

The *Aspergillus fumigatus* azole knowledge  
repository

Sibbe Bakker

2023-10-03



# Contents

<b>Glossary</b>	<b>5</b>
Knowledge Graph . . . . .	5
Ontology . . . . .	5
SPARQL Protocol And RDF Query Language (SPARQL) . . . . .	5
R programming language . . . . .	6
Resource Description Framework (RDF) . . . . .	6
<b>1 Introduction</b>	<b>9</b>
1.1 What is FAIR data? . . . . .	9
1.2 Why would FAIR data be better data? . . . . .	10
<b>2 How can you play FAIR?</b>	<b>11</b>
2.1 Introduction . . . . .	11
2.2 Choosing the right packages for your experiments . . . . .	12
2.3 Filling in the Excel form. . . . .	14
2.4 FAIRification using ASPAR_KR. . . . .	15
2.5 Submission of your worksheet and RDF data. . . . .	15
<b>3 Worked examples</b>	<b>17</b>
3.1 Air Samples according to Kortenbosch et al. (2022). . . . .	17
<b>4 Applications: How can you benefit of FAIR play?</b>	<b>27</b>
4.1 The SPARQL endpoint. . . . .	27
<b>5 Appendix</b>	<b>29</b>
5.1 Appendix A – Description of ASPAR_KR packages. . . . .	29
5.2 Appendix B – Extend ASPAR_KR with your own packages. . . . .	30
<b>References</b>	<b>33</b>



# Glossary

## Knowledge Graph

A collection of subject, predicate object relation triples.

## Ontology

Studer, Benjamins, and Fensel (1998) describes an ontology as being a ‘formal, explicit specification of a shared conceptualisation’. This refers to the understanding of a topic, being a machine readable version of a model obtained by consensus. An ontology is commonly used to store the rules of a graph database

## SPARQL Protocol And RDF Query Language (SPARQL)

A query language for RDF files. An example of an SPARQL query might be:

```
PREFIX abc: <nul://sparql/exampleOntology#> .
SELECT ?capital ?country
WHERE {
  # All things that have a city name and are a capital.
  ?x abc:cityname ?capital ;
    abc:isCapitalOf ?y.
  # All things that have a country name and are in Africa.
  ?y abc:countryname ?country ;
    abc:isInContinent abc:Africa.
}
```

Which retrieves a list of African countries and their capital city. For further reading see

**“Wikidata:SPARQL Tutorial - Wikidata” (n.d.)** The wikidata SPARQL tutorial, rich with examples to learn with.

**“Course: Querying SIB Resources with SPARQL”** (n.d.) Overview of different biological SPARQL endpoints.

**Sima et al. (2020)** A guide on how to query for evolutionary relationships.

## R programming language

A scientific programming language oriented towards statistics and data visualisation. Installation guide [here](#). An entry level guide on R is available [here](#).

## Resource Description Framework (RDF)

A file from the W3C organisation. This file format can store a large variety of information in a self describing manner. Information is stored in the form of subject, predicate, object relations (see example of Figure 1).



Figure 1: A RDF graph encoding the contact details of Erik Miller. The statement 'Erik is a doctor' is encoded as 'eric:me contact:personalTitle Dr.'

An rdf file of Figure 1 is:

```
@prefix eric:    <http://www.w3.org/People/EM/contact#> .
@prefix contact: <http://www.w3.org/2000/10/swap/pim/contact#> .
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

eric:me contact:fullName "Eric Miller" .
eric:me contact:mailbox <mailto:e.miller123(at)example> .
```

```
eric:me contact:personalTitle "Dr." .  
eric:me rdf:type contact:Person .
```

This file begins with *namespace* definitions, indicating where more information about that group of variables can be found.

Further reading:

- “RDF 1.2 Concepts and Abstract Syntax” (n.d.)





# Chapter 1

## Introduction

### 1.1 What is FAIR data?

To understand the spread of azole resistance traits within the *Aspergillus fumigatus* population, a lot of data must be combined: where is an isolate found? How well can it deal with azoles? What is its genotype? etc. While a lot of data about these subjects is collected, a clear overview of all data is not easy to obtain. Many researchers use different protocols, data formats, and do not share their data in a easy-to-find manner.

This method of (not) sharing research data is limiting the understanding of azole resistance in *A. fumigatus*. The solution is to engage in data sharing that adheres to the FAIR principles (“FAIR Principles” n.d.):

**Findable** The metadata is machine readable, making sure that the dataset can be found by a computer using relevant search terms.

**Accessable** The dataset is stored in a public database, where anyone can access it.

**Interoperable** Metadata should be formal, and it should be possible to combine it with other FAIR datasets.

**Reusable** The dataset is described in a way that it can be reused by another user for a different analysis. Licencing should be set up to make the reuse legal.

In this book, ASPAR\_KR is introduced, a system for managing *A. fumigatus* observations in a FAIR manner. The book consists of a few chapters.

