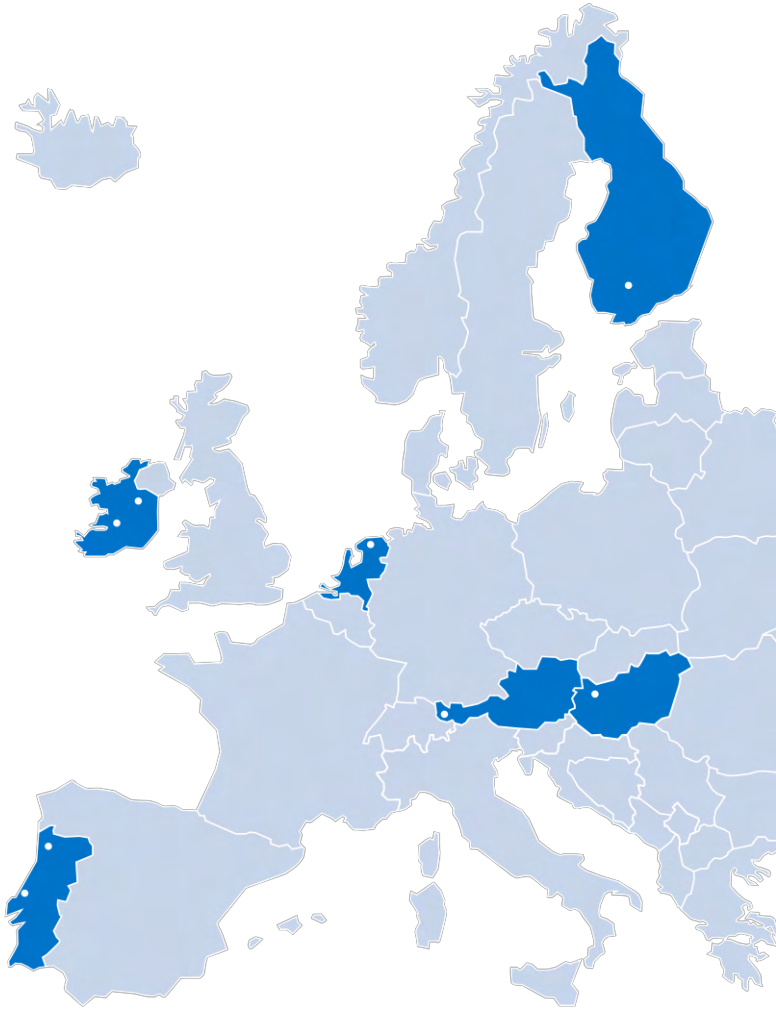
REGIONAL UNIVERSITY NETWORK PLUS

EUROPEAN UNIVERSITY



WELCOME TO THE WORKSHOP FAIR DATA

We start at 13:00 hours CEST

This workshop will be recorded

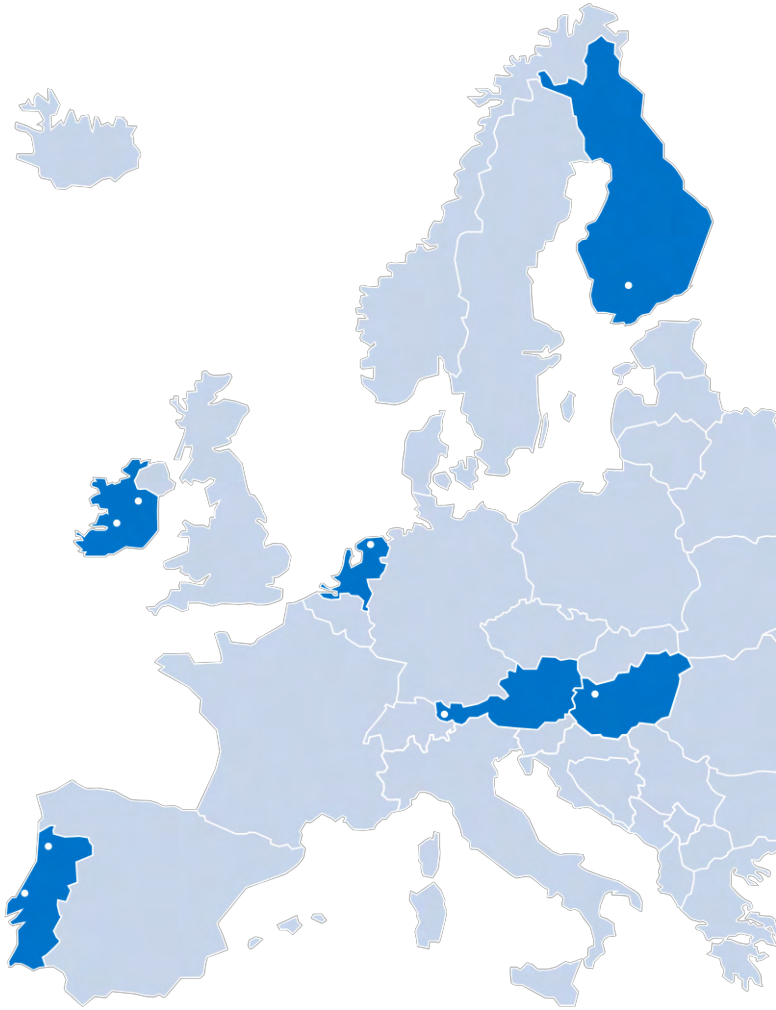
Co-funded by the Erasmus+ Programme of the European Union

The content of this publication represents the views of the author only and is his/her sole responsibility. The European Commission and the Agency do not accept any responsibility for use that may be made of yhe information it contains. Grant Agreement Number: 101035816

POLITÉCNICO DE LEIRIA · IPCA · TUS Technological University of the Shannon: Midlands Midwest · HAMK · SZÉCHENYI EGYETEM · NHL STENDEN university of applied sciences · FH Vorarlberg University of Applied Sciences

REGIONAL UNIVERSITY NETWORK PLUS

EUROPEAN UNIVERSITY



WORKSHOP FAIR DATA

2022, June 2nd

This workshop will be recorded

Co-funded by the Erasmus+ Programme of the European Union

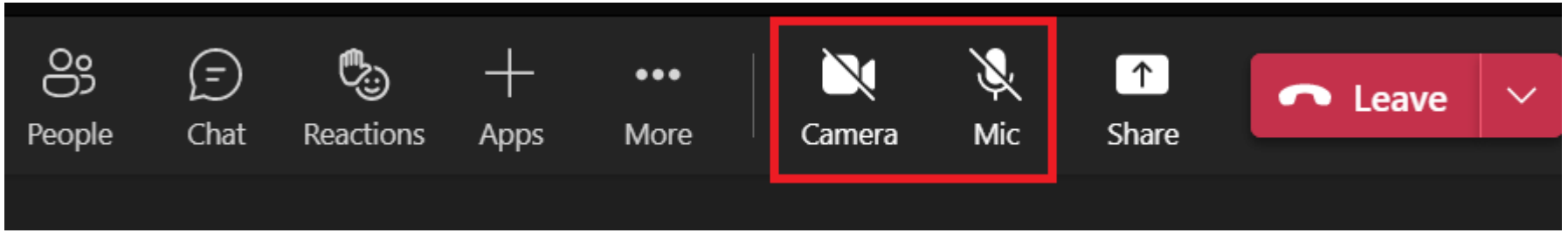
The content of this publication represents the views of the author only and is his/her sole responsibility. The European Commission and the Agency do not accept any responsibility for use that may be made of yhe information it contains. Grant Agreement Number: 101035816

