



# **A FAIR, Domain-aware (Meta-)Data Sharing Platform for (Experimental) Material Science**

Markus Scheidgen

[markus.scheidgen@physik.hu-berlin|fhi.mpg.de](mailto:markus.scheidgen@physik.hu-berlin|fhi.mpg.de)

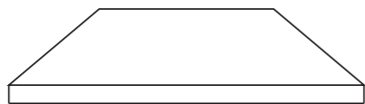
# Agenda

- data, meta-data, and domains
- **FAIR** and *domain aware* (meta-)data sharing
- **NOMAD**'s domain aware search for material science data

# data and meta-data



# data and meta-data



# data and meta-data

***data***

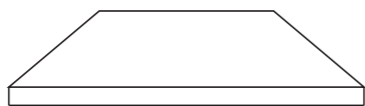
*raw data*

010101

101100

010011

110111



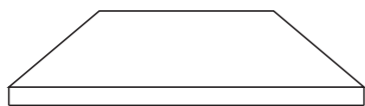
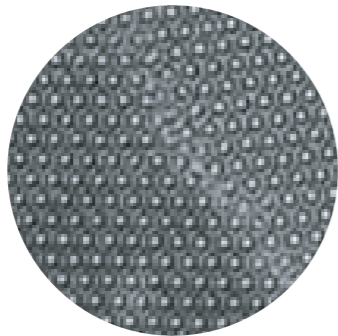
# data and meta-data

***data***

*raw data*

*processed data*

010101  
101100  
010011  
110111



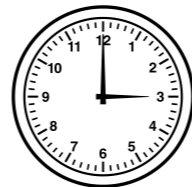
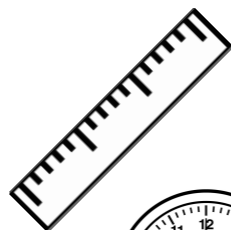
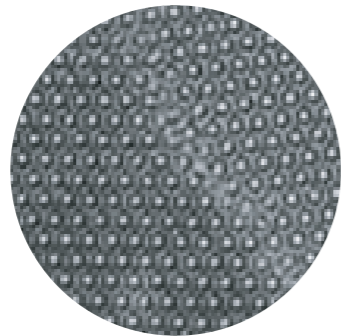
# data and meta-data

***data***

*raw data*

*processed data*

010101  
101100  
010011  
110111



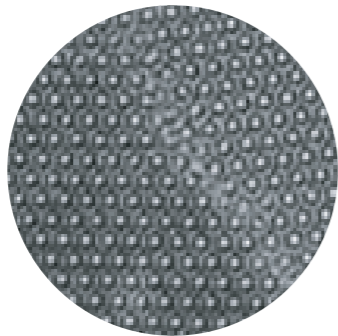
# data and meta-data

**data**

*raw data*

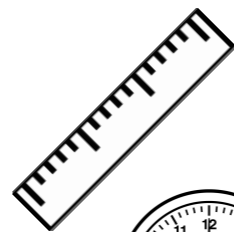
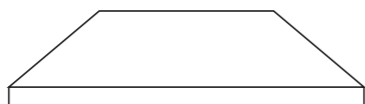
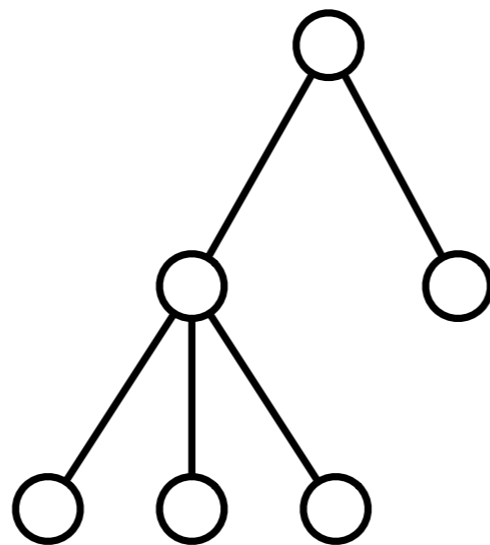
010101  
101100  
010011  
110111

*processed data*



**meta-data**

*instrumentation*





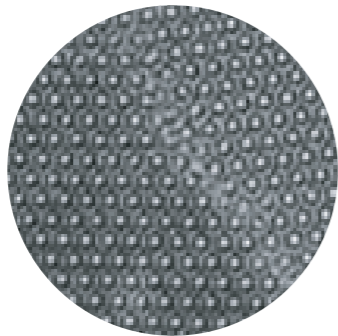
# data and meta-data

## **data**

*raw data*

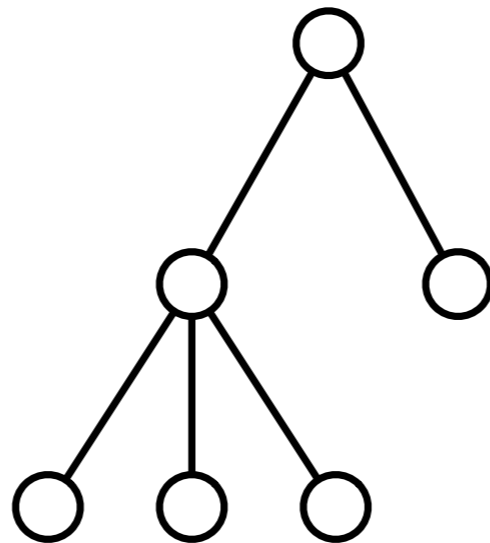
010101  
101100  
010011  
110111

*processed data*

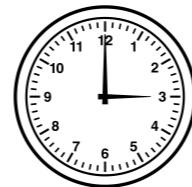
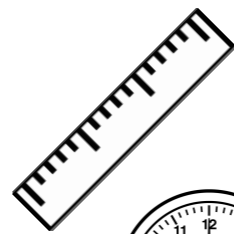
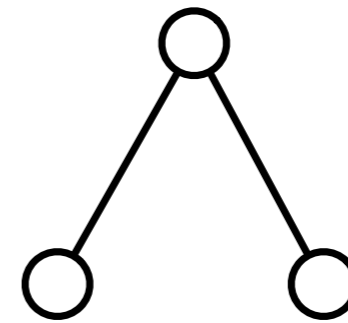


## **meta-data**

*instrumentation*



*human entry*



# data and meta-data

**data**

*raw data*

*processed data*

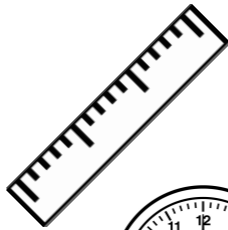
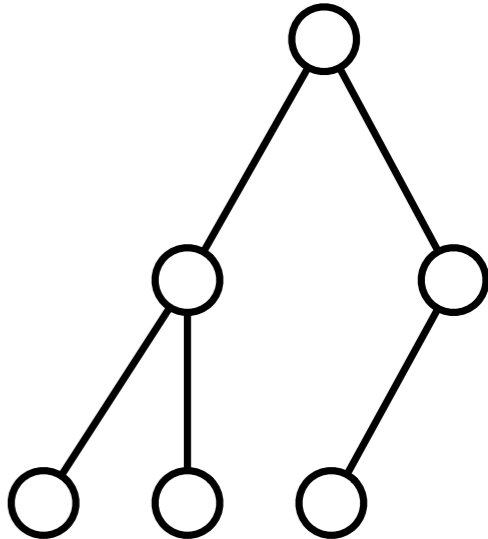
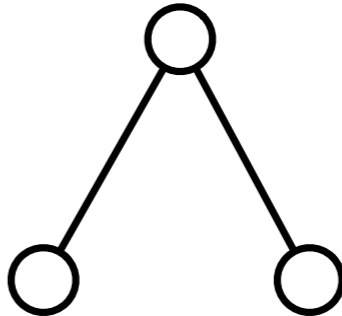
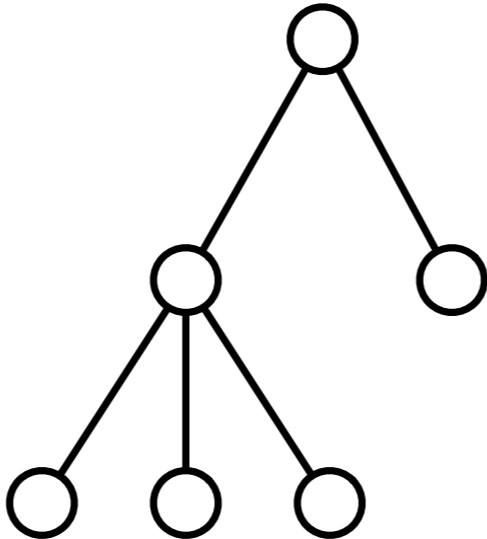
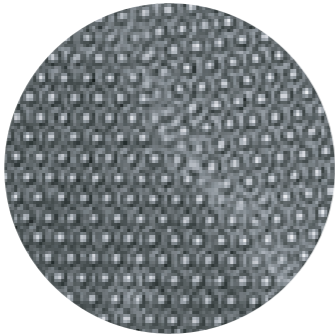
**meta-data**

*instrumentation*

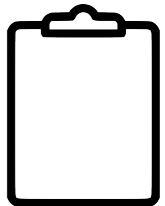
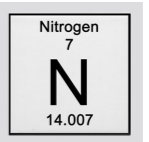
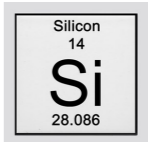
*human entry*

*electronic lab notebooks*

010101  
101100  
010011  
110111



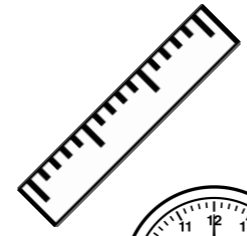
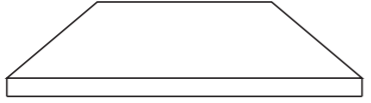
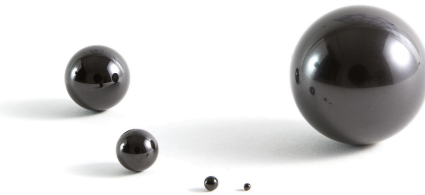
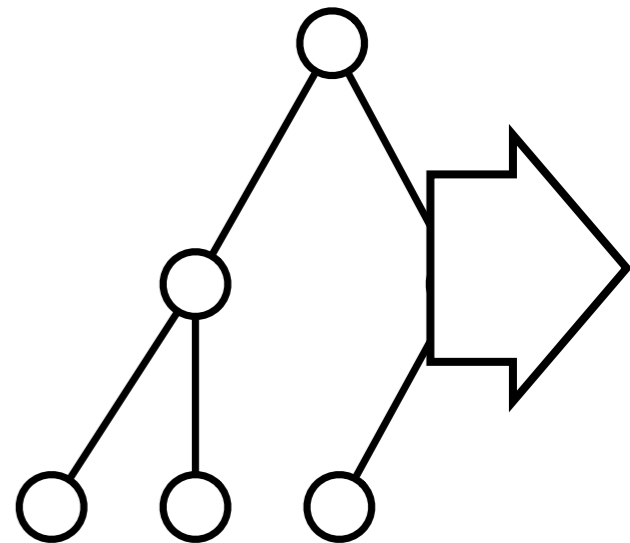
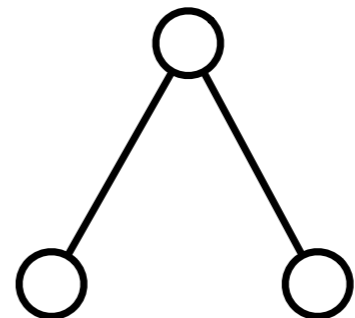
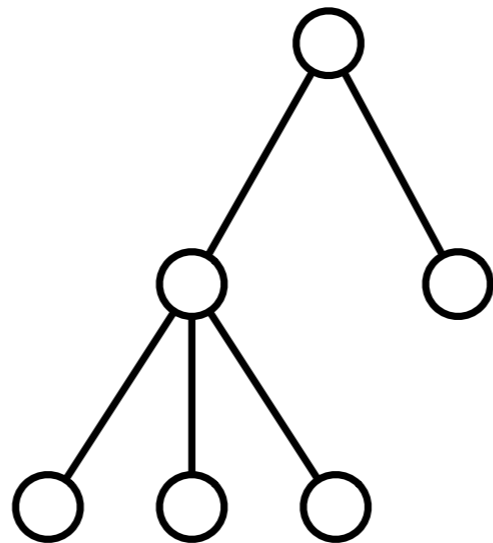
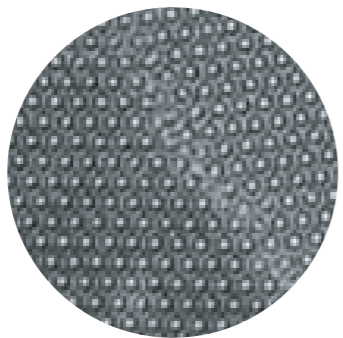
**preparation**