1. Introduction *This chapter.*
2. How can you play FAIR? *An example of how you can start to use ASPAR\_KR for your own needs, with worked code examples.*
3. Other chapters.
4. Appendix

## 1.2 Why would FAIR data be better data?

Making data more FAIR—FAIRification—is in the interest of the researcher and the wider research community. For the researcher, it is beneficial as FAIR data is becoming more important for grand organisations, with funding becoming conditional on FAIR data management (NWO n.d.). Additionally, the research whose data is published in a FAIR format, is more likely to get cited, since the first aspect of FAIR data is being easily *findable* in a public database. For the research community at large, FAIR data is important since FAIR data can be reused more easily, this means that less money is wasted to collect data that already exist. Furthermore, it is easier to replicate FAIR studies, as the meta-data is easier to understand.

---

### Storytime



Throughout this book, we'll follow the story of Marie and Peter—both are mycologist interested in how azole resistance spreads throughout the *A. fumigatus* population. Marie is using ASPAR\_KR to make her new experimental data FAIR while Peter is using ASPAR\_KR for FAIRification of his existing datasets, as his publisher now requires FAIR data management. As we go along, we'll find that Marie has a way easier time in her FAIRification journey than Peter.

---

Besides the direct benefits of FAIR data, you'll also learn about interesting technologies, such as SPARQL, that can make your analysis pipeline more easy.

## Chapter 2

# How can you play FAIR?

### 2.1 Introduction

If you want to make your *A. fumigatus* observations more FAIR, it is best to begin the process from the beginning. This ensures that you do not forget to write down any critical details. To start data entry, you have to be aware of the data structure of the ASPAR\_KR database (Figure 2.1):

1. Investigation: Here you should record the purpose of your study and the author and publishing information. Example: Investigating azole resistance of *Aspergillus fumigatus* spores in Dutch hospitals.
2. Study: Here you record the sub-investigations of your study. Example: Resistance occurrence in the Rijnstate hospital.
3. Observation units: Here you indicate what observations were made. Any experimental factors are indicated at this level. Example: The ward for cancer treatment.
4. Sample: Here you indicate what you sampled. Repeated measures or experimental time point samples can be indicated here. Example: The culture taken from floor near patient room A2 and its MIC dilution series.
5. Assay: Here you indicate what was measured and when. Assay results cannot be analysed further. Example: The PCR amplification of *CYP51A* taken from the culture.



Figure 2.1: Classes available within the ‘ASPAR KR’ database. Each class ‘owns’ lower level classes. For example, a sample has associated assays.

ASPAR\_KR has versions of the Sample and Assay class, called *packages*. These packages are like pre-defined templates for your experiments. They allow ASPAR\_KR to check whether anything is missing and whether everything is filled in correctly, leading to a FAIRer dataset. The aforementioned data structures, excel templates and validation programme, are based on the FAIRDS Nijse, Schaap, and Koehorst (2023).

---

### Storytime



During their 10:15 coffee break Marie and Peter are discussing the subject of FAIR data. “Now the NWO wants researchers to use FAIR data practices! How am I supposed to do that while I am also busy with data analysis?” “Well”, says Marie, “I’ve recently heard of the ASPAR\_KR project, and how it aims for a solution for researchers just like you.” “I’ve already started using it”, she noted as she got out her laptop. When you have verified the basic syntax of your FAIR data template;

---

## 2.2 Choosing the right packages for your experiments

To start, go to ASPAR\_KR metadata configurator. You will be greeted with a webpage as is shown (Figure 2.2). On this webpage you’ll find two fields to fill in. *Project specific metadata* and *Experimental metadata*. When these are filled in, you can click on **GENERATE WORKBOOK**, to make an excel file where the metadata can be filled in.

---

### Storytime



Marie Logs on to her computer and goes to <https://aspar-kr.bioinformatics.nl/template>, where she fills out the web form together with Peter for his research.

---

Using the excel sheet generated by this webpage, the FAIRification process can begin.

## 2.2. CHOOSING THE RIGHT PACKAGES FOR YOUR EXPERIMENTS<sup>13</sup>

The screenshot displays the UNLOCK Metadata Configurator interface. On the left is a sidebar with the UNLOCK logo and a list of navigation links: About, Metadata Configurator (highlighted), Validate Metadata, BioProjects Export, ENA Submission (Beta), Terms Overview, and iRODS Login. The main content area is titled 'Step 1 of 3. Project specific metadata' and contains a section for 'Investigation Information'. This section includes three input fields: 'Investigation identifier' with a placeholder 'e.g. NWO\_UNLOCK', 'Investigation Title', and 'Investigation Description'. Below these fields is a checkbox labeled 'I understand that the metadata will be according to the GDPR requirements'. At the bottom of the main area, there are three definitions: 'Investigation identifier: The identifier comprising the unique name of the institution/database hosting the submission of the investigation data, and the accession number of the investigation in that institution.', 'Investigation title: Human-readable string summarising the investigation.', and 'Investigation description: Human-readable text'. A footer bar at the bottom right contains the text 'UNLOCK (Download / Source code / Help) - v1.0.121-dev'.