POLITÉCNICO DE LEIRIA · IPCA INSTITUTO POLITÉCNICO DO CÁVADO E DO AVE · TUS Technological University of the Shannon: Midlands Midwest · HAMK · SZÉCHENYI EGYETEM · NHL STENDEN university of applied sciences · FH Vorarlberg University of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Please consider

- Camera off
- Microphone off



- Use chat if you have questions

- Technical problems: datasupport@nhlstenden.com

Mainstreaming Open Science Practices

REGIONAL UNIVERSITY NETWORK PLUS
EUROPEAN UNIVERSITY

CERTIFICATE OF ATTENDANCE

we hereby recognise

JELLIE VISSER

as an attendee of the
WORKSHOP ON FAIR DATA

2.JUNE.2022
13h00-17h00 CET

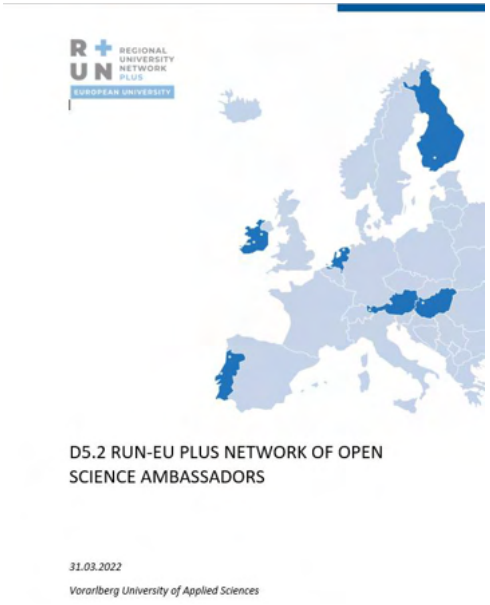
Dr Siobhán Moane
RUN-EU PLUS Project

Agenda

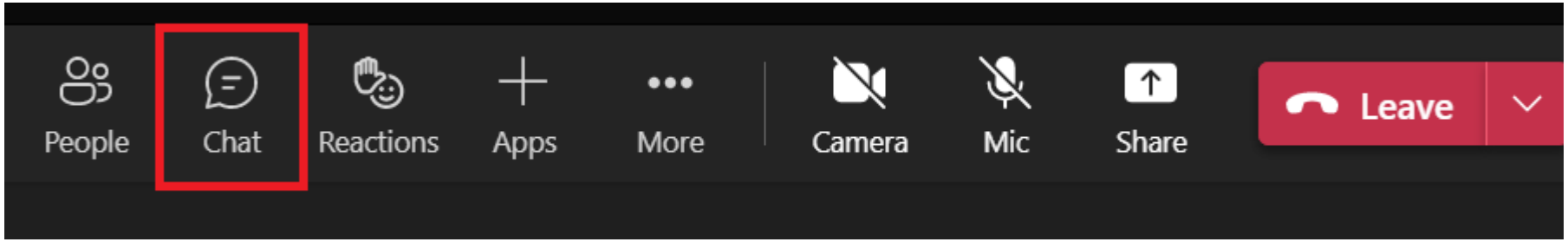
13.00 – 13:05	Welcome	Ingrid van Gorkum	NHL Stenden
13:05 – 14:45	Datamanagementplanning: what, why, how?		
	Group Exercise	Anna Mikkonen, Toni Pullianen, Nina Hynna	HAMK
14:45 – 15:00	BREAK		
15:00 – 16:30	FAIR data	Ingrid van Gorkum	NHL Stenden
16:30 – 17:00	Questions and Wrap UP	Ingrid van Gorkum, Jellie Visser	NHL Stenden

RUN-EU PLUS: WP 5 Mainstreaming Open Science Practices

- Audit map and GAP analysis in training and education on open science skills
- Establishment of the RUN-EU+ network of Open Science Ambassadors
- Development, design and implementation of workshop training programs for new practices of Open Science skills
- Annual report on the implementation