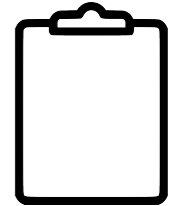
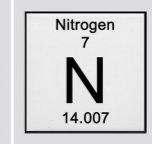
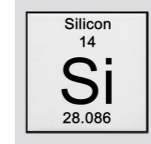
# data and meta-data

**data**  
*raw data*    *processed data*    **meta-data**  
*instrumentation*    *human entry*    *electronic lab notebooks*

010101  
101100  
010011  
110111



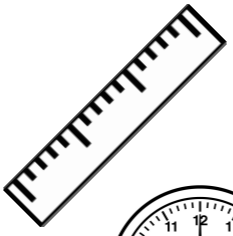
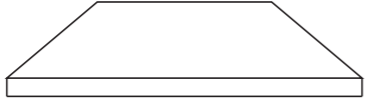
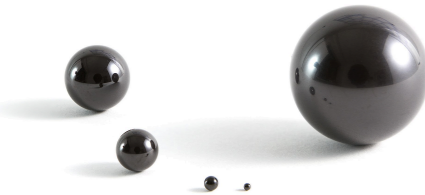
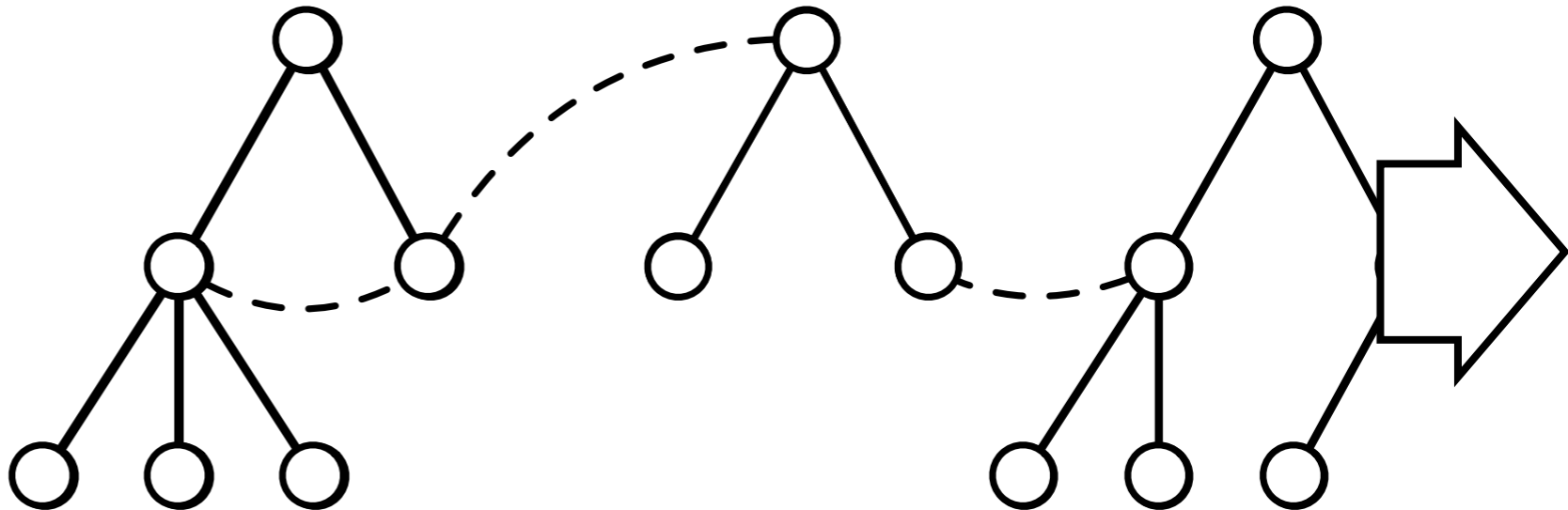
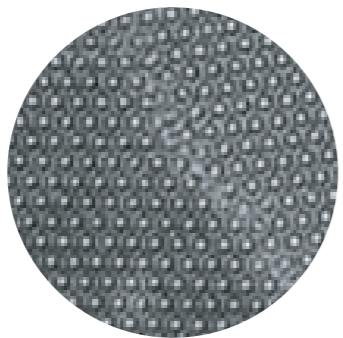
**preparation**



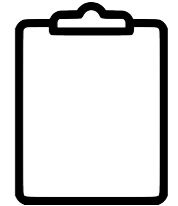
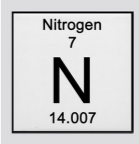
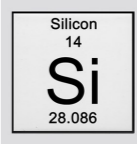
# data and meta-data

**data**  
*raw data*      *processed data*      **meta-data**  
*instrumentation*      *human entry*      *electronic lab notebooks*

010101  
101100  
010011  
110111



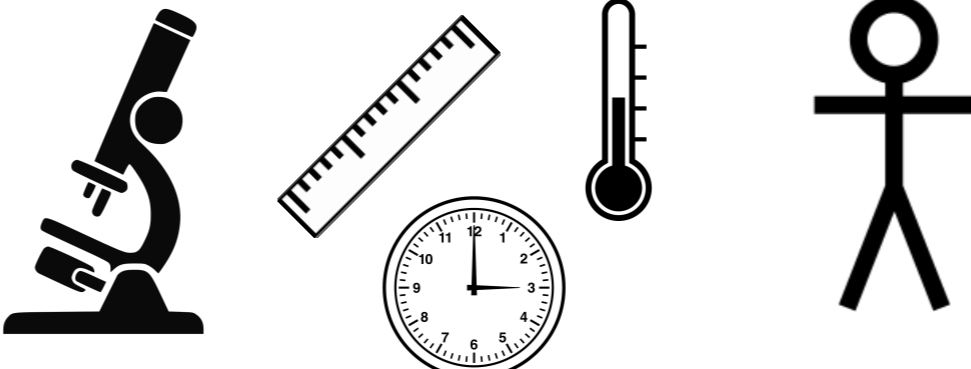
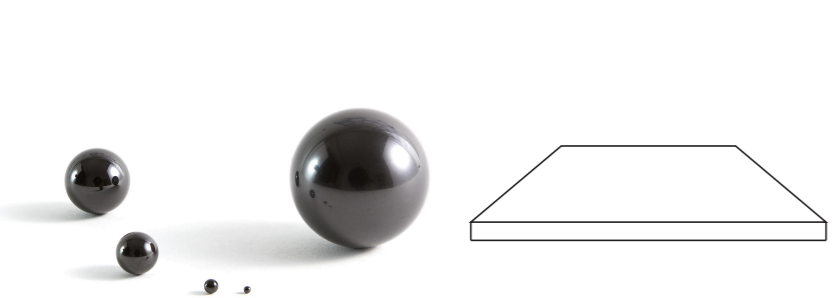
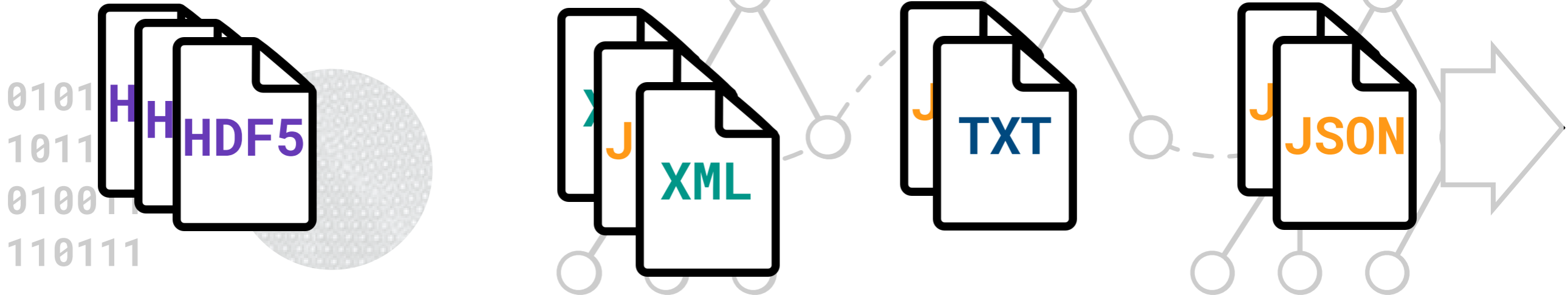
**preparation**



# data and meta-data

**data**      **meta-data**

*raw data*    *processed data*    *instrumentation*    *human entry*    *electronic lab notebooks*



**preparation**

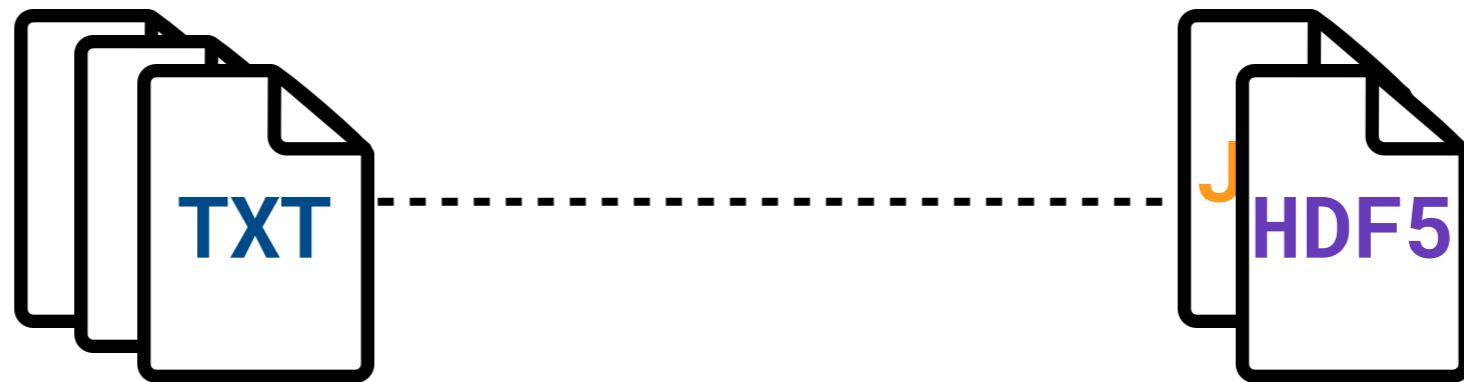
1      2      3      4

Silicon 14 <b>Si</b> 28.086	Nitrogen 7 <b>N</b> 14.007
--------------------------------------	-------------------------------------