**UNLOCK**  
MICROBIAL POTENTIAL

About  
Metadata Configurator  
Validate Metadata  
BioProjects Export  
ENA Submission (Beta)  
Terms Overview  
iRODS Login

Metadata Configurator

Step 1 of 3. Project specific metadata

Investigation Information

Investigation identifier  
e.g. NWO\_UNLOCK

Investigation Title

Investigation Description

☐ I understand that the metadata will be according to the GDPR requirements

**Investigation identifier:** The identifier comprising the unique name of the institution/database hosting the submission of the investigation data, and the accession number of the investigation in that institution.  
**Investigation title:** Human-readable string summarising the investigation.  
**Investigation description:** Human-readable text

UNLOCK (Download / Source code / Help) - v1.0.121-dev

Figure 2.2: Screenshot of the FAIRDS interface for generating a metadata excel file.

### 2.2.1 Investigation specific meta data

The investigation describes the field of study. If your research is (going to be) published, you can fill in the title and abstract of your study here. Otherwise you should come up with a title and abstract. It is important to write clear description fields, so others can easily use your dataset. In *Investigation identifier* it is important that the identifier will be unique to the database.

Whenever you feel like ASPAR\_KR cannot for fill your needs, you can suggest a package. Besides suggesting it is also possible to extend an existing package

### 2.2.2 Experiment specific metadata

What you fill in here depends on your research and the results you obtain. There are packages for ObservationUnits, Samples and Assays, they are documented in Chapter 5.1. At minimum, the required fields of a package should be able to describe your research well enough that somebody can understand the dataset without extra context.

---

#### Storytime



Peter completed a differential study of azole resistance in *A. fumigatus* from compost made from farms *vs* garden waste.

---

## 2.3 Filling in the Excel form.

After creating the form for your investigation, you can start to go to work. The **Excel** sheet generated by the template web app makes a spreadsheet for each class shown in Figure 2.2. It is important to take the time to fill in all of the fields carefully. This makes sure that any future user of the data set can easily understand your experimental design. If you have any questions about the meaning for different column names, check the terms overview on the FAIRDS website.

### 2.3.1 Study

A Study is a set of observational units of which samples are measured to address a specific question. This question, together with the experimental factors used to address it should be specified in the *study description*.

### 2.3.2 Observation unit

In the observation unit, the experimental factors should be indicated in columns of the following format:

id	fcr_watered	fct_sun_hours
1	yes	20
2	no	5

This makes it easy for users of the data set to understand what experimental treatments were applied over the experimental units.

### 2.3.3 Sample

Repeated measures are indicated in the Sample class. Any experimental factor that changes between repeated sampling of the same observation unit is indicated here.

Some packages in of the Sample class ask for `latitude` and `longitude` fields. These should be specified in the WGS84 format, which is also used by open street maps (OSM) or Google earth.

### 2.3.4 Assay

Different measurements of the same sample are indicated here. Class is special, as it also allows references to files that you've previously uploaded to ASPAR\_KR.

## 2.4 FAIRification using ASPAR\_KR.

When the excel sheet is filled in, comes the validation step. Here, the ASPAR\_KR programme makes sure that all syntax is as expected and that the numbers, such as coordinates are valid. If everything is determined to be correct, ASPAR\_KR offers the button **DOWNLOAD RDF**. When this button is clicked, the RDF dataset is downloaded.

If there is an issue with the document, ASPAR\_KR will detect it, and report it in the text box **Description** (Figure 2.3). If you have any questions about the error messages given by ASPAR\_KR, please reach out to Sibbe Bakker.

## 2.5 Submission of your worksheet and RDF data.

When you have verified the basic syntax of your FAIR data template using the web-tool. You can submit it, by sending it to `sibbe<dot>bakker<at>wur.nl`. Then the dataset will be reviewed and if deemed of sufficient quality, will be made part of the next ASPAR\_KR release.