Feedback



In the chat you'll find a link

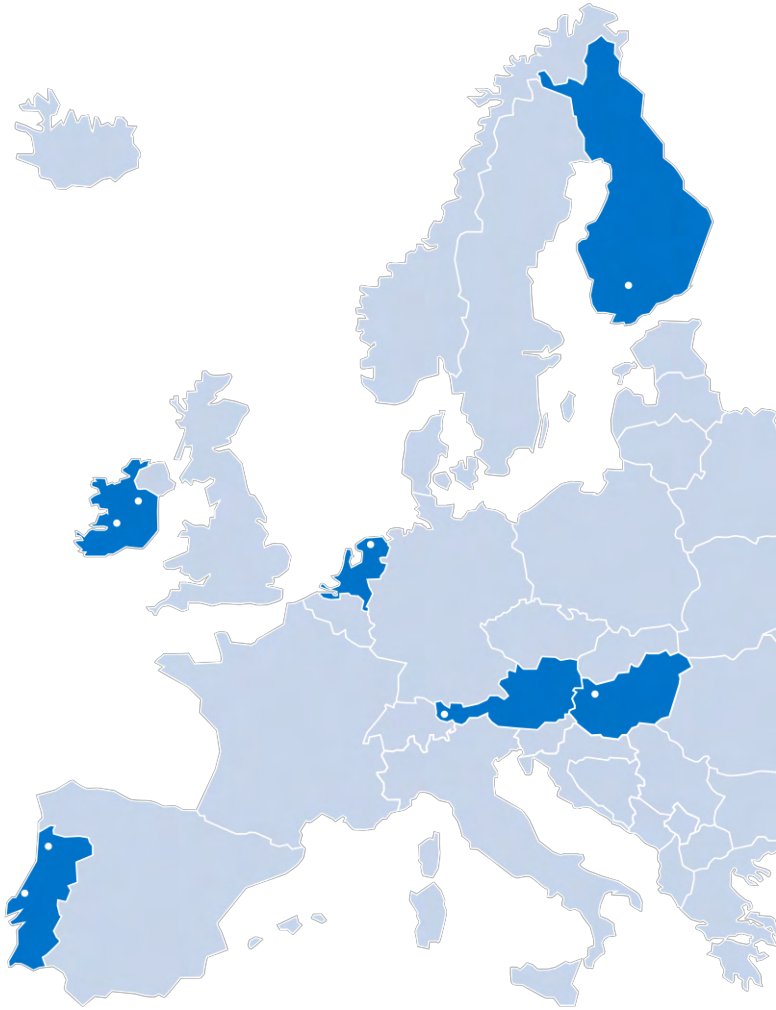
https://forms.office.com/r/iQmvnv2wM8

My level of knowledge of FAIR data is

- ○ Very low
- ○ Low
- ○ Moderate
- ○ High
- ○ Very high

REGIONAL UNIVERSITY NETWORK PLUS

EUROPEAN UNIVERSITY



RESEARCH DATA MANAGEMENT PLAN

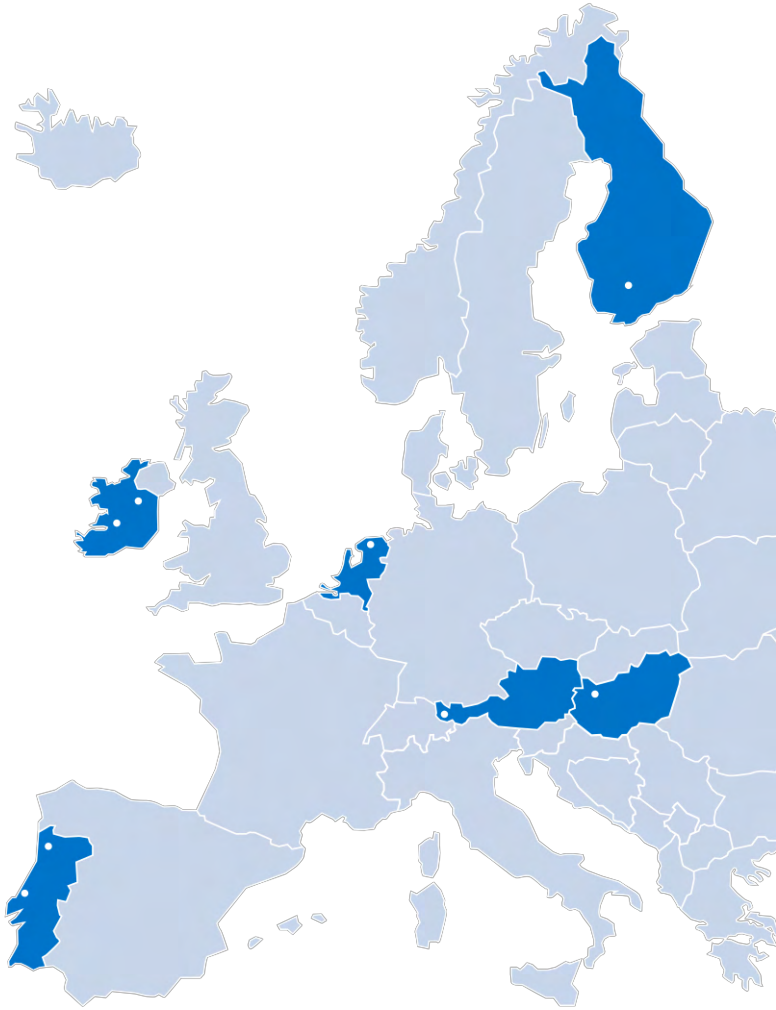
Co-funded by the Erasmus+ Programme of the European Union

The content of this publication represents the views of the author only and is his/her sole responsibility. The European Commission and the Agency do not accept any responsibility for use that may be made of yhe information it contains. Grant Agreement Number: 101035816

POLITÉCNICO DE LEIRIA · IPCA INSTITUTO POLITÉCNICO DO CÁVADO E DO AVE · TUS Technological University of the Shannon: Midlands Midwest · HAMK · SZÉCHENYI EGYETEM · NHL STENDEN university of applied sciences · FH Vorarlberg University of Applied Sciences

REGIONAL UNIVERSITY NETWORK PLUS

EUROPEAN UNIVERSITY



FAIR DATA

Co-funded by the Erasmus+ Programme of the European Union

The content of this publication represents the views of the author only and is his/her sole responsibility. The European Commission and the Agency do not accept any responsibility for use that may be made of yhe information it contains. Grant Agreement Number: 101035816

POLITÉCNICO DE LEIRIA · IPCA INSTITUTO POLITÉCNICO DO CÁVADO E DO AVE · TUS Technological University of the Shannon: Midlands Midwest · HAMK · SZÉCHENYI EGYETEM · NHL STENDEN university of applied sciences · FH Vorarlberg University of Applied Sciences

Open Science:

Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable **reuse, redistribution and reproduction of the research** and its underlying data and methods.

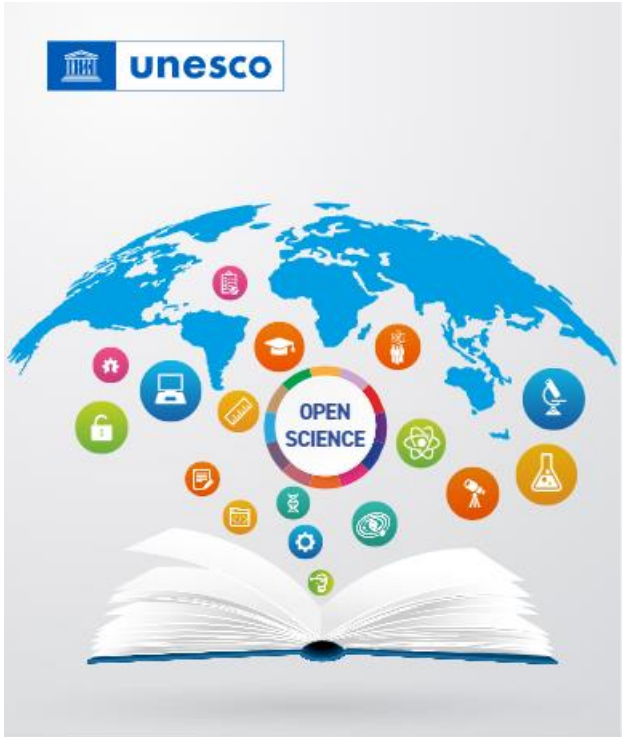
New funding

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

"Results paid for by public funds should be freely accessible worldwide. This applies to both scientific publications and other forms of scientific output. In principle, it must be possible to share the research data with others as well. In this way, valuable knowledge can be utilised by **researchers, businesses and civil society organisations**".

Faster knowledge development

- COVID
- Research is more visible
- Citizen science
  - Patient’s organisations



UNESCO Recommendation on Open Science

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Article 19

Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.

Article 27

1. Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits.
2. Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author.

United Nations

R U N REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Sustainable Development GOALS:

-reduce inequality within and among countries -no paywalls-

-quality education - everyone has access to the latest developments -

- knowledge development is fast – all goals have a scientific part

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Among others:

- Research institutions and organisations support proper infrastructure for the management and protection of data and research materials in all their forms (encompassing qualitative and quantitative data, protocols, processes, other research artefacts and associated metadata) that are necessary for reproducibility, traceability and accountability

- Research institutions and organisations reward open and reproducible practices in hiring and promotion of researchers.

Open Science:

Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable **reuse, redistribution and reproduction of the research** and its underlying data and methods.

FAIR data:

FAIR data are data which are

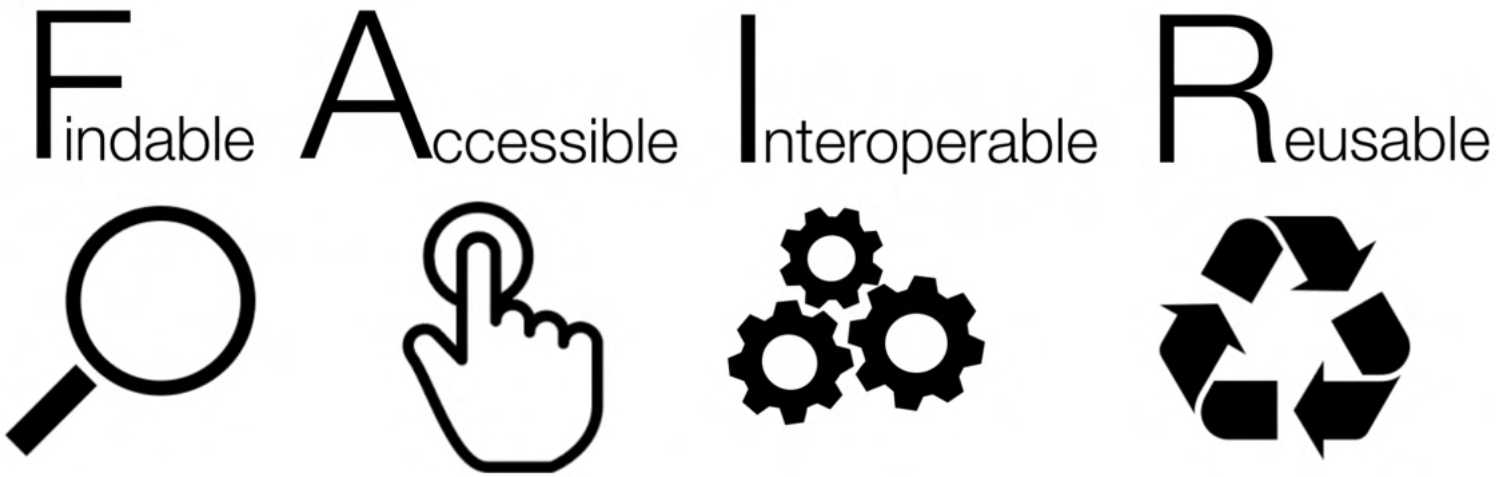


Image: CC-BY-SA by SangyaPundir

What is FAIR not?

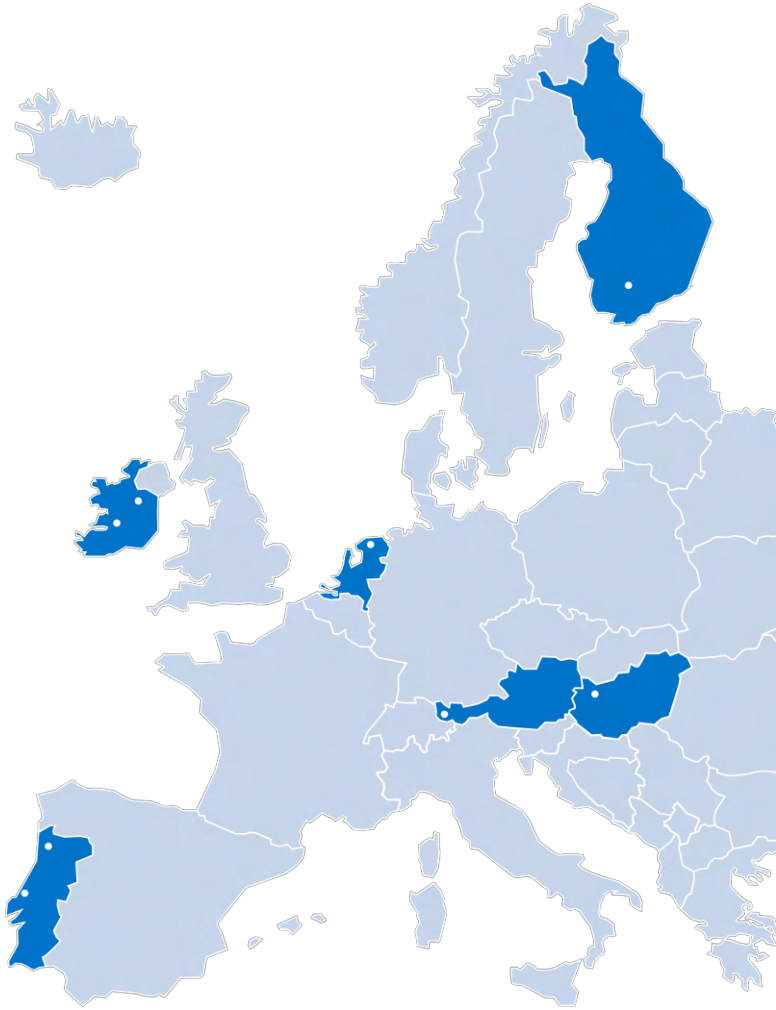
FAIR data is not a standard
FAIR is based on principles with the goal to reuse valuable research objects.

FAIR data are not always open, open data are not always FAIR
FAIR data are data which meet principles of findability, accessibility, interoperability, and reusability.

FAIR data is not just about humans being able to find, access, reformat and reuse data.
Computers should also be able to access a data publication, without help from human operators.

REGIONAL UNIVERSITY NETWORK PLUS

EUROPEAN UNIVERSITY



F - Findability

Co-funded by the Erasmus+ Programme of the European Union

The content of this publication represents the views of the author only and is his/her sole responsibility. The European Commission and the Agency do not accept any responsibility for use that may be made of yhe information it contains. Grant Agreement Number: 101035816

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

Persistent Identifiers

- A **persistent identifier** (PI or PID) is a long-lasting reference to a document, file, web page, or other object.

For example:

- Books — ISBN 978 81 212 6239 2
- Articles and datasets — DOI DOI 10.1108/JTF-04-2017-0022
- Persons — ORCID 0000-0001-9189-9434

Go to https://orcid.org/ , search for John Smith.

Which numbers do you recognise as persistent identifiers, when you check one of the John Smiths? (clicking on employment and show more detail.

Of course, checking your own name (or that of one your collegues) is allowed

Showing **50** of **88559** results.

Items per page: 50 1 – 50 of 88559

ORCID ID	First Name	Last Name	Other Names
0000-0003-1660-3511	John	Smith	
0000-0002-9315-0678	John	Smith	
0000-0002-4216-1107	John	Smith	
0000-0001-5376-4926	John	Smith	
0000-0001-7793-0079	John	Smith	

Irish Management Institute: Dublin, Dublin, IE

1998 to 2000 | Certificate in Supervisory Management
Education

Organization identifiers
GRID: grid.435607.3
Irish Management Institute: Dublin, IE
http://www.imi.ie/

Other organization identifiers provided by GRID
ISNI: 0000000404888940
ORGREF: 5757917
ROR: https://ror.org/011hy5f81
WIKIDATA: Q6070946
WIKIPEDIA_URL: https://en.wikipedia.org/wiki/Irish_Management_Institute (preferred)

Added
2020-06-24

Last modified
2020-06-24

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

Standardised metadata

- Does a (future) user or computer have sufficient information to be able to find the data and understand what the dataset entails?

A	B	C	D	E	F
person	postal code	ducks counted	doves counted	oystercatchers counted	pinguins counted
a	9717 DD	2	3	4	0
b	**1012 JS**	0	87	0	0
c	2513 AA	10	2	0	0
d	8926 XE	8	2	0	25

Standardised metadata

- Does a (future) user or computer have sufficient information to be able to find the data and understand what the dataset entails?

- Metadata are:
  - 1. Standardised
  - 2. Structured
  - 3. Machine and human readable
  - 4. They are a subset of documentation

Standardised metadata : 4 types

Look for the kinds of data at https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:230140;

- Descriptive metadata
  - Author, Title, Abstract, Date, Contextual metadata as methods of data collection
- Structural metadata
  - Links to related objects (for example the article written or reused datasets)
- Technical metadata
  - Software used, file formats
- Administrative metadata
  - User licenses, embargo, etc

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

Repositories: What are repositories

- Repositories is a large database used to store, document and publish all kinds of digital objects.
- Data repositories can help make a researcher's data more discoverable and accessible, and lead to potential reuse. Using a repository can lead to increased citations of your work
- Depending on the discipline requirements - publisher, funders, institutional policies, national policies - researchers may be required to store their data in certain repositories.