# data and meta-data



# data and meta-data



# data and meta-data

**data**

*raw data*

*processed data*

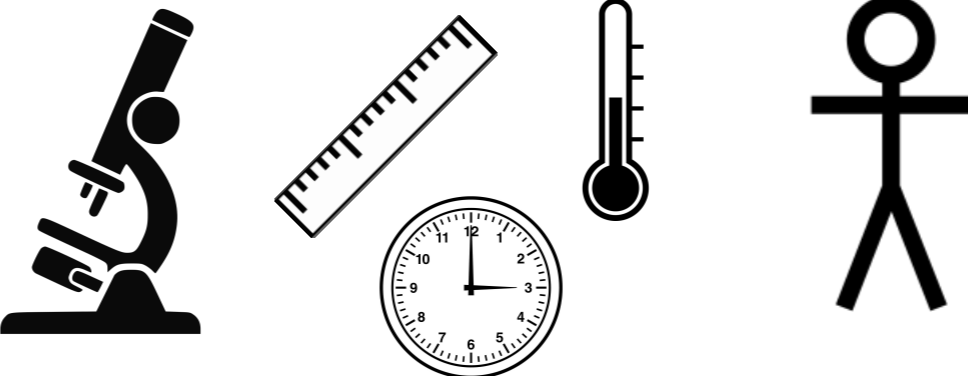
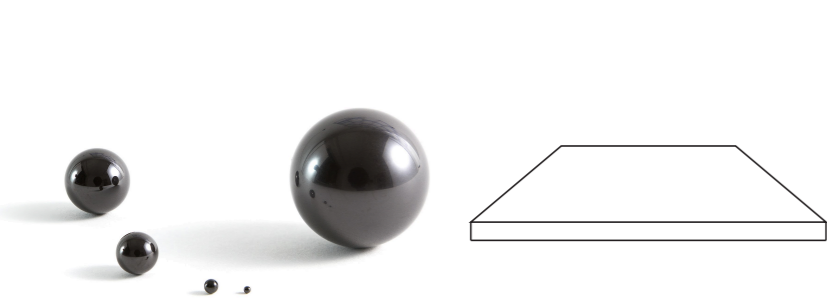
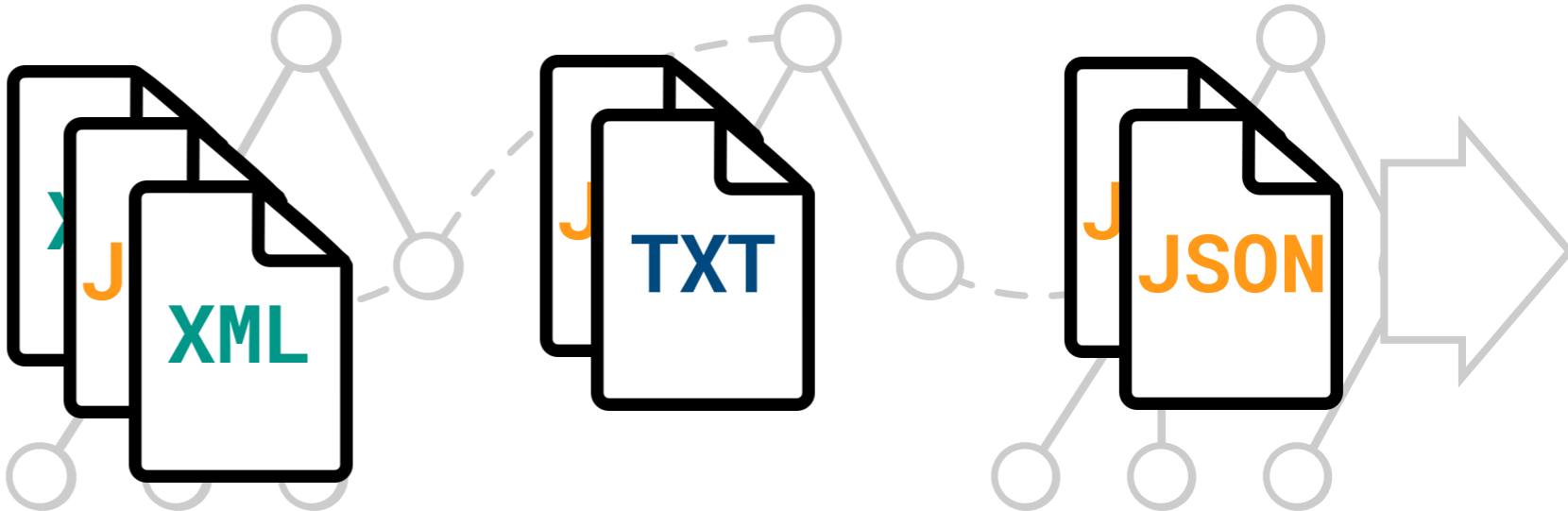


**meta-data**

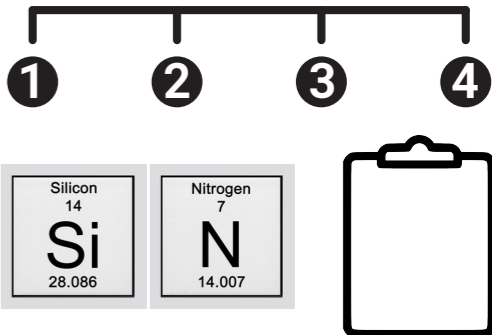
*instrumentation*

*human entry*

*electronic lab notebooks*



**preparation**





# data and meta-data

*structure, data types, quantity definitions*

**data formats**

**schemas**

**databases**

**specifications**

**documentation**

**data**

*raw data*

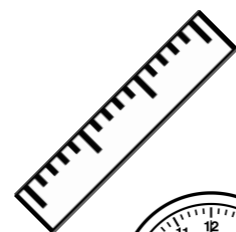
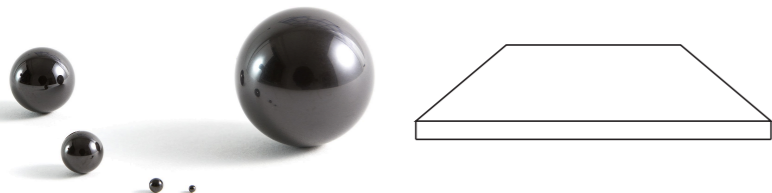
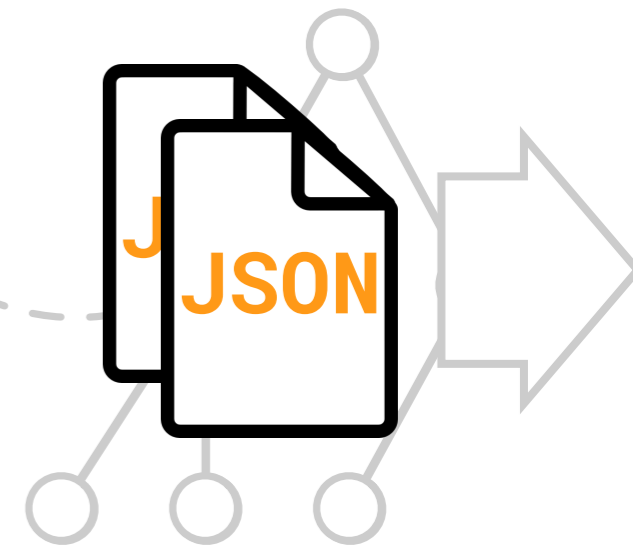
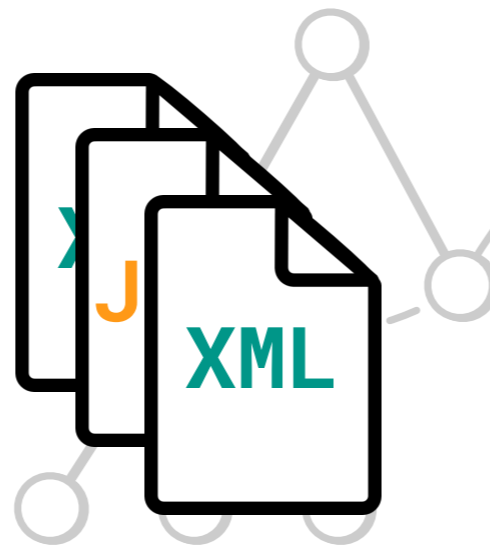
*processed data*

**meta-data**

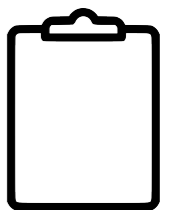
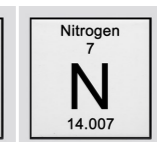
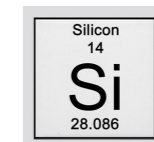
*instrumentation*

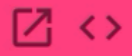
*human entry*

*electronic lab notebooks*



**preparation**







# run

section 


DESCRIPTION +


SECTIONS -

 frame sequence

 method

 sampling method

 single configuration calculation

 system

 topology

FEATURES -

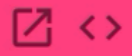
ALL (24)

OTHERS (3)

message\_debug (4)

parsing\_message\_debug (4)

accessory\_info (13)





# run

section 


DESCRIPTION +


SECTIONS -


 frame sequence

 method

 sampling method

 single configuration calculation

 system

 topology

FEATURES -

ALL (24) OTHERS (3)

message\_debug (4)

parsing\_message\_debug (4)

accessory\_info (13)

# data and meta-data

*structure, data types, quantity definitions*

**data formats**

**schemas**

**databases**

**specifications**

**documentation**

**data**

*raw data*

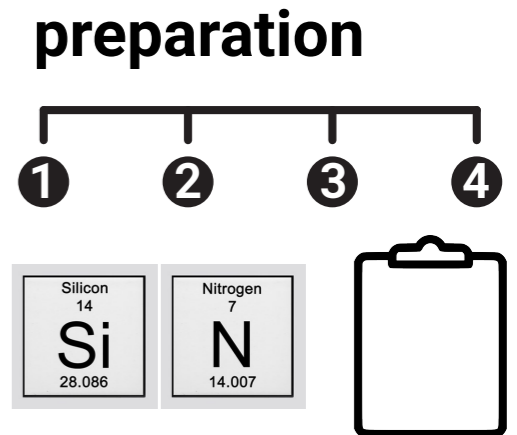
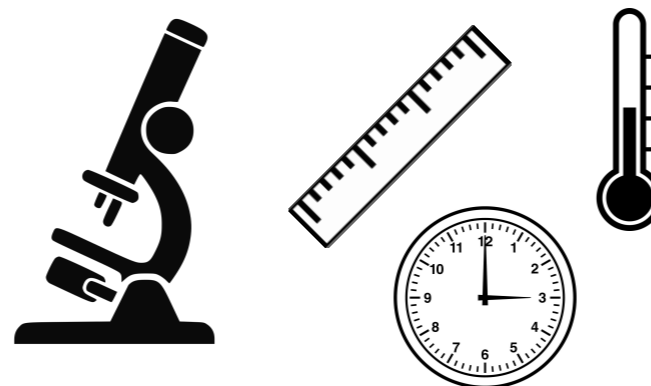
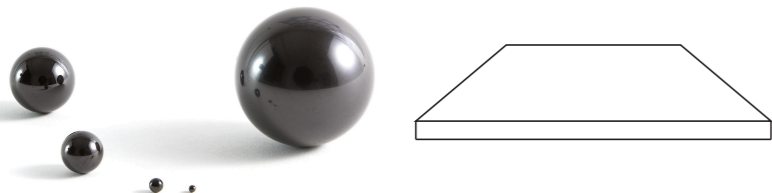
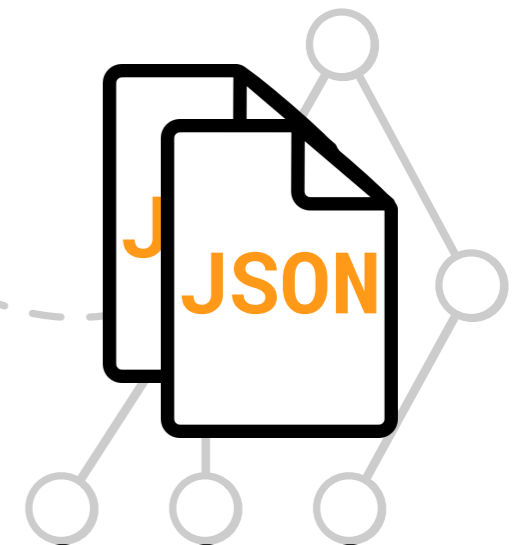
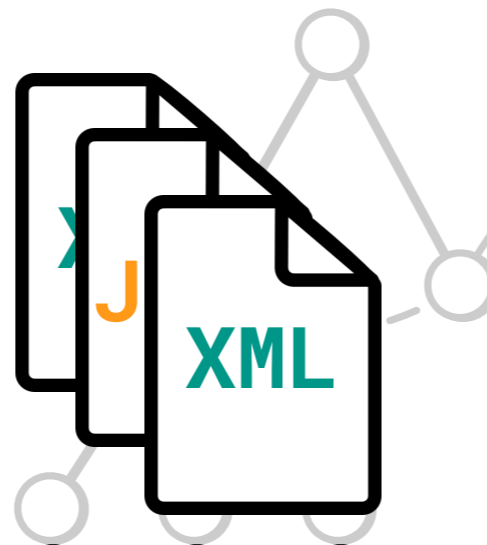
*processed data*