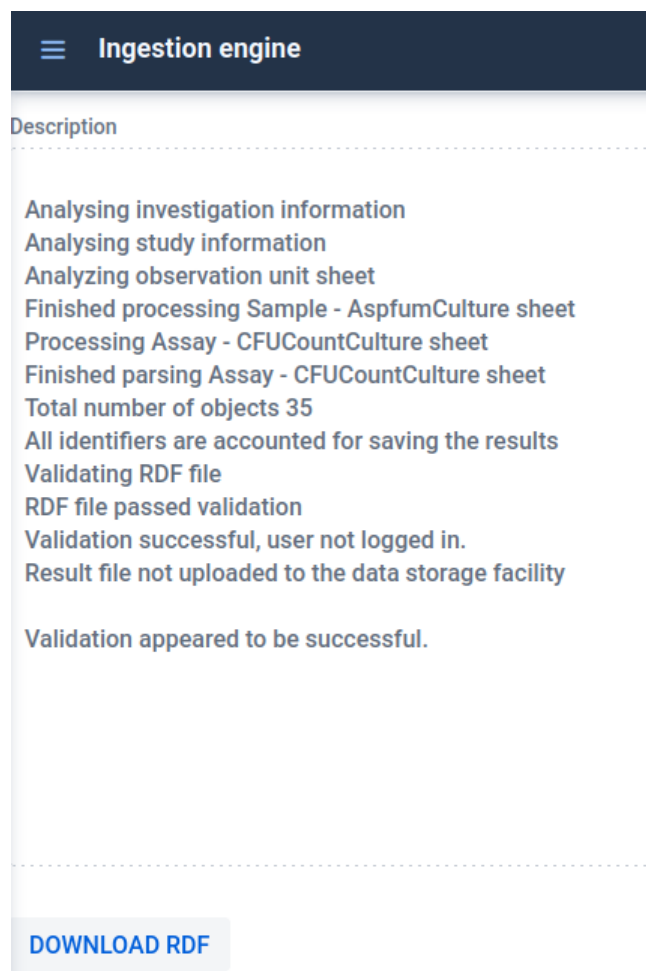


Figure 2.3: Log of the validation actions of ASPAR KR. The FAIRified dataset can be obtained by clicking 'DOWNLOAD RDF'



## Chapter 3

# Worked examples

In this chapter, short examples are given on how to apply the ASPAR\_KR metadata standardisation programme. A download link to the standardised excel sheet is given at the end of each section.

### 3.1 Air Samples according to Kortenbosch et al. (2022).

“Following the example of Peter, we see what it takes to fill in the meta data for a limited air sample study”.

Kortenbosch et al. (2022) reports a method of cheaply taking air samples of *A. fumigatus*. In this technique, **Delta traps** are assigned to *observation units* and left for a certain number of *days*. Then the Delta traps are transferred to a flamingo agar (Zhang et al. (2021)) plate with azole, and without and the **colony forming units** are counted.

Peter is planning to determine the azole resistance of *A. fumigatus* in Arnhem *vs* Nijmegen. Per town, he hangs 4 air sample units, and he plans to plate them on flamingo agar and flamingo agar with itraconazole (4 mg/L).

To record these measurements using ASPAR\_KR, take the following steps:

#### 3.1.1 Set up the metadata sheet

1. Go to <https://aspar-kr.bioinformatics.nl/template>. And fill in step 1 according to Figure 3.1.
  - a. Fill in *Investigation information*:
    - **Identifier**: Arnhem\_vs\_Nijmegen;
    - **Title**: Azole resistance in *A. fumigatus* in Arnhem compared to Nijmegen.

- **Description:** The Itraconazole resistance fraction was studied using the airsampling method of Hylke et al. The Fractions did not differ much. Four airsampling units were placed in each town.
- b. Fill in person information (click **add** when finished).
    - **First name:** Peter.
    - **Last name:** Doe.
    - **email:** peter.doe@example.com.
    - **Organisation:** Example company.
    - **Role:** Being a good example.
    - **ORCHID:** <left empty>.
    - **Department:** Example department
  2. Add Study information (Click: **Study information**). Since the design is simple, we can copy the information from investigation.
  3. Add the standard observational unit template. Click: **Observation unit information > Search a package > default**.
  4. Fill in the experimental metadata. Since we need to keep track of Delta traps, and their corresponding culture we use airstrip samples, and CFU count samples.
    - a. Add the AirStrip sample package: Click: **Sample Information > search a package > Scroll down and add AirStrip > Add template**.
    - b. Add the CFU sample package: Click: **Sample Information > search a package > Scroll down and add CFUCountCulture > Add template**.
    - c. Adjust your optional metadata requirements.
  5. No assay information is needed since Peter did not plan to do genotyping or another test.
  6. Download the excel sheet (**Generate workbook**). You can download the resulting sheet [here](#), to circumvent steps 1-5.

### 3.1.2 Fill in the data

### 3.1.3 The common metadata tables

As mentioned in Section 2.1, between each class, objects are linked by *identifiers*. In the *Arnhem\_vs\_Nijmegen.xlsx* sheet things are no different, you'll find the following 5 tables in this sheet:

1. Table 3.1: The investigation. Links to Study via *investigation identifier*.
2. Table 3.2: The study. Links to observational unit via *study identifier*.

3.1. AIR SAMPLES ACCORDING TO KORTENBOSCH ET AL. (2022). 19

Metadata Configurator

2C393ACD6A0765FAFDA4C488

Step 1 of 3. Project specific metadata

Investigation Information

Investigation identifier

Arnhem\_vs\_Nijmegen

Investigation Title

Azole resistance in A fumigatus in Arnhem compared to Nijmegen

Investigation Description

The Itraconazole resistance fraction was studied using the airsampling method of Hylke et al. The Fractions did not differ much. Four ~~airsampling~~ units were placed in each town.