Repositories : types of repositories

- Domain-specific
  - ICPSR - Social Sciences
  - The CESSDA archives – Social Sciences
  - PANGAEA - Earth and space science data
  - Crystallography Open Database (COD) - Chemistry & Crystallography
- Institutional repositories/National repositories
- Cross-discipline repositories
  - Zenodo
  - Figshare

Repositories: which one to choose?

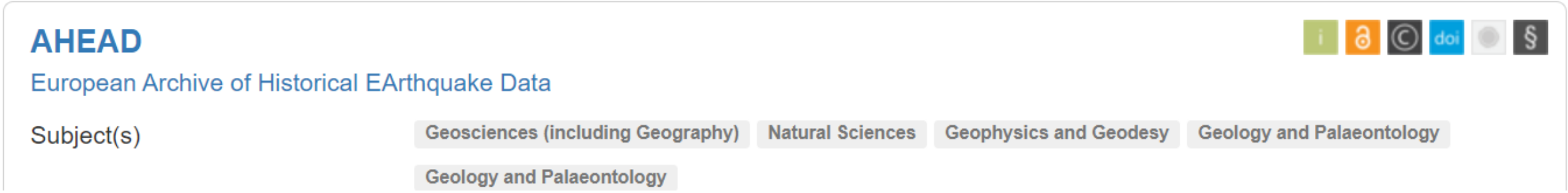
- Look at your organisation policy!
- Is it certified? Core Trust Seal
- Is ideally internationally recognised, commonly used and endorsed by the respective community.
- Matches your particular data needs (e.g. formats accepted; access, back-up and recovery, and sustainability of the service).
- Offers clear terms and conditions that meet legal requirements (e.g. for data protection) and allow reuse without unnecessary licensing conditions (e.g. restricted vs open).
- What are the costs?

Assignment

- Look at https://www.re3data.org/

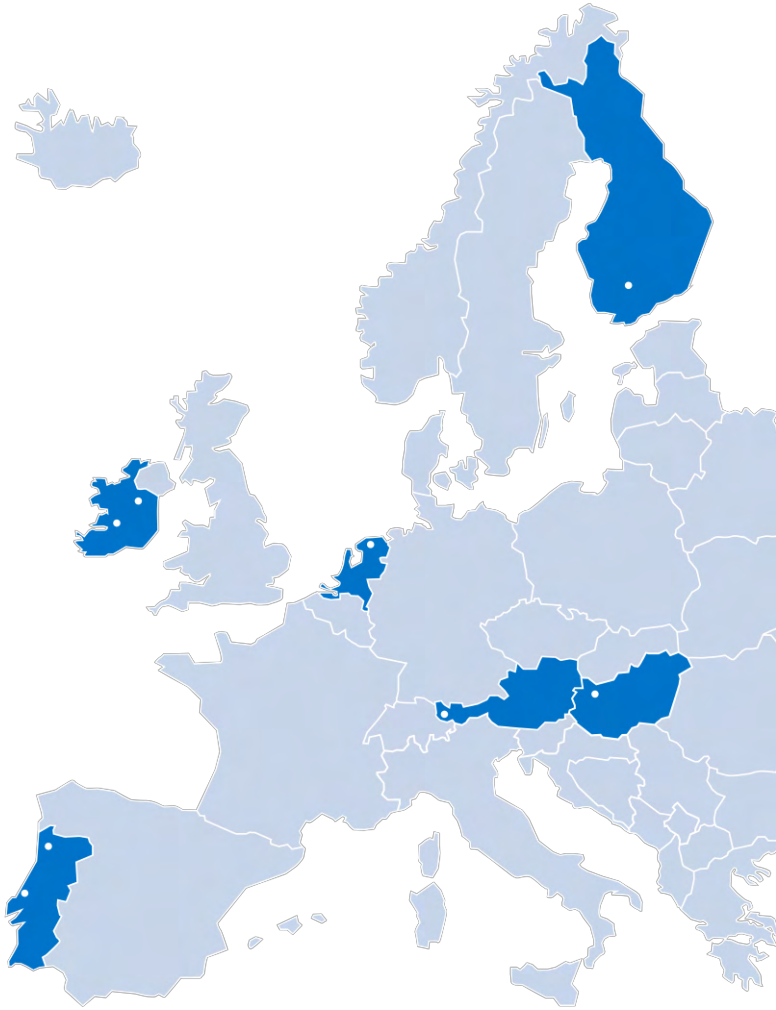
- Find the repositories mentioned in your country or your research field

- Do they provide DOI's?
- Is it certified or does it support repository standards?



REGIONAL UNIVERSITY NETWORK PLUS

EUROPEAN UNIVERSITY



A - Accessiblity

Co-funded by the Erasmus+ Programme of the European Union

The content of this publication represents the views of the author only and is his/her sole responsibility. The European Commission and the Agency do not accept any responsibility for use that may be made of yhe information it contains. Grant Agreement Number: 101035816

Accessible

It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

- ☐ Following the persistent ID will take you to the data or associated metadata
- ☐ The protocol by which data can be retrieved follows recognised standards e.g. http
- ☐ The access procedure includes authentication and authorisation steps, if necessary
- ☐ Metadata are accessible, wherever possible, even if the data aren't

'How FAIR are your data?' checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

Accessible

It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

- ☐ Following the persistent ID will take you to the data or associated metadata
- ☐ The protocol by which data can be retrieved follows recognised standards e.g. http
- ☐ The access procedure includes authentication and authorisation steps, if necessary
- ☐ Metadata are accessible, wherever possible, even if the data aren't

'How FAIR are your data?' checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

Access: categories

Data access categories can be

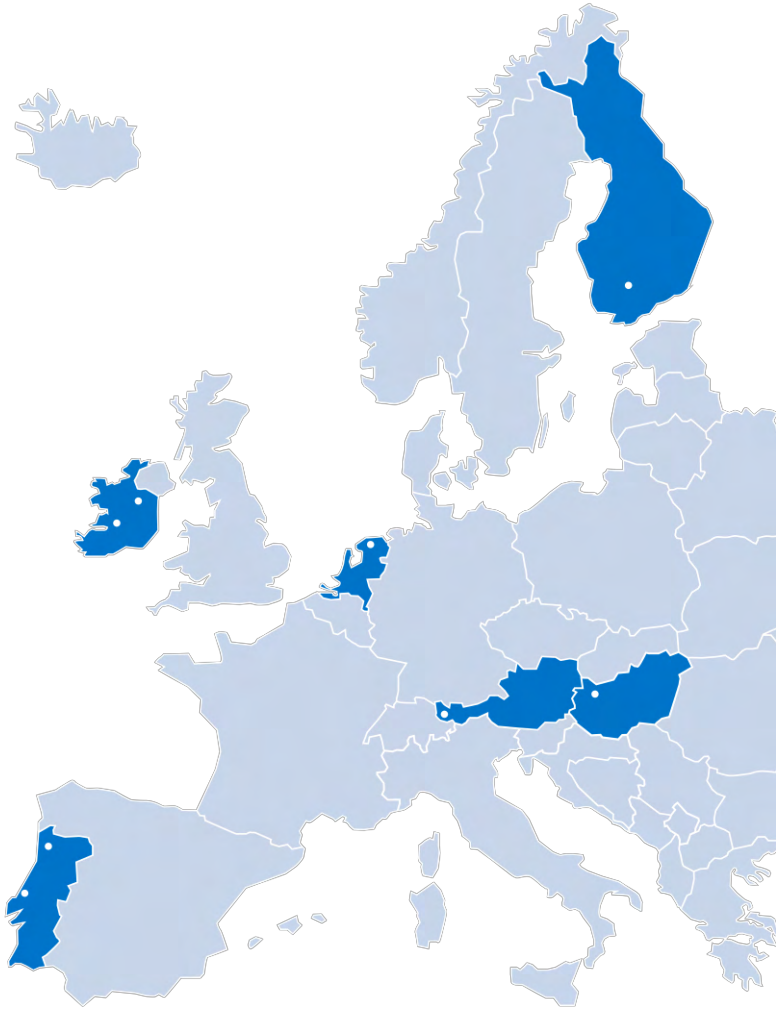
- Open access
- Restricted access : https://www.lifelines.nl/researcher/how-to-apply
- Embargo
- Closed access