**meta-data**

*instrumentation*

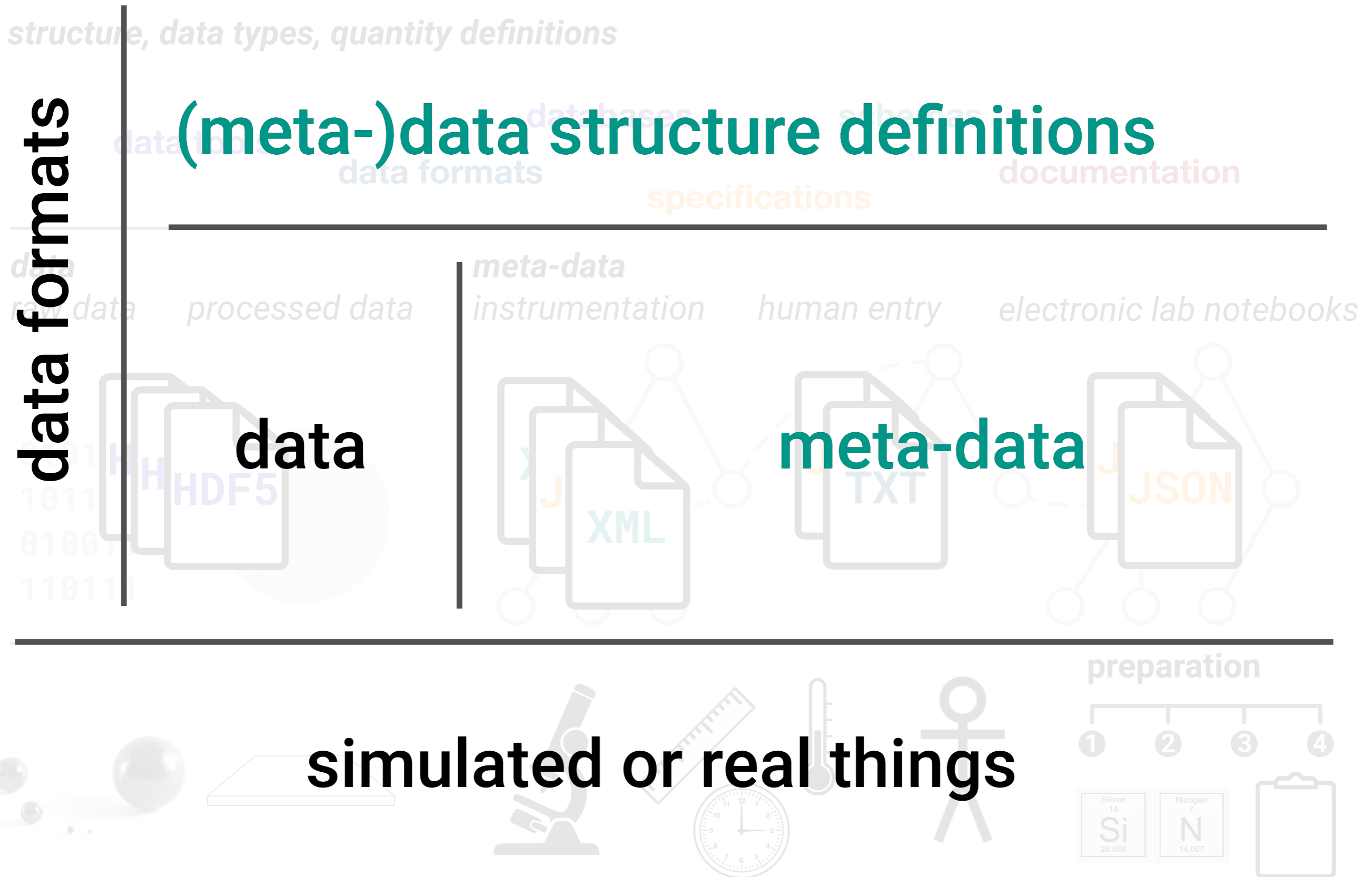
*human entry*

*electronic lab notebooks*

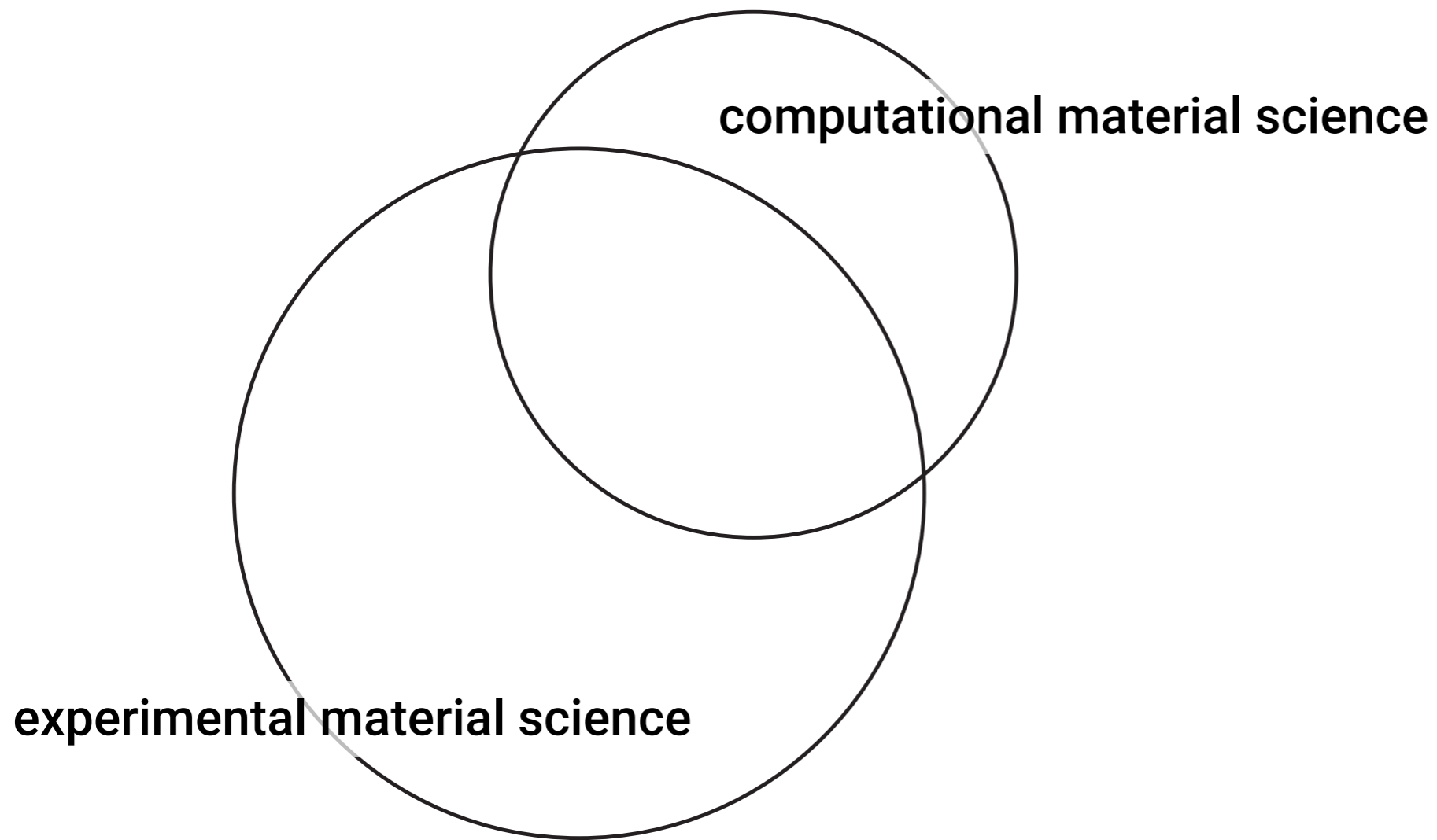




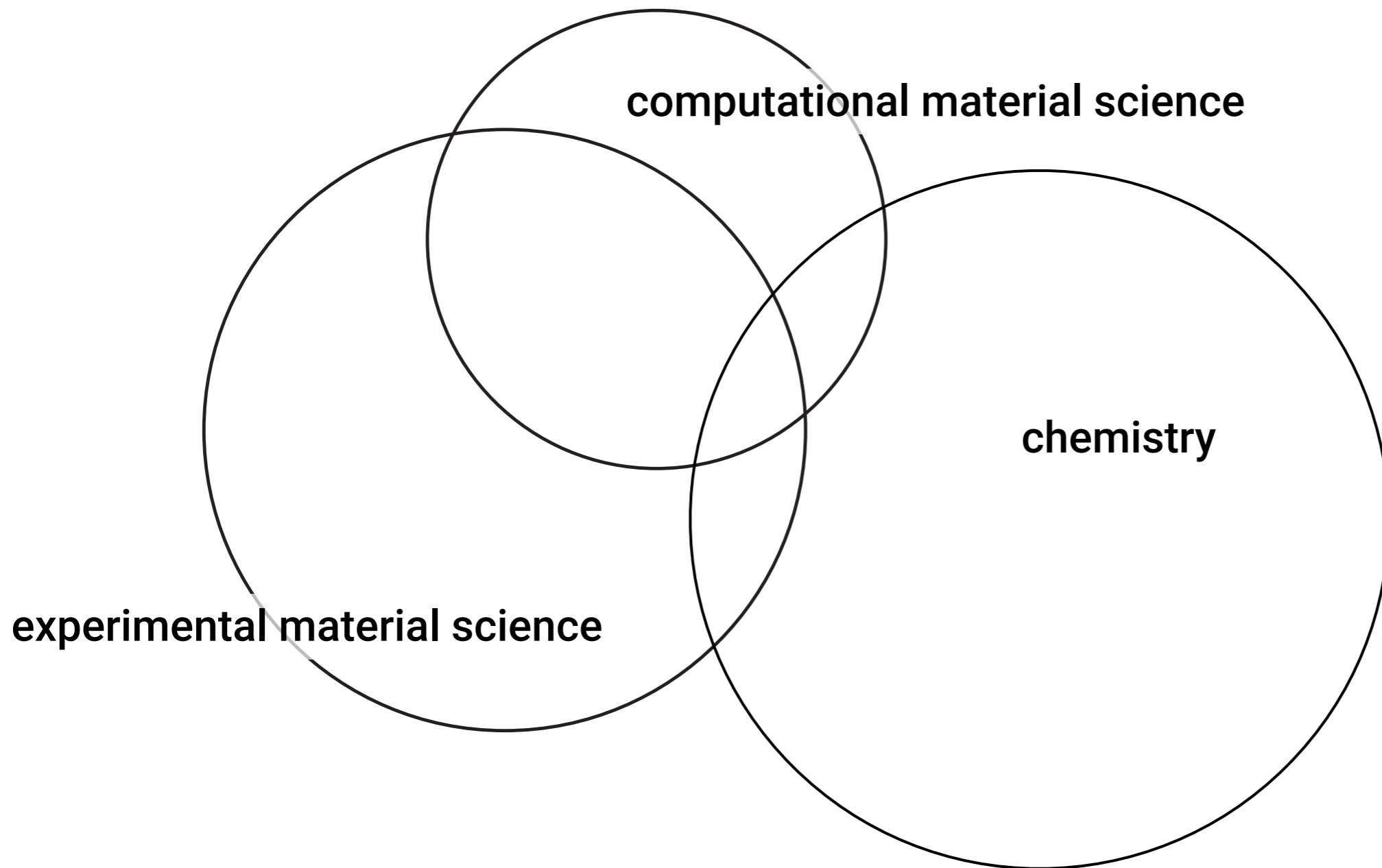