☒ I understand that the metadata will be according to the GDPR requirements

**Investigation identifier:** The identifier comprising the unique name of the institution/database hosting the submission of the investigation data, and the accession number of the investigation in that institution.

**Investigation title:** Human-readable string summarising the investigation.

**Investigation description:** Human-readable text describing the investigation in more detail.

First name •

Last name •

E-mail

ORCID

Organization •

Department •

Role

Add persons involved in this investigation.

Add

Clear

First Na...	Last Name	E-Mail	ORCID	Organiza...	Departm...	Role
Peter	doe	peter.d...		Exempl...	exempl...	Being a ...

Figure 3.1: Filled in investigation information

3. Table 3.3: The observational unit. Links to Sample via *observational unit identifier*.

These three tables will always be present. The actual research data will is always part of the sample and assay tables.

Table 3.1: Investigation table. Investigation has the investigation identifier column.

investigation iden- ti- fier	investigation title	investigation description	firstname	lastname	email ad- dress	orcid	organization	department
Arnhem- vs-Nijmegen	Azole resis- tance in A fumi- ga- tus in Arn- hem com- pared to Ni- jmegen.	Itraconazole resistance fraction was studied using the airsampling method of Hylke et al. The Fractions did not differ much. Four airsampling units were placed in each town.	Peter	Doe	p. p@e x.co m		org	s

Table 3.2: The Study table.

study identifier	study description	study title	investigation identifier
Arnhem-vs- nijmegen	4 samples in each town were made using hylkes method	Azole resistance in A fumigatus in Arnhem compared to Nijmegen.	Arnhem_vs_Nijmegen

Table 3.3: Observational unit table.

observation unit identifier	observation unit name	observation unit description	study identifier
arnhem	Arnhem	Measurements in Arnhem	Arnhem-vs-nijmegen
nijmegen	Nijmegen	Measurements in Nijmegen	Arnhem-vs-nijmegen

### 3.1.4 Experimental data tables

#### 3.1.4.1 Air samples

To fill in these tables Peter hangs the Air sample strips up for 4 weeks: between the 5<sup>th</sup> of November and the 26<sup>th</sup> of November. He records the minimal information for the air sample strips:

1. start end end date.
2. Collection location.
  - Lat
  - Long

For taxonomy ID he picks 32644: unidentified, since Peter does not know which species he will collect on his Delta Trap. For the culture made using the Delta Trap, he selects for *A. fumigatus*, so 746128

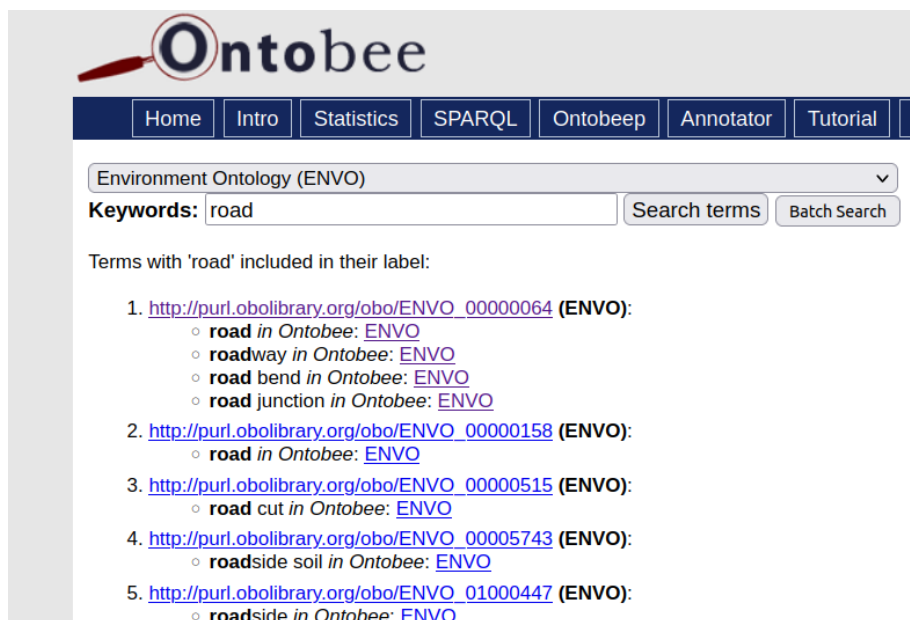
He also collects the recommended and optional data:

- Altitude
- Environmental context:
  - Broad scale. Overall environment where the sample is found? What has the biggest influence on the environment of the sample?
  - Local scale. Things in the vicinity of the sample that *may* influence the outcome of the experiment.
  - Environmental medium. Things that directly surrounded your sample prior to sampling.

In Arnhem & Nijmegen he hangs some samples in trees of the the park, others at the train station at traffic poles. Since he wants to record these terms as part of the environmental context, he looks up the terms on ontobee/envo (Figure 3.2). He finds that:

- The Broad scale environmental context is “City”
- The local environmental context is “Road” or “Park”.
- The medium for samples in a park is “tree” and for the train station: “pole”.