Because of sensitive information, intellectual property, personal information and confidentiality agreements

BREAK

REGIONAL UNIVERSITY NETWORK PLUS

EUROPEAN UNIVERSITY



I - Interoperability

Co-funded by the Erasmus+ Programme of the European Union

The content of this publication represents the views of the author only and is his/her sole responsibility. The European Commission and the Agency do not accept any responsibility for use that may be made of yhe information it contains.
Grant Agreement Number: 101035816

POLITÉCNICO DE LEIRIA · IPCA INSTITUTO POLITÉCNICO DO CÁVADO E DO AVE · TUS Technological University of the Shannon: Midlands Midwest · HAMK HÄMEEN AMMATTIKORKEAKOULU · SZÉCHENYI EGYETEM UNIVERSITY OF GYŐR · NHL STENDEN university of applied sciences · FH Vorarlberg University of Applied Sciences

Interoperable

Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

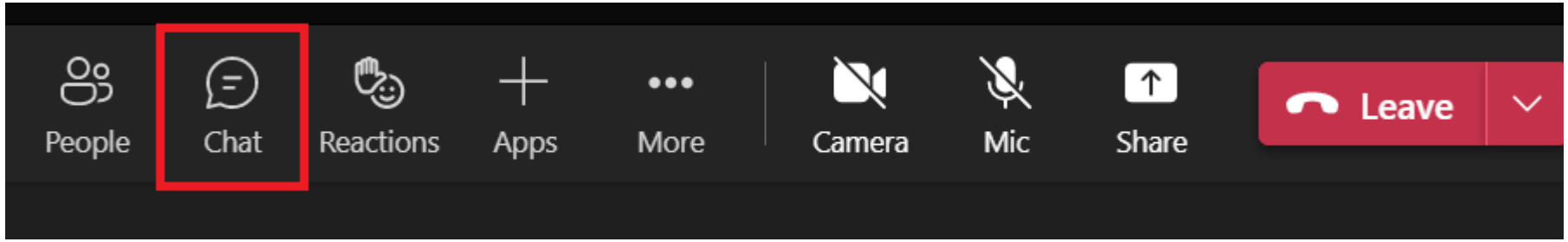
Interoperable

Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

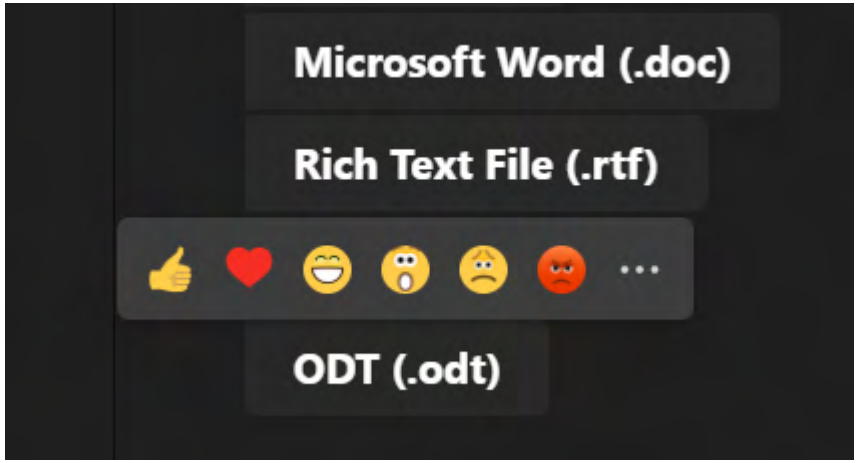
- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

Formats



- Which format will be preferred for text documents?
  - Microsoft Word (.doc)
  - Rich Text File (.rtf)
  - PDF/A (.pdf)
  - ODT (.odt)



  - Think about changes over time…

Formats

- Which format will be preferred for text documents?
  - Microsoft Word (.doc)
  - Rich Text File (.rtf)
  - PDF/A (.pdf)
  - ODT (.odt)

Type	Preferred format(s)	Non-preferred format(s)
Text documents	• PDF/A (.pdf) • ODT (.odt)	• Microsoft Word (.doc) • Office Open XML (.docx) • Rich Text File (.rtf) • PDF other than PDF/A (.pdf)

File formats | DANS (knaw.nl))

Formats

- Which format will be preferred for spreadsheet?
  - Microsoft Excel (.xls)
  - Office Open XML Workbook (.xlsx)
  - PDF/A (.pdf)
  - ODS (.ods)
  - CSV (.csv)

- File formats | DANS (knaw.nl)

Interoperable

Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

Dublin Core : metadata standard

Logo image of DCMI, which formulates Dublin Core

1. Title
2. Creator
3. Subject
4. Description
5. Publisher
6. Contributor
7. Date
8. Type
9. Format
10. Identifier
11. Source
12. Language
13. Relation
14. Coverage
15. Rights

Interoperable

Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

How many words do you know for dog?

Thesaurus / *dog*

dog See definition of *dog* on Dictionary.com

noun **canine mammal** | *verb* **chase after; bother**

SYNONYMS FOR *dog* **Compare Synonyms**

pup	doggy	pooch	fido
puppy	hound	stray	flea bag
bitch	mongrel	tyke	man's best friend
cur	mutt	bowwow	tail-wagger

