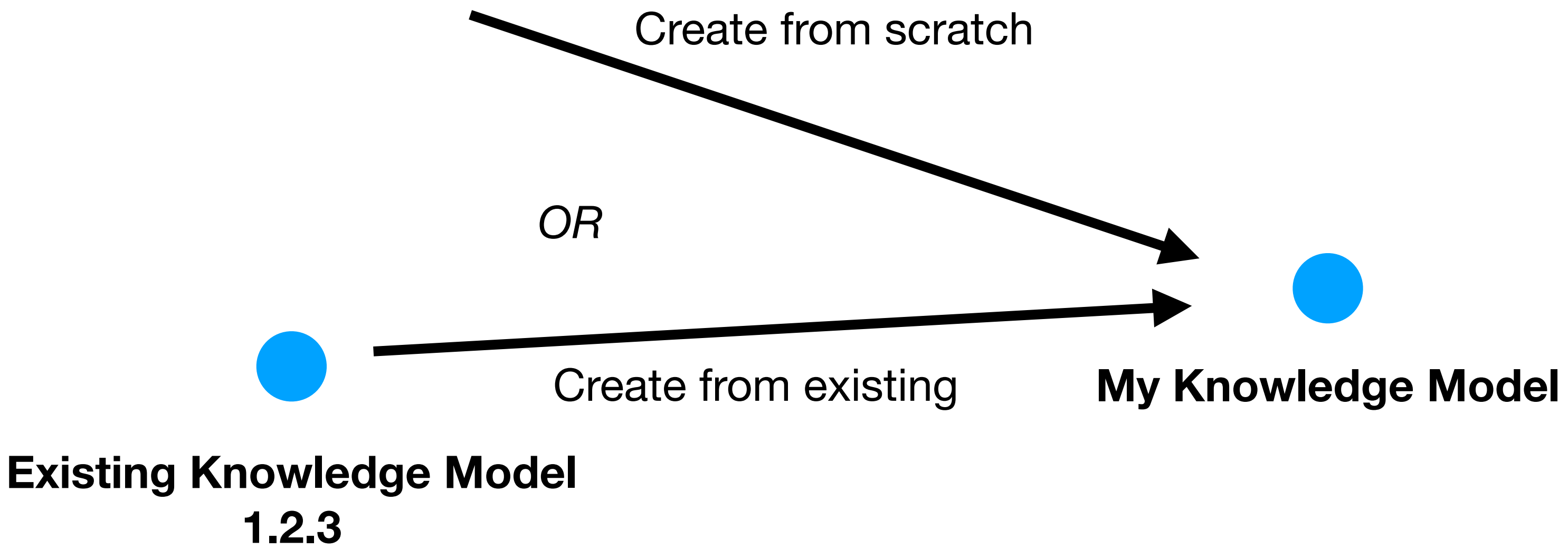
Publishing the Knowledge Model

- Before a Knowledge Model can be used, it has to be **published**
- **Version, description, license, 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