Using this information, he fills in the Delta Trap excel sheet.



**Ontobee**

Home Intro Statistics SPARQL Ontobee Annotator Tutorial F

Environment Ontology (ENVO) ▾

Keywords: road Search terms Batch Search

Terms with 'road' included in their label:

1. [http://purl.obolibrary.org/obo/ENVO\\_00000064](http://purl.obolibrary.org/obo/ENVO_00000064) (ENVO):
  - road in Ontobee: ENVO
  - roadway in Ontobee: ENVO
  - road bend in Ontobee: ENVO
  - road junction in Ontobee: ENVO
2. [http://purl.obolibrary.org/obo/ENVO\\_00000158](http://purl.obolibrary.org/obo/ENVO_00000158) (ENVO):
  - road in Ontobee: ENVO
3. [http://purl.obolibrary.org/obo/ENVO\\_00000515](http://purl.obolibrary.org/obo/ENVO_00000515) (ENVO):
  - road cut in Ontobee: ENVO
4. [http://purl.obolibrary.org/obo/ENVO\\_00005743](http://purl.obolibrary.org/obo/ENVO_00005743) (ENVO):
  - roadside soil in Ontobee: ENVO
5. [http://purl.obolibrary.org/obo/ENVO\\_01000447](http://purl.obolibrary.org/obo/ENVO_01000447) (ENVO):
  - roadside in Ontobee: ENVO

Figure 3.2: Lookup of a machine readable term [on Ontobee](<https://ontobee.org/search?ontology=ENVO&keywords=station&submit=Search+terms>).

#### 3.1.4.2 Culture Samples

After filling in the air sample sheet, he fills in the information for the two layer culture:

As a selection medium, he adds itraconazole to a concentration of 4 g/L ( $x$  molar)<sup>1</sup>, which has CHEMBL1835949 as the standard identifier.

#### 3.1.5 Convert the sheet to ttl format

1. Go to <https://aspar-kr.bioinformatics.nl/validate>.
2. Upload the finalised sheet.
3. Click “DOWNLOAD RDF”.

#### 3.1.6 Analysis of the FAIR data with R

To analyse the data we can use the `rdflib` package in R together with the `tidyverse`.

<sup>1</sup>Molecular weight of itraconazole is 705.6 g/mol, he added 4 g to 1 liter of flamingo medium. Therefore  $((1 * \text{mole}) / (705.6 * \text{gram})) * (4 * \text{gram}) = \text{approx. } 5.6689342 \text{ mmol}$ .

### 3.1. AIR SAMPLES ACCORDING TO KORTENBOSCH ET AL. (2022). 23

```
base::dir.create("hylke_air_method_example")
destfile="./hylke_air_method_example/data.ttl"
fileURL <-
  "https://git.wur.nl/aspar_kr/aspar/-/raw/main/example_files/Arnhem_vs_Nijmegen.ttl?ref_type=heads"
if (!base::file.exists(destfile)) {
  utils::download.file(fileURL ,destfile,method="w")
}

rdf <- rdflib::rdf_parse("hylke_air_method_example/data.ttl",
                        format = "turtle")
rdf
```

```
## Total of 416 triples, stored in hashes
## -----
## <http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/stu_arnhemVsNijmegen/obs_nijme
## <http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/stu_arnhemVsNijmegen/obs_nijme
## <http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/stu_arnhemVsNijmegen> <http://
## <http://fairbydesign.nl/ontology/selection_medium> <http://www.w3.org/1999/02/22-rdf-syntax-ns
## <http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/stu_arnhemVsNijmegen/obs_nijme
## <http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/stu_arnhemVsNijmegen/obs_arnhe
## <http://fairbydesign.nl/ontology/biosafety_level> <http://schema.org/valueRequired> "true"^^<h
## <http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/stu_arnhemVsNijmegen/obs_nijme
## <http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/stu_arnhemVsNijmegen/obs_nijme
## <http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/stu_arnhemVsNijmegen/obs_arnhe
##
## ... with 406 more triples
```

After reading the RDF object we can perform simple SPARQL queries to get tibble data.

```
sparql_query <-
  ' # Select the first 5 of everything...
  SELECT * WHERE {?s ?p ?o} LIMIT 5
  '

result <- rdflib::rdf_query(rdf, sparql_query)
result
```

```
## # A tibble: 5 x 3
##   s                                     p      o
##   <chr>                                <chr> <chr>
## 1 http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/st~ http~ 2
## 2 http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/st~ http~ A tw~
## 3 http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/st~ http~ arnh~
## 4 http://fairbydesign.nl/ontology/selection_medium                    http~ http~
## 5 http://fairbydesign.nl/ontology/inv_arnhemVsNijmegenComparison/st~ http~ CHEM~
```