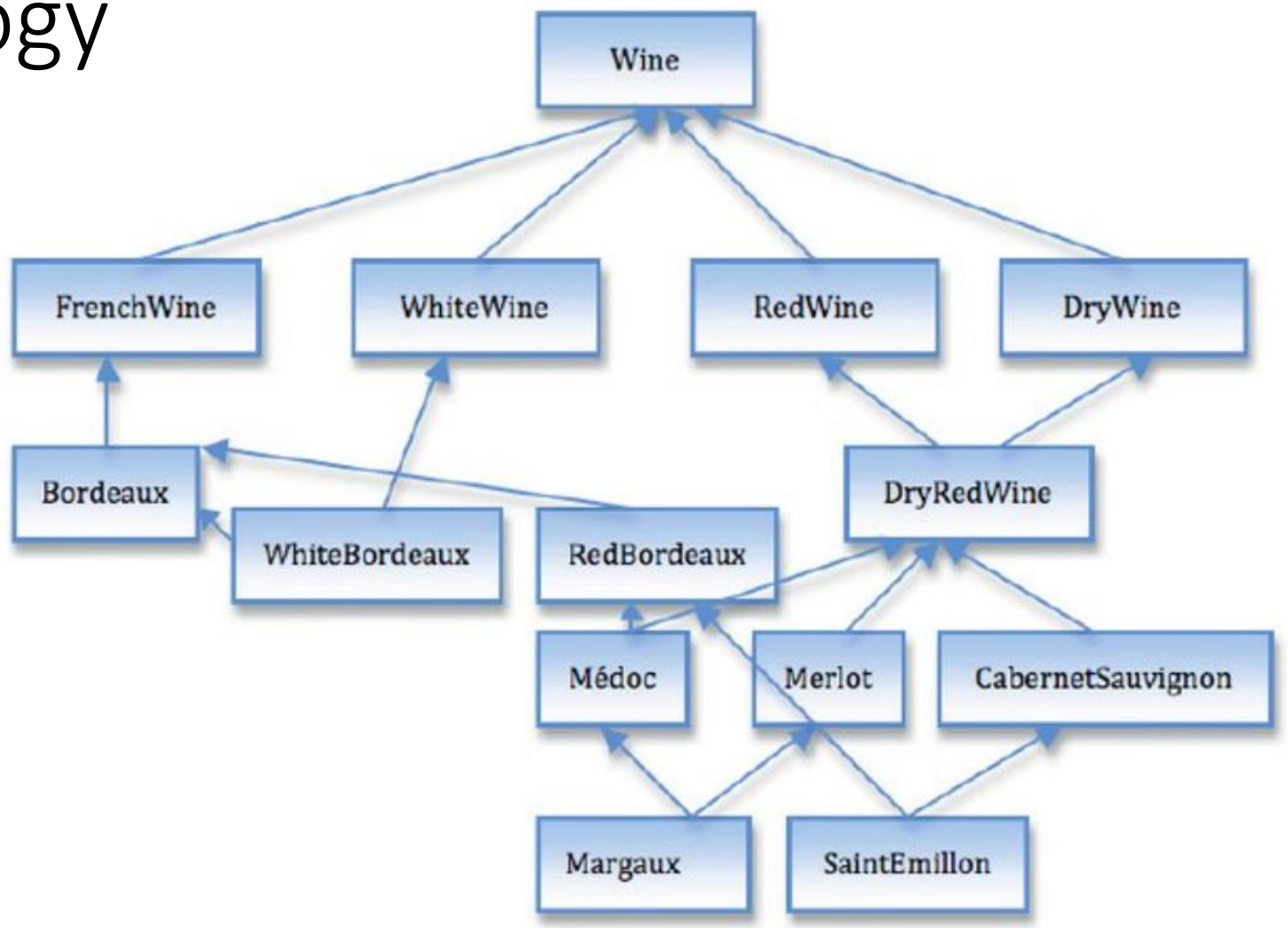
See also synonyms for: **dogged / dogging / dogs**

Roget's 21st Century Thesaurus, Third Edition Copyright © 2013 by the Philip Lief Group.

Plant versus plant

CC BY-SA 4.0 Tiia Monto

Ontology


Wine
FrenchWine
WhiteWine
RedWine
DryWine
Bordeaux
WhiteBordeaux
RedBordeaux
DryRedWine
Médoc
Merlot
CabernetSauvignon
Margaux
SaintEmillon

Controlled vocabulary thesaurus : NLM Pubmed

Entry Terms:

- Dog
- Canis familiaris

All MeSH Categories
Organisms Category
Eukaryota
Animals
Chordata
Vertebrates
Mammals
Eutheria
Carnivora
Caniformia
Canidae
Dogs

Standardisation : measurement unit

- Temperature
- degree Celsius vs Fahrenheit

- Length
- Meters and centimeters vs inches and feet

- Windspeed
- m/s vs knots/s

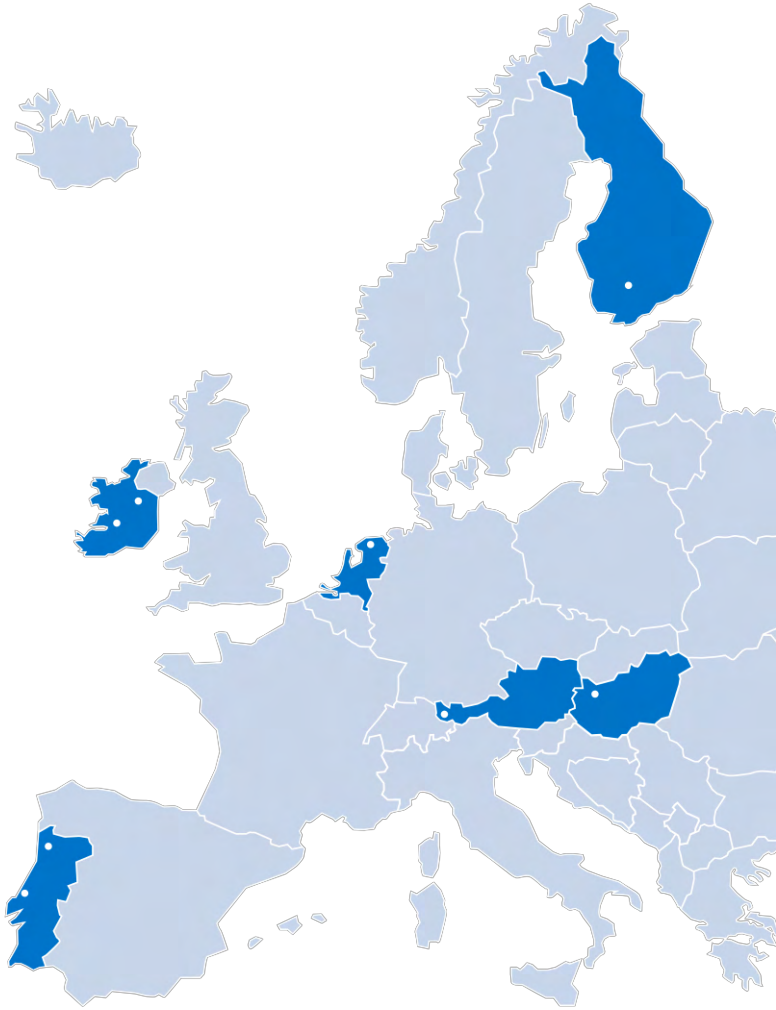
Standardisation : raw data is preferred

- Bio Mass Index
    - kg / (length in meters * length in meters)

- Square Meter
    - Length * width

REGIONAL UNIVERSITY NETWORK PLUS

EUROPEAN UNIVERSITY



R - Reusability

Co-funded by the Erasmus+ Programme of the European Union

The content of this publication represents the views of the author only and is his/her sole responsibility. The European Commission and the Agency do not accept any responsibility for use that may be made of yhe information it contains. Grant Agreement Number: 101035816

POLITÉCNICO DE LEIRIA · IPCA INSTITUTO POLITÉCNICO DO CÁVADO E DO AVE · TUS Technological University of the Shannon: Midlands Midwest · HAMK · SZÉCHENYI EGYETEM · NHL STENDEN university of applied sciences · FH Vorarlberg University of Applied Sciences

Reusable

Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

- ☐ The data are accurate and well described with many relevant attributes
- ☐ The data have a clear and accessible data usage license
- ☐ It is clear how, why and by whom the data have been created and processed
- ☐ The data and metadata meet relevant domain standards

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

Reusable

Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

- ☐ The data are accurate and well described with many relevant attributes
- ☐ The data have a clear and accessible data usage license
- ☐ It is clear how, why and by whom the data have been created and processed
- ☐ The data and metadata meet relevant domain standards