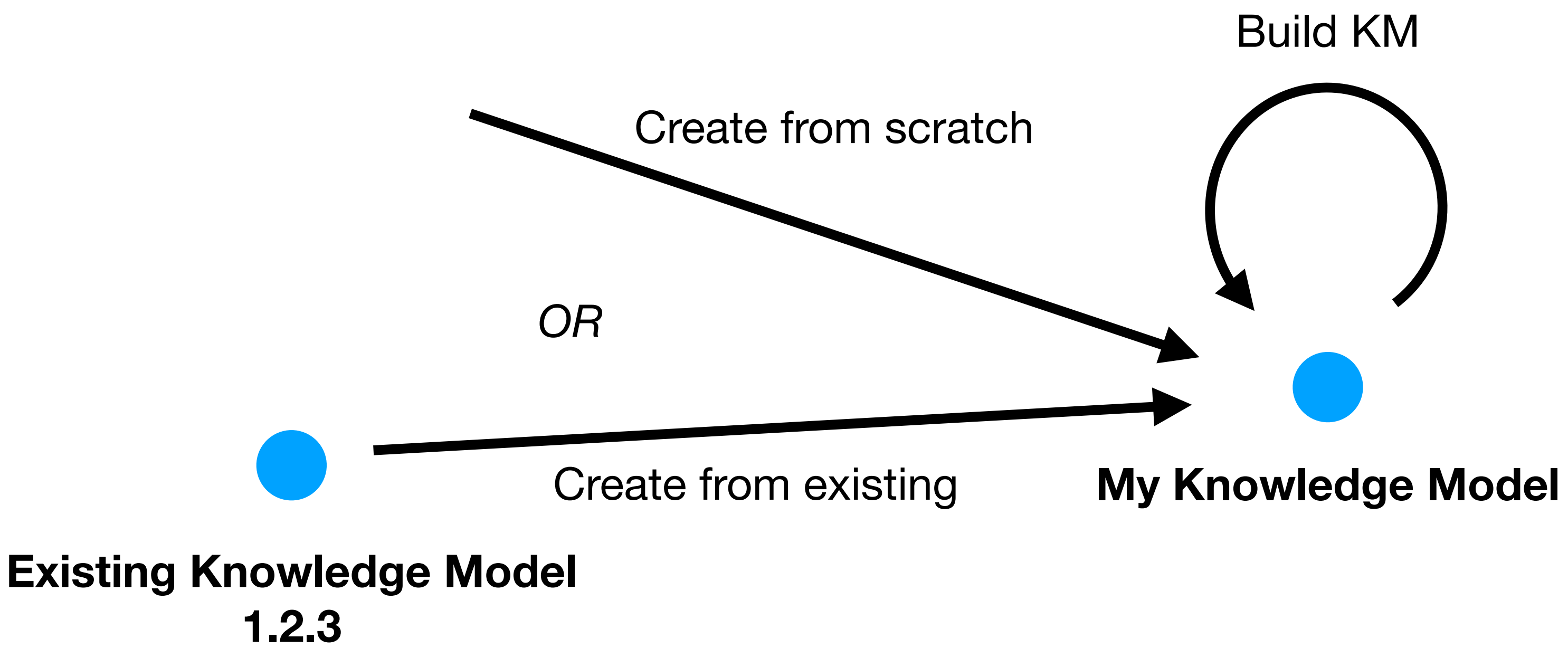
Knowledge Model Workflow



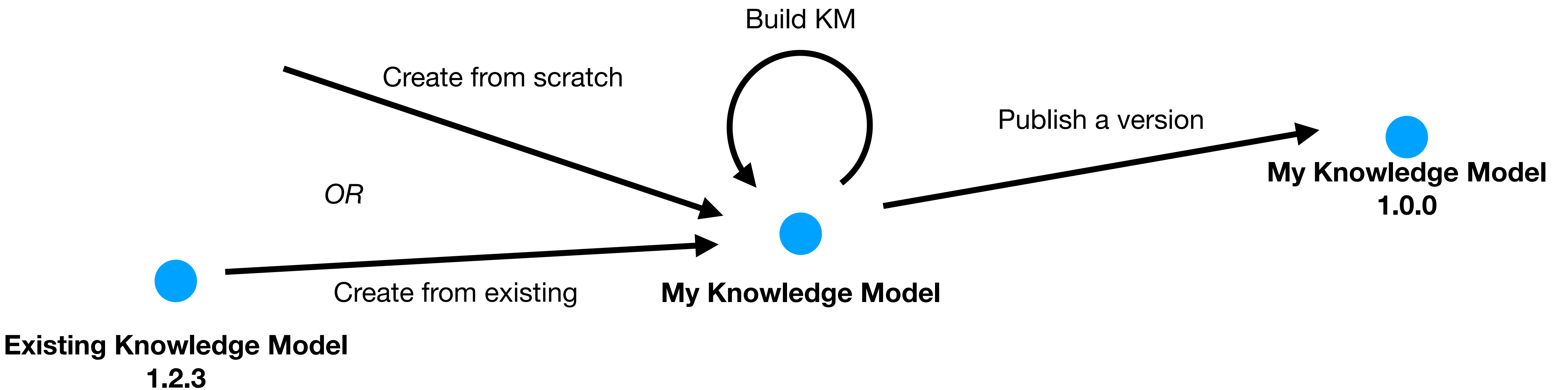
Knowledge Model Workflow



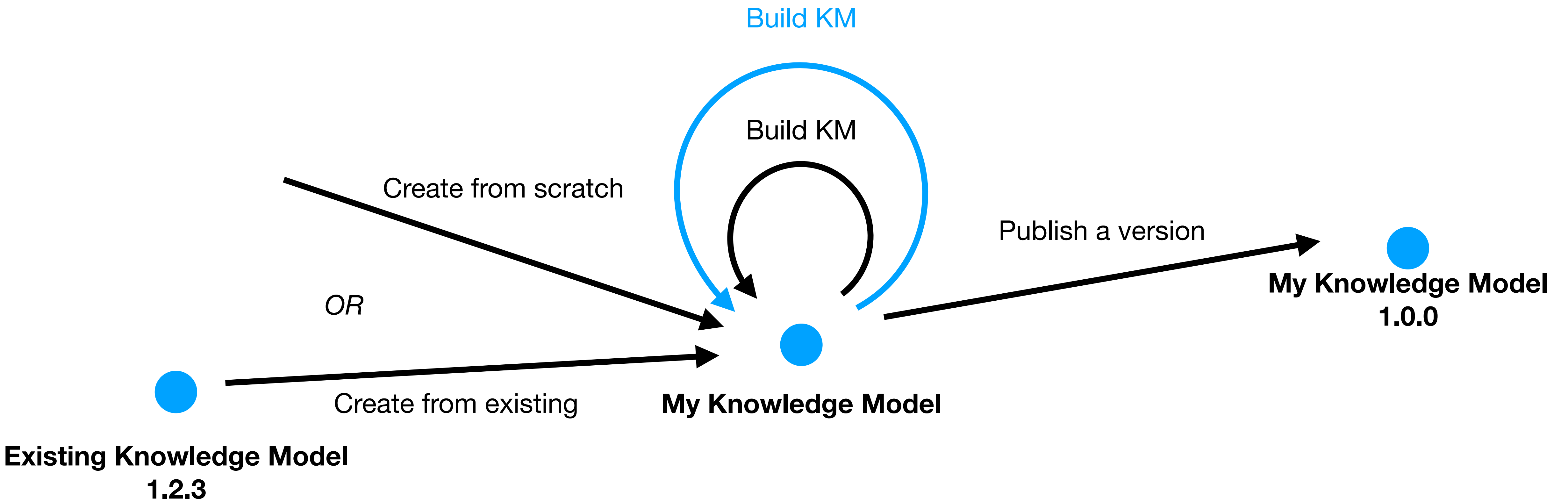
Knowledge Model Workflow



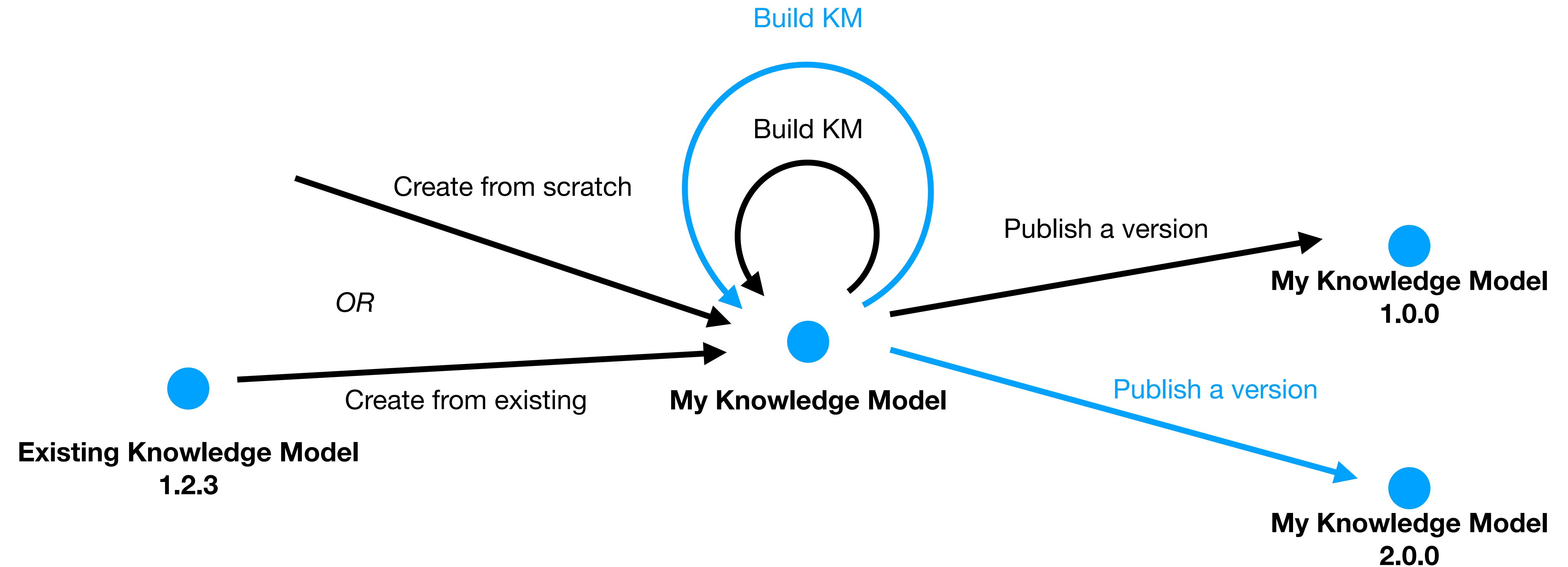
Knowledge Model Workflow



Knowledge Model Workflow



Knowledge Model Workflow



Knowledge Model Import & Export

- From a file
- From DSW Registry



Demo

- Setting up an **integration question**
- **Project** and **document templates**
- Configuring a **submission service**



Learn more

- Data Stewardship Wizard Workshop
<https://doi.org/10.5281/zenodo.3689221>
- DSW Template Development Kit: The First Tutorial
<https://doi.org/10.5281/zenodo.4286272>



Questions & Discussion

Try DS Wizard yourself!



- ds-wizard.org
- info@ds-wizard.org
- https://twitter.com/dswizard_org