# data and meta-data



# domains

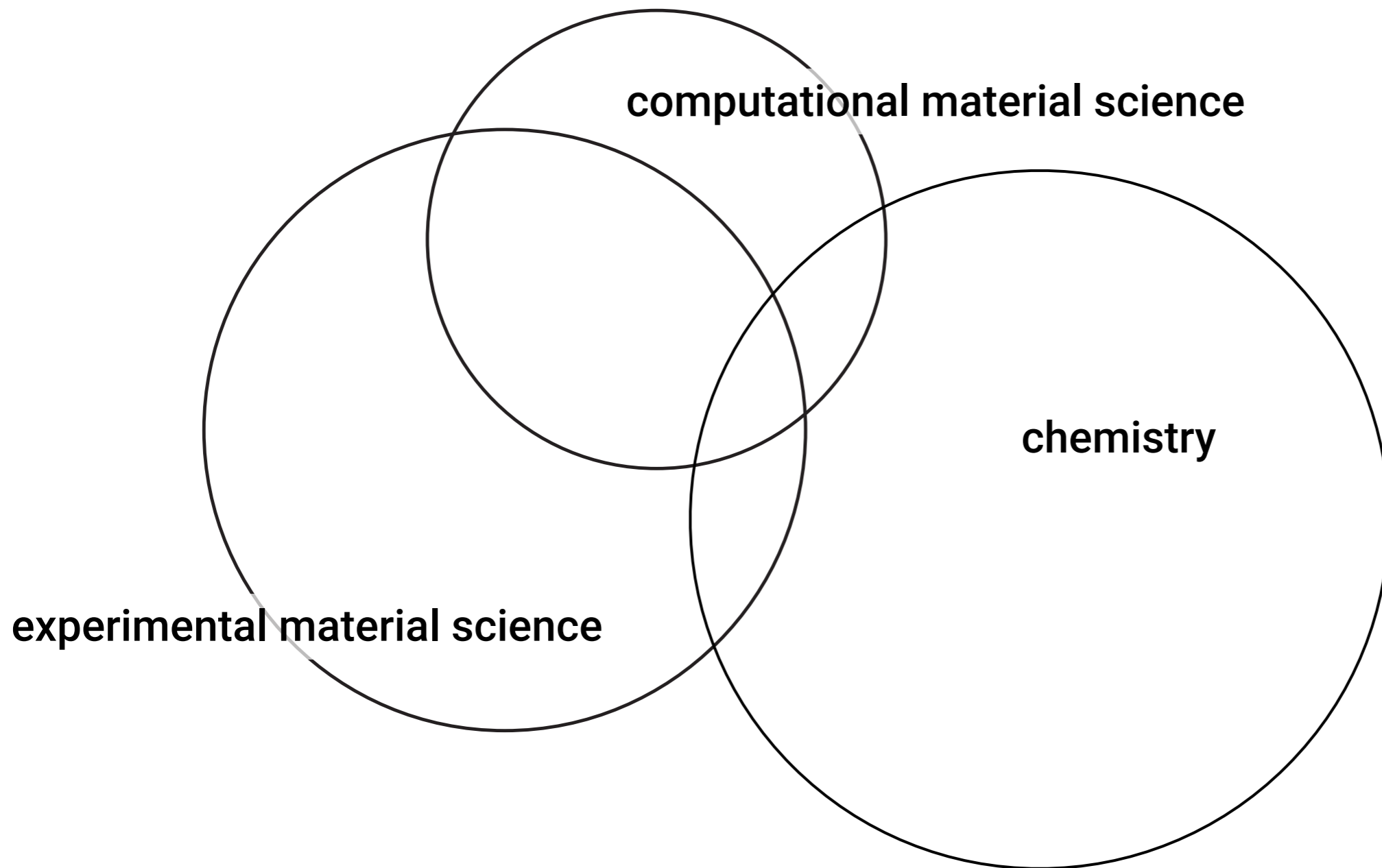


# domains

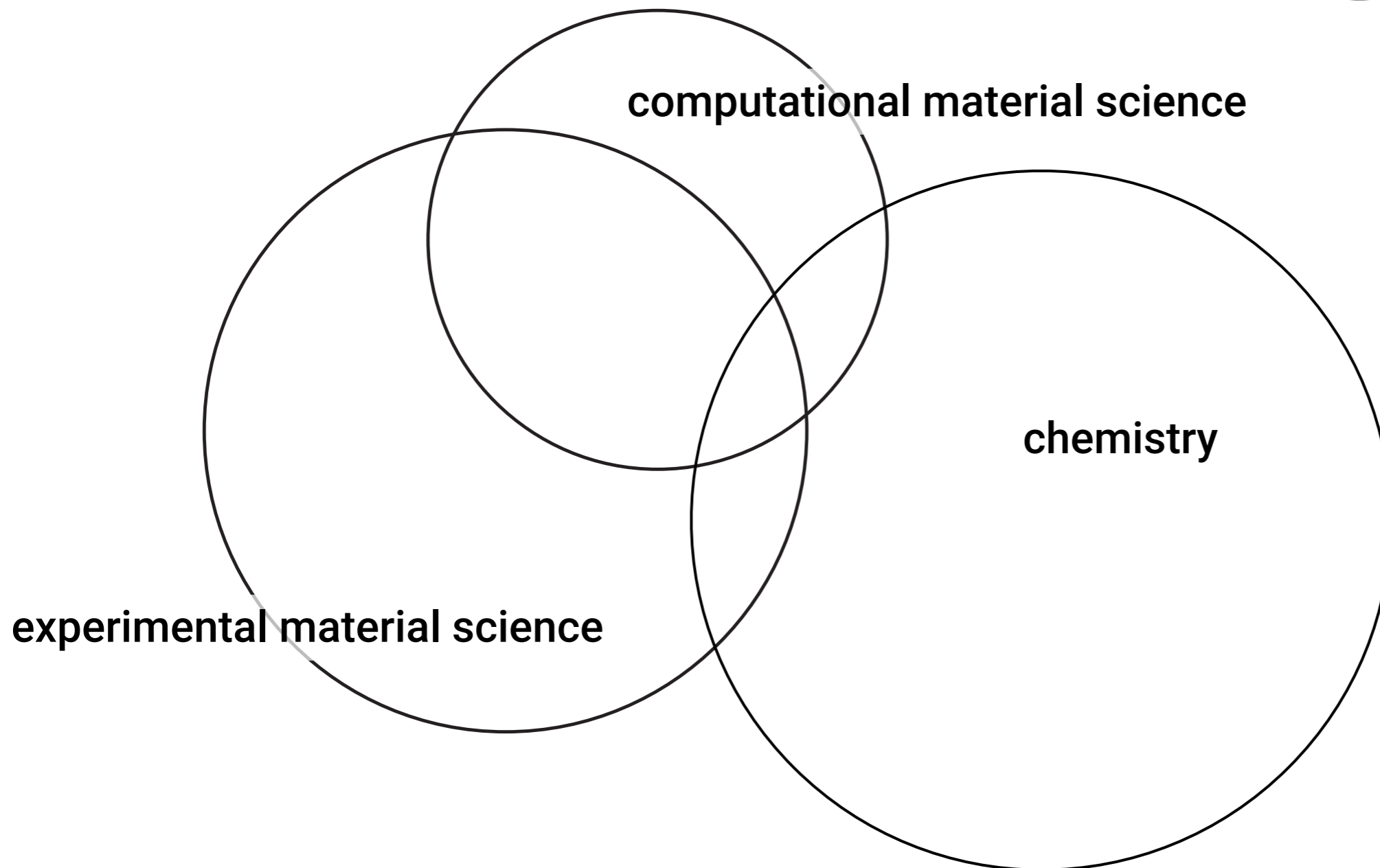
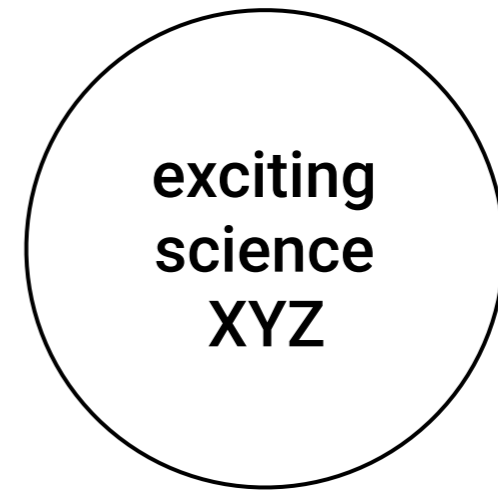