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

What information do you need to provide when you contact the copyrightholder

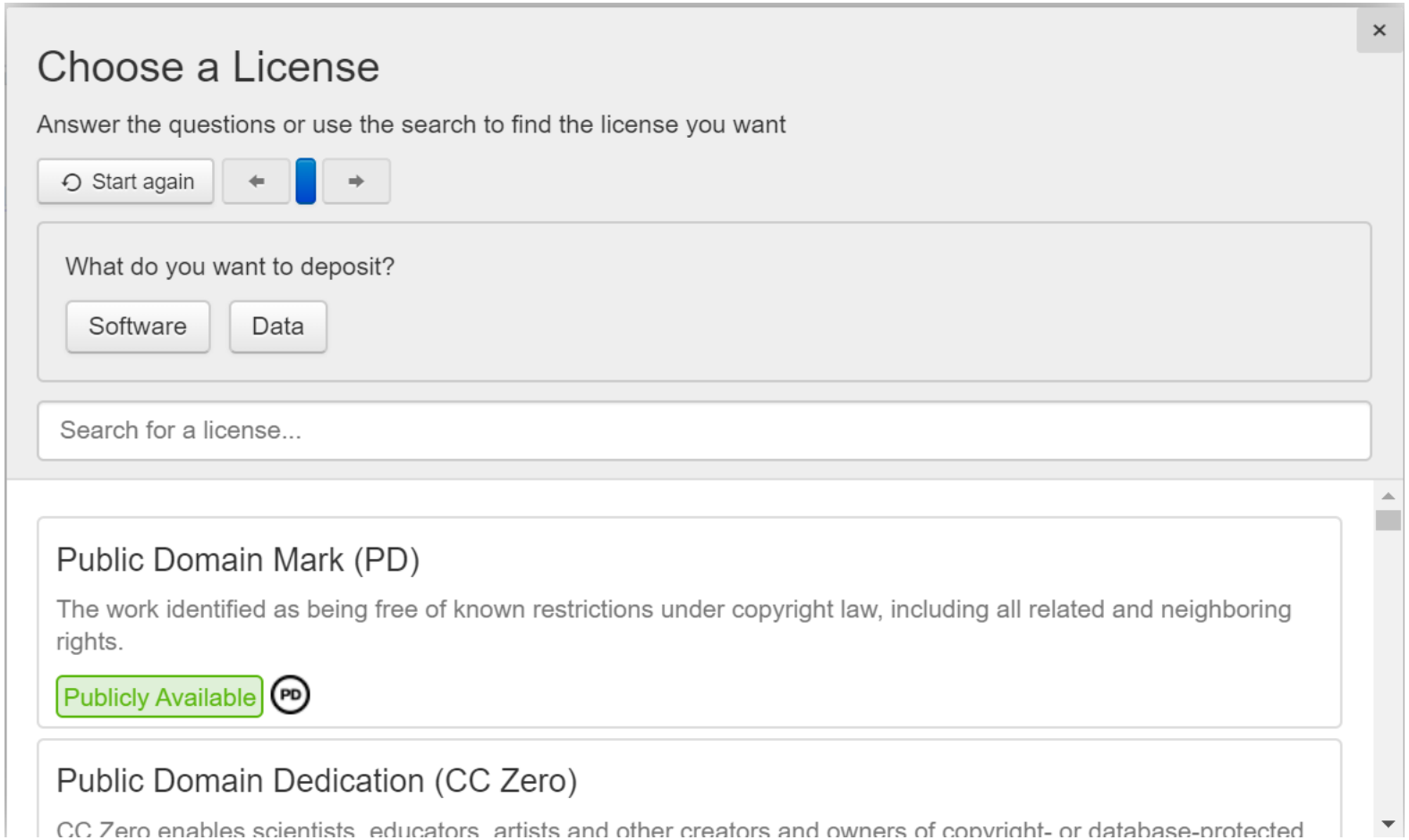
- What you will be using (amount and content)
- The context their work will be used in
- Where you will be using the work (e.g publicly online)
- For what for purposes (e.g educational, commercial, personal)
- How they will be attributed

CC - licenses

Video CC BY NC SA 4.0 U of G library

Licenses for software or data: Choose the right license

- GitHub



Reusable

Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

- ☐ The data are accurate and well described with many relevant attributes
- ☐ The data have a clear and accessible data usage license
- ☐ It is clear how, why and by whom the data have been created and processed
- ☐ The data and metadata meet relevant domain standards

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT

README files

- What do README.txt files contain?

Cornell

AUTHOR_DATASET_ReadmeTemplate.txt

This DATSETNAMEreadme.txt file was generated on YYYY-MM-DD by NAME
<help text is included in angle brackets, and can be deleted before saving>

GENERAL INFORMATION

1. Title of Dataset:

2. Author Information
    A. Principal Investigator Contact Information
        Name:
        Institution:
        Address:
        Email:

    B. Associate or Co-investigator Contact Information
        Name:
        Institution:
        Address:
        Email:

    C. Alternate Contact Information
        Name:
        Institution:

READ ME files

Take a look at data set 1 and discuss: :

- Does the README file contain all the information needed?
- Are the data files shared in a preferred format?

Copy the data set links and the readme file template from the chat to your computer before entering the break-out room.

Time: 15 min.

If time left take a look at another dataset data and compare.

What is FAIR not?

FAIR data is not a standard
FAIR is based on principles with the goal to reuse valuable research objects.

FAIR data are not always open, open data are not always FAIR
FAIR data are data which meet principles of findability, accessibility, interoperability, and reusability.

FAIR data is not just about humans being able to find, access, reformat and reuse data.
Computers should also be able to access a data publication, without help from human operators.

Are there any questions?

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Feedback

Feedback is given anonymously

Feedback will be used to develop future training programmes
Responses will be compiled into a report for the EU
Feedback will be shared FAIR with IMPACTLAB | Universiteit Utrecht (uu.nl)

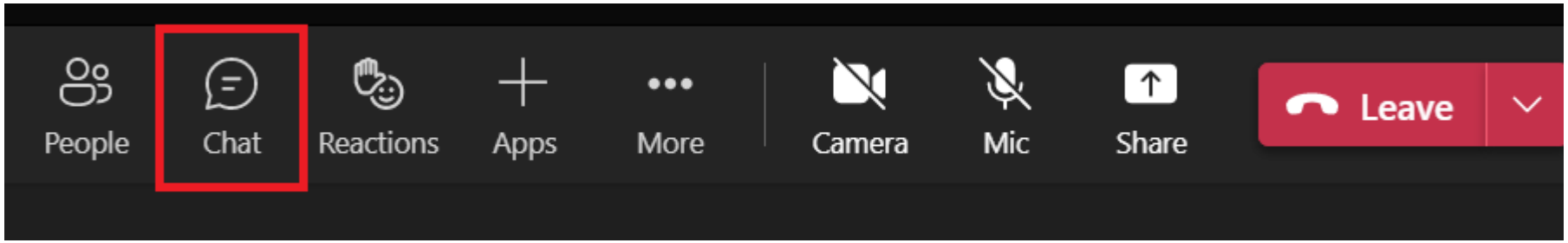
Participants can stop providing feedback at every moment

If you want to know more about privacy protection in this research you can ask:

- Ingrid van Gorkum (projectleader) via datasupport@nhlstenden.com]
- or Floor May (data protection officer) via privacy-en-security@nhlstenden.com

Feedback

In the chat you'll find a link



https://forms.office.com/r/L3qwUhGDFV

Research Career Development

Register for the June 9th workshop
https://run-eu.eu/2022/05/16/online-workshop-on-attractive-researcher-career-paths/

More information:
https://run-eu.eu/2022/05/20/researcher-career-development-training-programme-2022/