Now, to plot an overview of Peter's experiment we can use:

```
sparql_query <-

' # Select the first 5 of everything...
prefix ppeo:      <http://purl.org/pp eo/PPEO.owl#>
prefix jerm:      <http://jermontology.org/ontology/JERMOntology#>
prefix fair:      <http://fairbydesign.nl/ontology/>
prefix rdfs:      <http://www.w3.org/2000/01/rdf-schema#>
prefix schema:    <http://schema.org/>
SELECT ?observation_label ?sample_label (COUNT(?sample_label) as ?n) ?total_cfu ?sel
  # Get the samples of interest
  ?observation_unit a ppeo:observation_unit .
  ?observation_unit jerm:hasPart ?parts .
  ?parts a jerm:Sample .
  ?parts fair:packageName "TwoLayerCulture" .
  ?observation_unit schema:name ?observation_label .
  ?parts schema:name ?sample_label .

  # Experimental data
  ?parts fair:total_cfu ?total_cfu .
  ?parts fair:selection_cfu ?selection_cfu .
} GROUP BY ?observation_unit
'

result <- rdflib::rdf_query(rdf, sparql_query)
result
```

```
## # A tibble: 2 x 5
##   observation_label sample_label      n total_cfu selection_cfu
##   <chr>            <chr>      <dbl>    <dbl>        <dbl>
## 1 The city of Arnhem Arnhem Centraal culture      4      66          28
## 2 The city of Nijmegen Nijmegen stationsplein cul~      4      57          21
```

The result table, we can modify and plot using the tidyverse.

```
result |>
  dplyr::mutate(resistance_fraction = selection_cfu / total_cfu) |>
  ggplot2::ggplot(ggplot2::aes(x = observation_label, y = resistance_fraction)) +
  ggplot2::geom_bar(stat="identity")
```



3.1. AIR SAMPLES ACCORDING TO KORTENBOSCH ET AL. (2022). 25





## Chapter 4

# Applications: How can you benefit of FAIR play?

Now comes the most interesting part of the ASPAR\_KR system, the usage of others data to come to a better scientific insight. In this chapter, data analysis workflows with R will be given.

### 4.1 The SPARQL endpoint.

Using the SPARQL endpoint

```
# Query to get the distance from a sample to a place.
# Returns the distance in km.
PREFIX jerm: <http://jermontology.org/ontology/JERMontology#>
PREFIX schema: <http://schema.org/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX uom: <http://www.opengis.net/def/uom/OGC/1.0/>
SELECT ?d_km
WHERE {
    # Get the samples
    ?p ?o jerm:Sample .

    # Take only one specific sample.
    ?p schema:identifier ?sample_label .
    FILTER(?sample_label = "Sample_D1_a_8_8") .
```

```

# Get the place
?p geo:hasGeometry ?geo .
?geo geo:asWKT ?point

# Where is den Helder?
SERVICE <https://query.wikidata.org/sparql> {
  # What things are a Dutch municipality?
  ?municipality wdt:P31 wd:Q2039348.
  # What things have a place?
  ?municipality wdt:P625 ?PlaceOfDenHelder .
  # Take only the thing that is Den Helder.
  FILTER(?municipality = wd:Q9911) .
}

# What is the distance?
BIND (geof:distance(?point, ?PlaceOfDenHelder,
uom:metre) as ?dist)
# Convert to kilo metre.
BIND (?dist / 1000 as ?d_km)
}

```

## Chapter 5

# Appendix

### 5.1 Appendix A – Description of ASPAR\_KR packages.

A package is a variation of the classes in Figure 2.1. A list of the packages, and the fields available in them is present in the the terms overview. If a package you need is not available, please refer to Annex 5.2.

#### 5.1.1 Observation unit packages

**Default** Standard observational unit. Here you fill in treatments that were applied on the experimental unit itself.

#### 5.1.2 Sample packages

**Compost** A sample of compost. It is important to indicate whether it is has azoles in it.

**CFUCountCulture** Results of a plate count assay. One can use this package if you put material which originates from a single Sample on plate culture.

**AirStrip** A airstrip sample taken according to Kortenbosch et al. (2022).

#### 5.1.3 Assay packages

Like a PCR test result, can be added upon request.

## 5.2 Appendix B – Extend ASPAR\_KR with your own packages.

### 5.2.1 Planning the addition.

Since it is not possible for the developer of ASPAR\_KR to know about all the research data that should be standardised with ASPAR\_KR, a package can be requested on gitlab<sup>1</sup> If you want to add the package yourself, it is possible, by understanding the class structure of ASPAR\_KR Figure 5.1.



Figure 5.1: Classes available within the ‘ASPAR KR’ database. Each class ‘owns’ lower level classes. For example, a sample has associated assays.