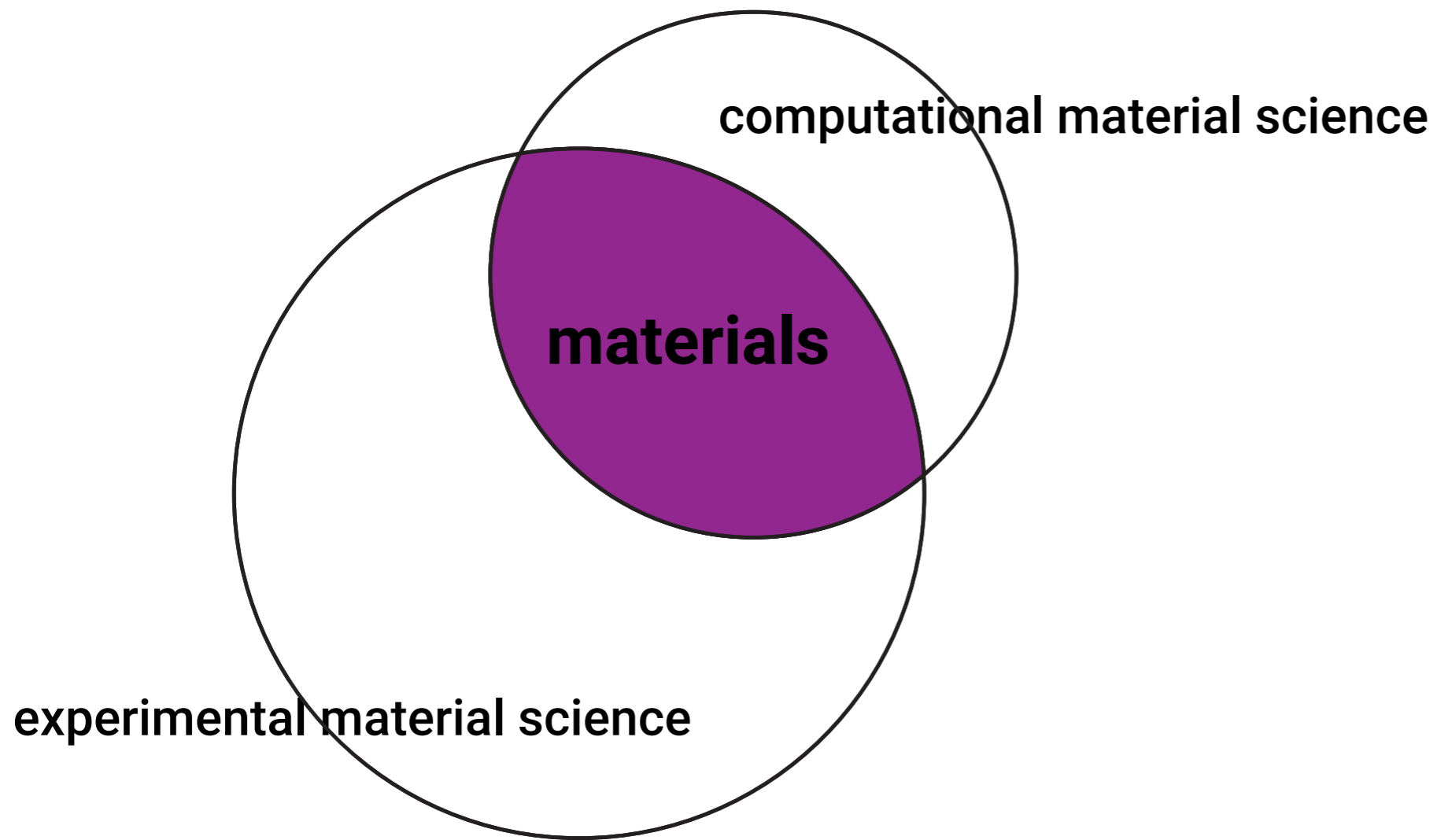
# domains



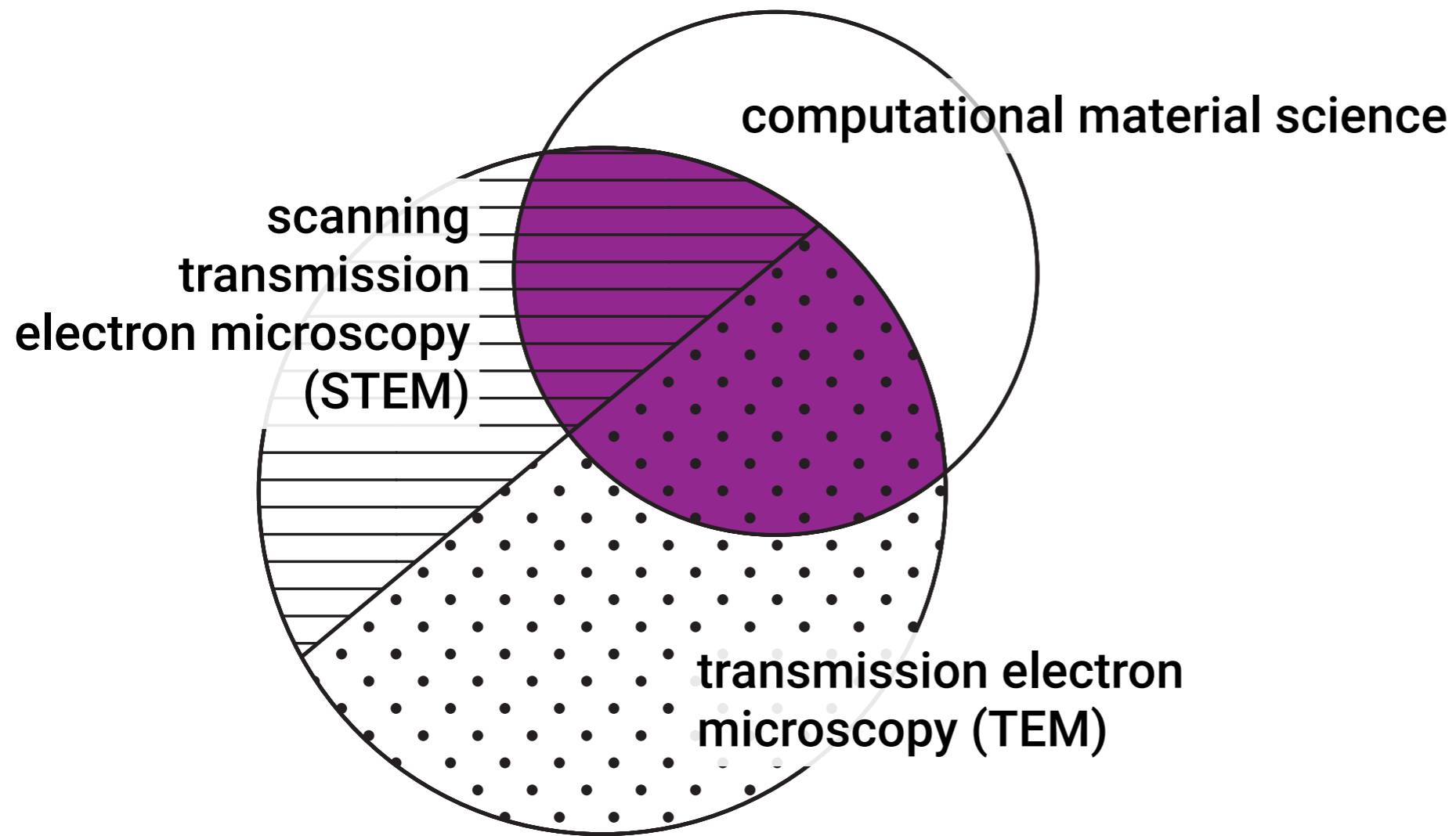
# domains



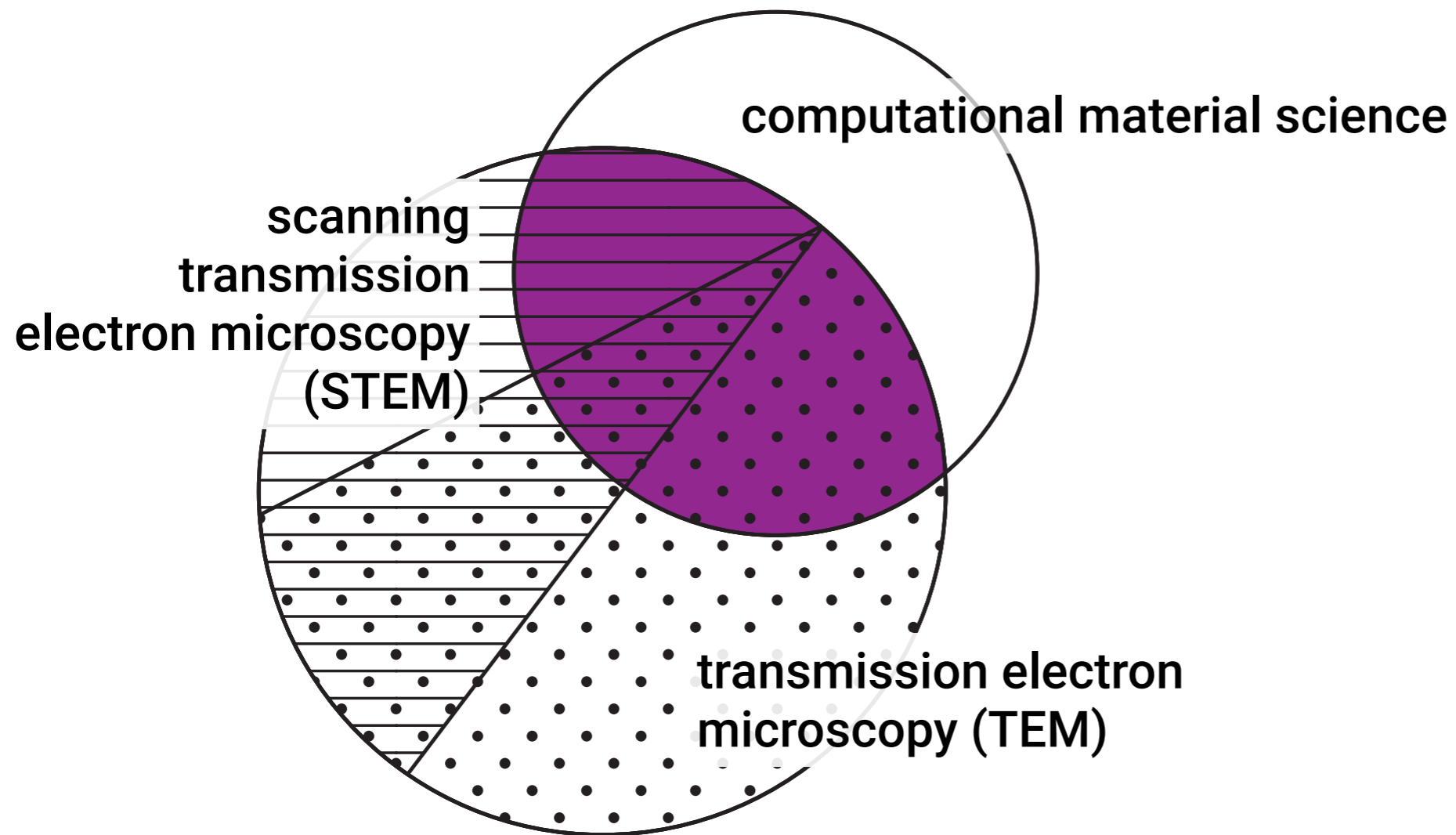
# domains



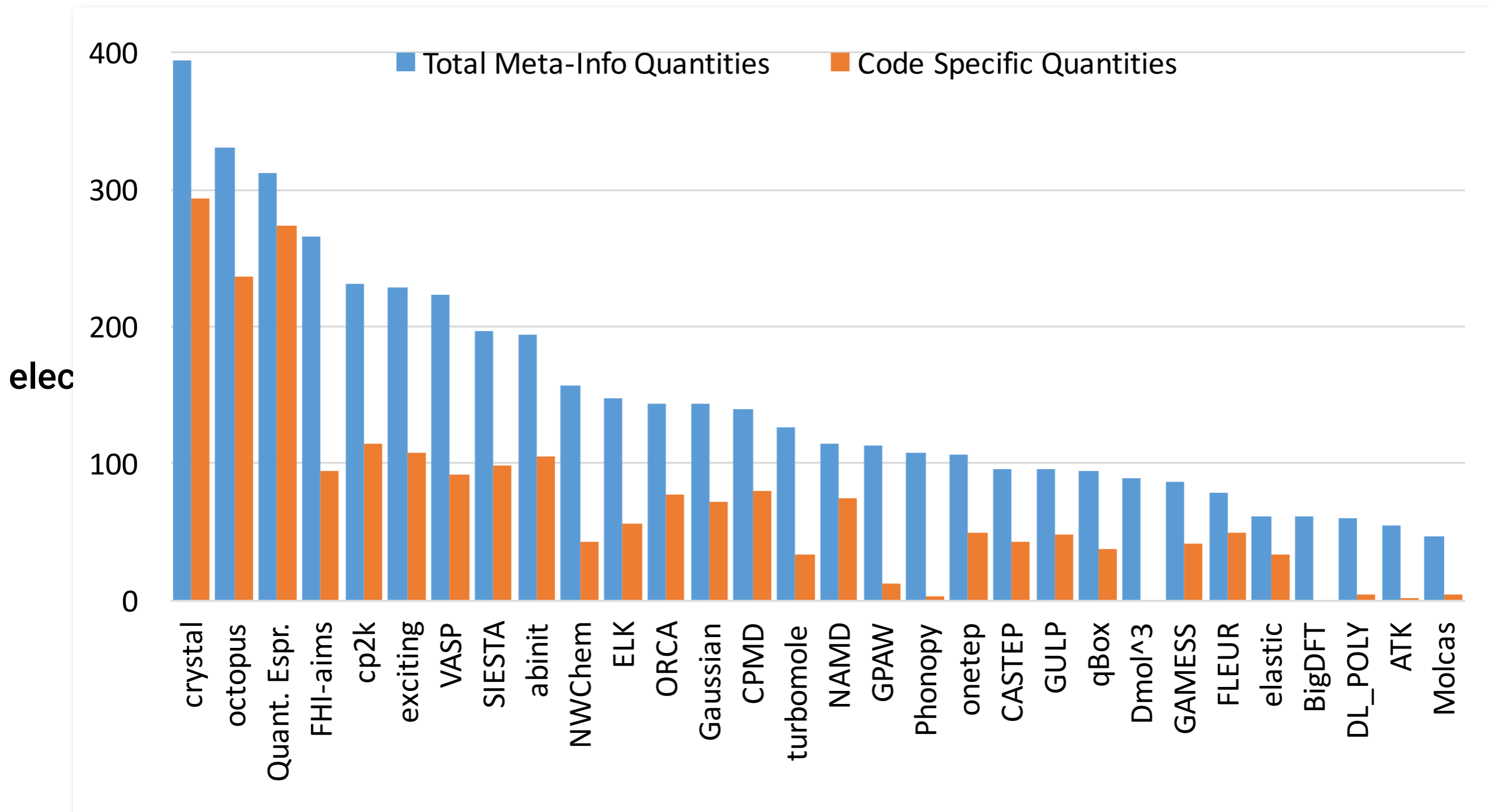
# domains



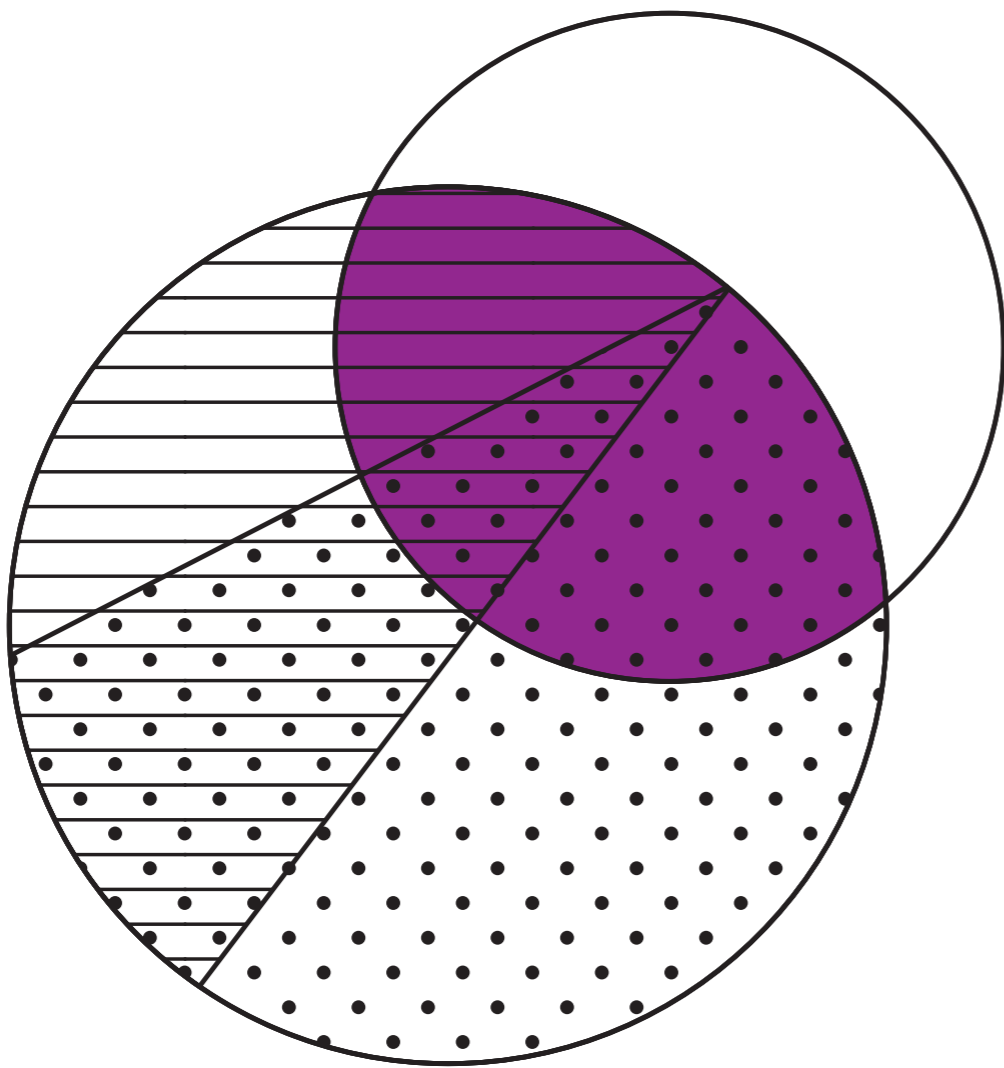
# domains



# domains

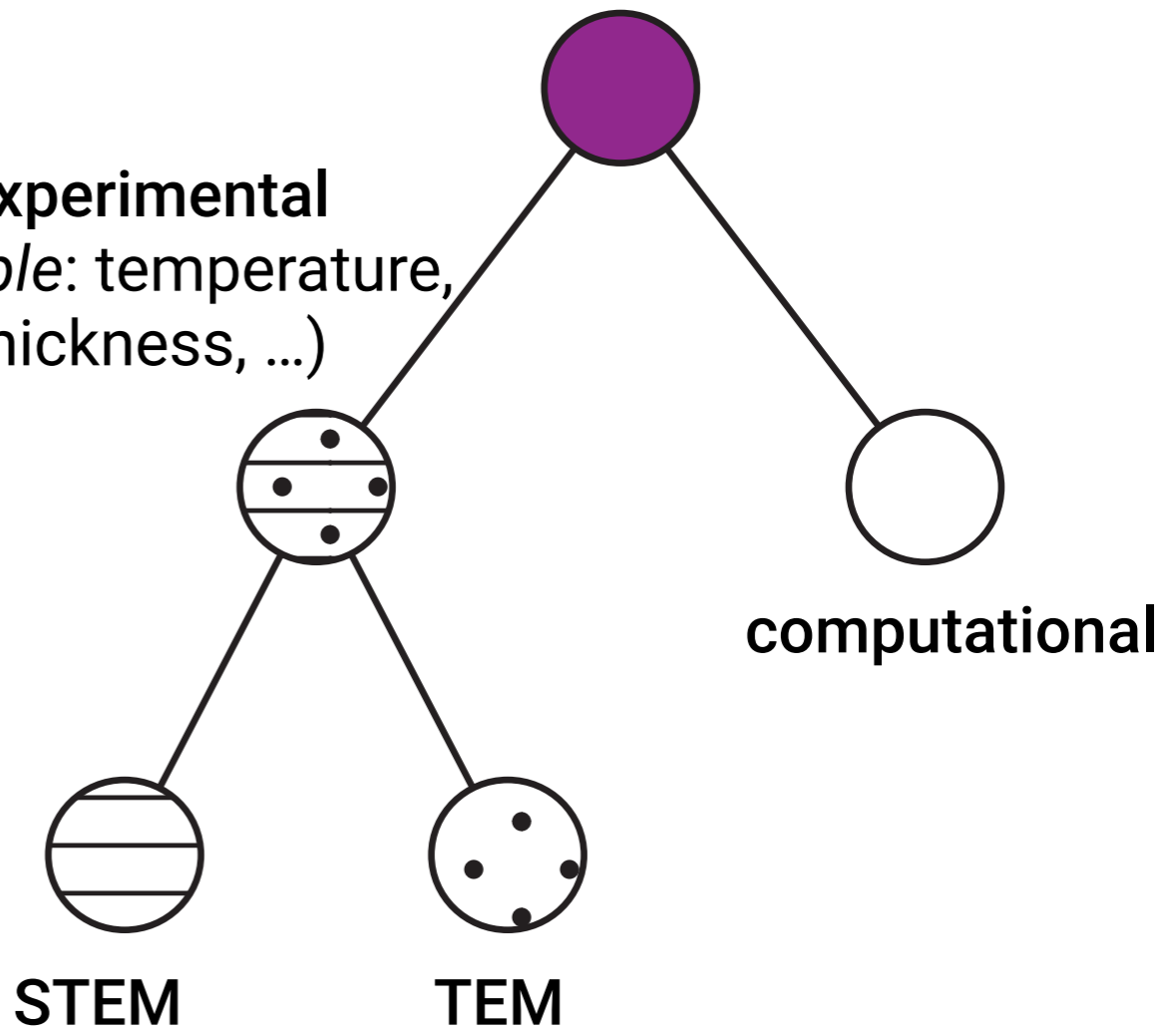


# domains



**material science**  
(*material*: atoms, symmetry, ...)

**experimental**  
(*sample*: temperature, thickness, ...)



# FAIR principles

- Findable
- Accessible
- Interoperable
- Re-usable



# FAIR principles

■ ~~Findable~~

**FILTERABLE**

■ Accessible

■ Interoperable

■ Re-usable

# FAIR principles

■ ~~Findable~~

**FILTERABLE**

■ Accessible

■ Interoperable

■ ~~Re-usable~~

**REPURPOSABLE**

# FAIR principles

■ ~~Findable~~

**FILTERABLE**

■ Accessible

■ Interoperable

■ ~~Re-usable~~

**~~REPURPOSABLE~~**

**RECYCLABLE**

# FAIR and *Domain Aware*

(meta-)data repositories



*domain agnostic*



Google  
Dataset Search Beta



# Google Dataset Search Beta


Nach Daten  suchen 

Probieren Sie [boston education data](#) or [weather site:noaa.gov](#)

[Weitere Informationen](#) über das Einfügen Ihrer Datensätze in der Datensatzsuche.



# Google Dataset Search Beta

Nach Daten  suchen 

Probieren Sie [boston education data](#) or [weather site:noaa.gov](#)

[Weitere Informationen](#) über das Einfügen Ihrer Datensätze in der Datensatzsuche.

# FAIR and *Domain Aware*

(meta-)data repositories

**FAIR**

*domain agnostic*



Google

Dataset Search Beta

# FAIR and *Domain Aware*

(meta-)data repositories

**FAIR**

*domain agnostic*



Google

Dataset Search Beta

(meta-)data formats

**IR**

*domain specific*



40 more



# FAIR and *Domain Aware*

(meta-)data repositories

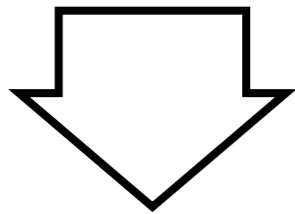
**FAIR**

*domain agnostic*



Google

Dataset Search Beta

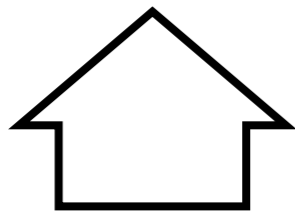


**FAIR**

*domain aware*



(meta-)data formats



**IR**

*domain specific*



40 more

Filter entries and show:

All entries  Only public entries  Only your entries  Only entries from your staging area

search

enter atoms, codes, functionals, or other quantity values



There are **4152027** entries.

About 4152027 results:

Formula	Code	Basis set	XT treatment	System	Crystal system	Spacegroup	Authors	References
Ac	VASP	plane waves	GGA	bulk	cubic	Fm-3m	Cooper, Sheldon	<i>no references</i>
Ac	VASP	plane waves	GGA	bulk	cubic	Fm-3m	Cooper, Sheldon	<i>no references</i>
Ac2	VASP	plane waves	GGA	bulk	hexagonal	P6_3/mmc	Cooper, Sheldon	<i>no references</i>
Ac2	VASP	plane waves	GGA	bulk	hexagonal	P6_3/mmc	Cooper, Sheldon	<i>no references</i>
Ac2Cl6	FHI-aims	numeric AOs	LDA	bulk	hexagonal	P6_3/m	Bieniek, Bjoern	<i>no references</i>
Ac2Cl6	FHI-aims	numeric AOs	GGA	bulk	hexagonal	P6_3/m	Bieniek, Bjoern	<i>no references</i>
Ac2Cl6	FHI-aims	numeric AOs	GGA	bulk	hexagonal	P6_3/m	Bieniek, Bjoern	<i>no references</i>

Filter entries and show:

All entries  Only public entries  Only your entries  Only entries from your staging area

search

enter atoms, codes, functionals, or other quantity values



There are **4152027** entries.