### 5.2.2 Adding the packages.<sup>2</sup>

First we obtain the latest version of the metadata files as such:

```
suppressMessages(require(tidyverse))
#' Constructs the download link to the FAIRDS metadata files.
#' @param type The type of metadata to obtain.
.meta_data_link_builder <- function(type = "terms") {
  .accepted <- c("terms", "Sample", "Investigation", "ObservationUnit", "Assay")
  if (!(type %in% .accepted)) {
    stop(paste("Given type", type, " is not within", .accepted))
  }
  paste0("https://gitlab.com/m-unlock/fairds-metadata/-/raw/main/",
        type, ".tsv?ref_type=heads", collapse = "")
}

# Download the terms sheet.
terms <- readr::read_delim(.meta_data_link_builder("terms"))

## Rows: 682 Columns: 7
## -- Column specification -----
## Delimiter: "\t"
## chr (7): Item label, Requirement, Value syntax, Example, Preferred unit, URL...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

<sup>1</sup>See ‘filing issues’ for more information on how to do this.

<sup>2</sup>Note: It is possible to add packages completely by your self after obtaining push permissions from Jasper Koehorst, but this is out of the scope of this guide.

```
sample_meta_data <- readr::read_delim(.meta_data_link_builder("Sample"))
```

```
## Rows: 2421 Columns: 4
## -- Column specification -----
## Delimiter: "\t"
## chr (4): # Level, Package name, Item label, Requirement
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Then we can add a sample:

```
# add a sample.

.add_package <- function(packages,
                          level = NA,
                          name = NA,
                          label = NA,
                          requirement = NA) {
  .accepted_requirement <- c("optional", "mandatory", "ignore");
  for (re in requirement) {
    if (!(requirement %in% .accepted_requirement)) {
      stop(paste("Given requirement", re,
                  " is not within", .accepted_requirement))
    }
  }

  df <- dplyr::tibble("# Level" = level,
                     "Package name" = name,
                     "Item label" = label,
                     "Requirement" = requirement)
  rbind(df, packages)
}

new_packages <- .add_package(sample_meta_data,
                             level = "Sample",
                             name = "test",
                             label = "label",
                             requirement = "optional")

new_packages
```

```
## # A tibble: 2,422 x 4
##   `# Level` `Package name` `Item label` Requirement
##   <chr>      <chr>          <chr>      <chr>
## 1 Sample   test            label      optional
```

##	2	Sample	default	sample identifier	mandatory
##	3	Sample	default	sample description	mandatory
##	4	Sample	default	sample name	mandatory
##	5	Sample	default	ncbi taxonomy id	mandatory
##	6	Sample	default	scientific name	mandatory
##	7	Sample	default	biosafety level	mandatory
##	8	Sample	default	sampling strategy	optional
##	9	Sample	default	sample treatment	optional
##	10	Sample	default	observation unit identifier	mandatory
##	# i	2,412	more rows		



# References

- “Course: Querying SIB Resources with SPARQL.” n.d. Accessed September 19, 2023. <https://edu.sib.swiss/course/view.php?id=440>.
- “FAIR Principles.” n.d. GO FAIR. Accessed September 19, 2023. <https://www.go-fair.org/fair-principles/>.
- Kortenbosch, Hylke H., Fabienne Van Leuven, Bas J. Zwaan, and Eveline Snelders. 2022. “Catching More Air: An Effective and Simple-to-Use Air Sampling Approach to Assess Aerial Resistance Fractions in *Aspergillus Fumigatus*.” Preprint. Microbiology. <https://doi.org/10.1101/2022.11.03.515058>.
- Nijse, Bart, Peter J Schaap, and Jasper J Koehorst. 2023. “FAIR Data Station for Lightweight Metadata Management and Validation of Omics Studies.” *GigaScience* 12 (January): giad014. <https://doi.org/10.1093/gigascience/gi ad014>.
- NWO. n.d. “Research Data Management | NWO.” Accessed September 19, 2023. <https://www.nwo.nl/en/research-data-management>.
- “RDF 1.2 Concepts and Abstract Syntax.” n.d. Accessed September 19, 2023. <https://www.w3.org/TR/rdf12-concepts/>.
- Sima, Ana Claudia, Christophe Dessimoz, Kurt Stockinger, Monique Zahn-Zabal, and Tarcisio Mendes de Farias. 2020. “A Hands-on Introduction to Querying Evolutionary Relationships Across Multiple Data Sources Using SPARQL.” *F1000Research* 8 (July): 1822. <https://doi.org/10.12688/f1000research.21027.2>.
- Studer, Rudi, V. Richard Benjamins, and Dieter Fensel. 1998. “Knowledge Engineering: Principles and Methods.” *Data & Knowledge Engineering* 25 (1): 161–97. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6).
- “Wikidata:SPARQL Tutorial - Wikidata.” n.d. Accessed September 19, 2023. [https://www.wikidata.org/wiki/Wikidata:SPARQL\\_tutorial](https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial).
- Zhang, Jianhua, Alfons J. M. Debets, Paul E. Verweij, and Sijmen E. Schoustra. 2021. “Selective Flamingo Medium for the Isolation of *Aspergillus Fumigatus*.” *Microorganisms* 9 (6, 6): 1155. <https://doi.org/10.3390/microorganism s9061155>.