About 4152027 results:

Formula	Code	Basis set	XT treatment	System	Crystal system	Spacegroup	Authors	References
Ac	VASP	plane waves	GGA	bulk	cubic	Fm-3m	Cooper, Sheldon	<i>no references</i>
Ac	VASP	plane waves	GGA	bulk	cubic	Fm-3m	Cooper, Sheldon	<i>no references</i>
Ac2	VASP	plane waves	GGA	bulk	hexagonal	P6_3/mmc	Cooper, Sheldon	<i>no references</i>
Ac2	VASP	plane waves	GGA	bulk	hexagonal	P6_3/mmc	Cooper, Sheldon	<i>no references</i>
Ac2Cl6	FHI-aims	numeric AOs	LDA	bulk	hexagonal	P6_3/m	Bieniek, Bjoern	<i>no references</i>
Ac2Cl6	FHI-aims	numeric AOs	GGA	bulk	hexagonal	P6_3/m	Bieniek, Bjoern	<i>no references</i>
Ac2Cl6	FHI-aims	numeric AOs	GGA	bulk	hexagonal	P6_3/m	Bieniek, Bjoern	<i>no references</i>

## Metadata



### summary

Characterization of excited-state circular dichroism of WSe2

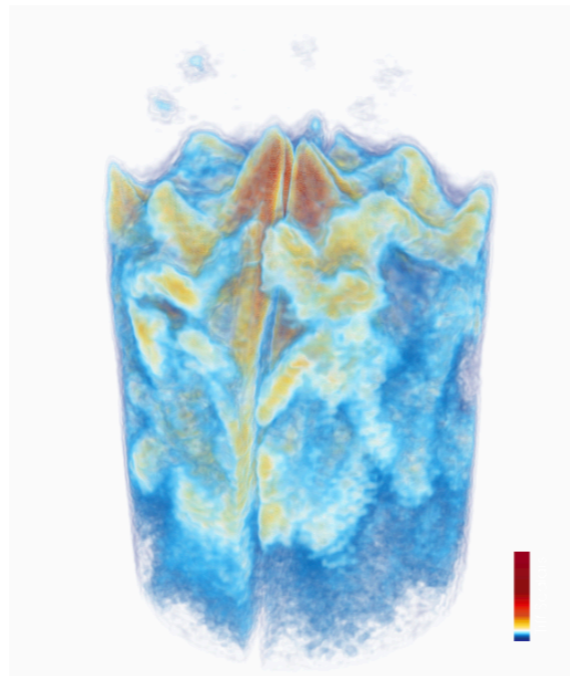
### sample formula

WSe2

### sample chemical

tungsten diselenide

### preview



### experimental method

multidimensional photoemission spectroscopy

### experiment location

Germany, Hamburg

### experiment time

01/04/2018, 02:00:00

### data

[zenodo.org](https://zenodo.org)

### comment

The first experimental data in nomad's experimental demonstrator

### references

<https://zenodo.org/record/2704788>

### authors

Xian, Patrick; Ernstorfer, Ralph

### datasets

*no datasets*

## Ids / processing

### PID

*not yet assigned*

### upload id

L-fbBWv4QcSrLCUIwh8unQ

### upload time

14/05/2019, 17:31:03

### experiment id

Uvras9A7BsoSZNbfYUiGHhh0iH9v

### mainfile

photoemission\_tungsten\_diselenide/201805\_WSe2\_LCPp...

### experiment hash

ZoD2dRhGTqkZM-2pX4XGq1mvRZ7h

### last processing

14/05/2019, 17:31:03

### processing version

0.4.4/df1721d05d13f9cab3efca2b5580e062e8745823

# FAIR and *Domain Aware*

(meta-)data repositories

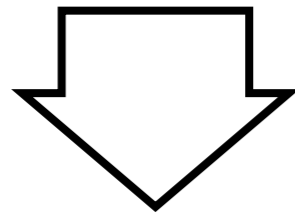
**FAIR**

*domain agnostic*



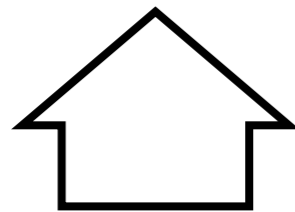
Google

Dataset Search Beta



**FAIR**

*domain aware*



**IR**

*domain specific*



(meta-)data formats

# FAIR and *Domain Aware*

(meta-)data repositories

**FAIR**

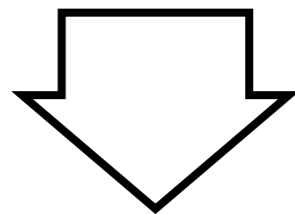
*domain agnostic*



Google

Dataset Search Beta

material science  
(*material*: atoms, symmetry, ...)



**FAIR**

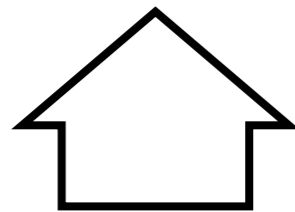
*domain aware*



experimental  
(*sample*: temperature, thickness, ...)

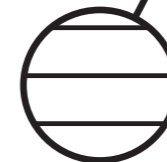


computational  
(code, functional)



**IR**

*domain specific*

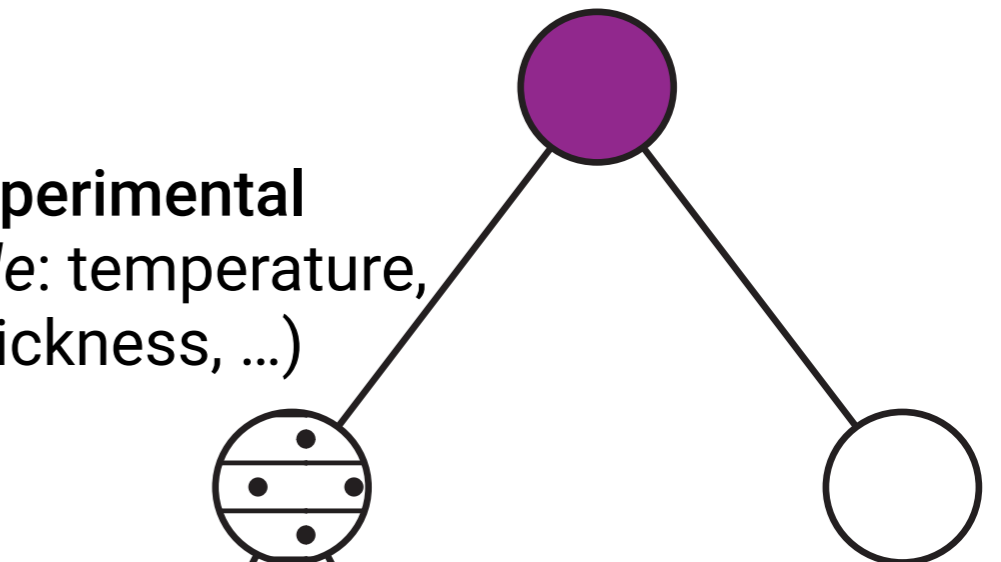


STEM



TEM

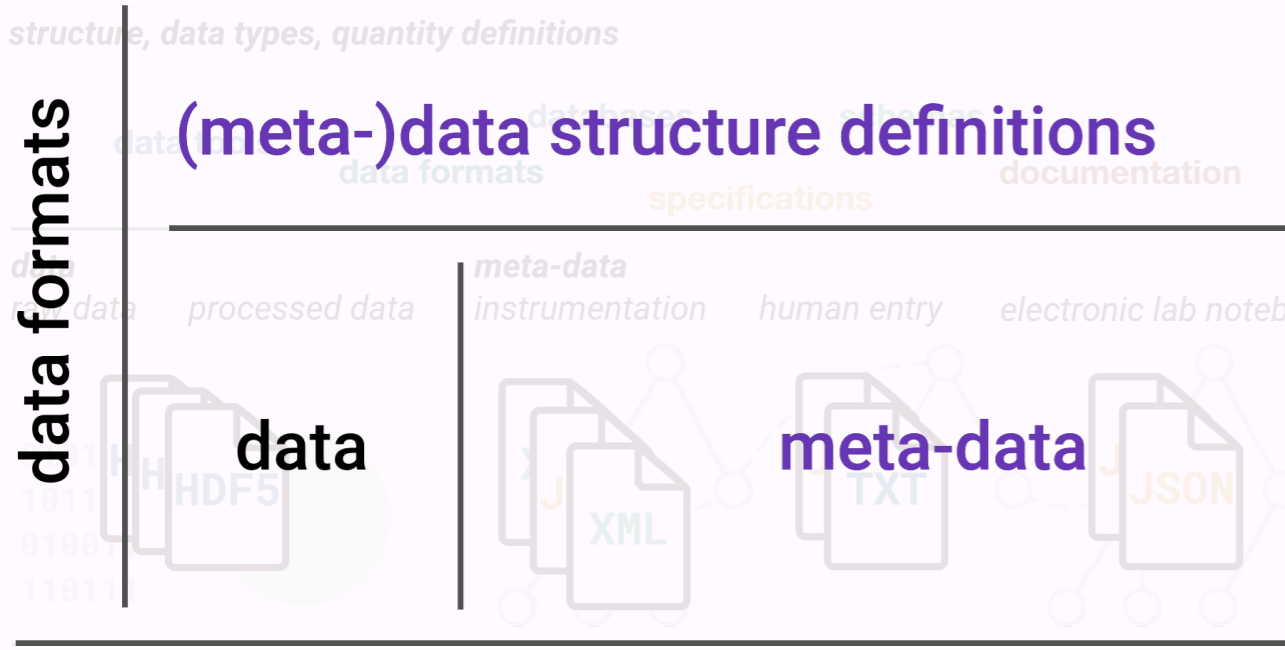
(meta-)data formats



# Summary

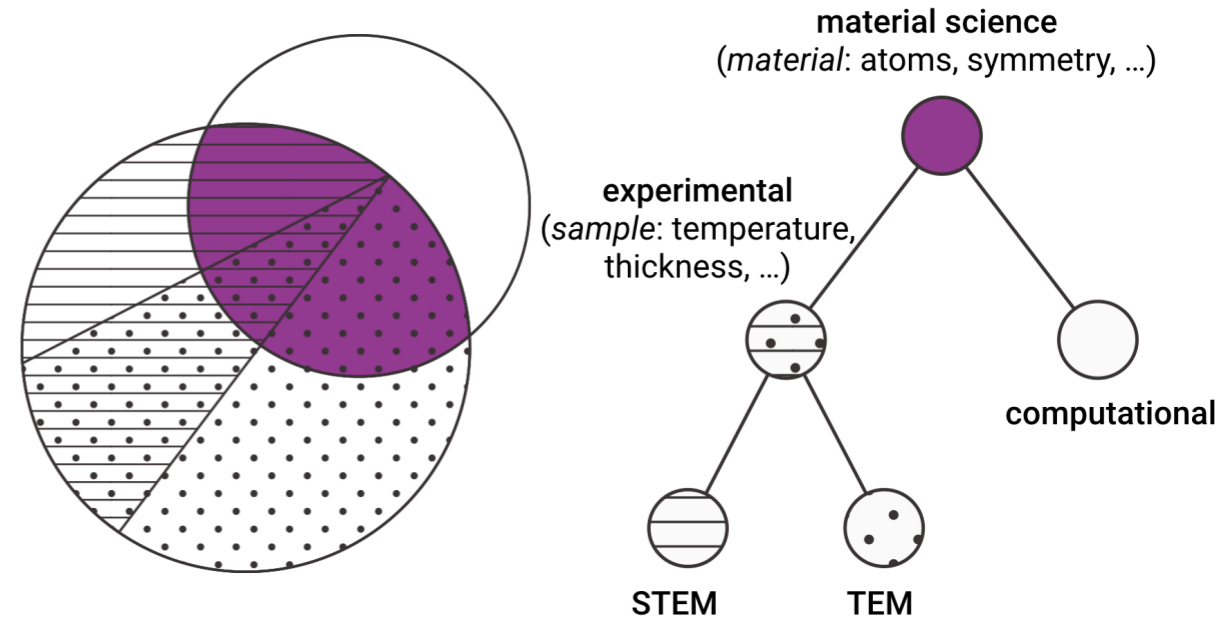
- meta-data paramount for data sharing
- homogenous domain meta-data for domain-aware **NOMAD** search
- integrating this meta-data from various material science sub-domains
- challenges:
  - agreeing on a schema system
  - maintaining suitable hierarchies of shared quantities

# data and meta-data



simulated or real things

# domains



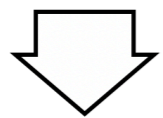
# FAIR and Domain Aware

(meta-)data repositories

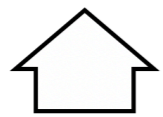
**FAI**  
domain agnostic



Google  
Dataset Search Beta



**FAIR**  
domain aware



**AIR**  
domain specific



(meta-)data formats

