HELMHOLTZ METADATA COLLABORATION

HZB Helmholtz Zentrum Berlin

Pilot training on Research Data Curation: implementation review

Luigia Cristiano, HMC

Helmholtz Zentrum Berlin

Why this contribution

Trigger the discussion on :

- Robust evaluation of a training offer
- Enrich the slides + jupyter notebook to new training tools and implementations
- the didactic planning approaches in training

Content

- Overview of the course plan
- **Didactic structure**
- Jupyter notebook
- **Evaluation board**

ALL WORK in PROGRESS !

Why the Pilot training

Test the effictiveness of the training offer respect to:

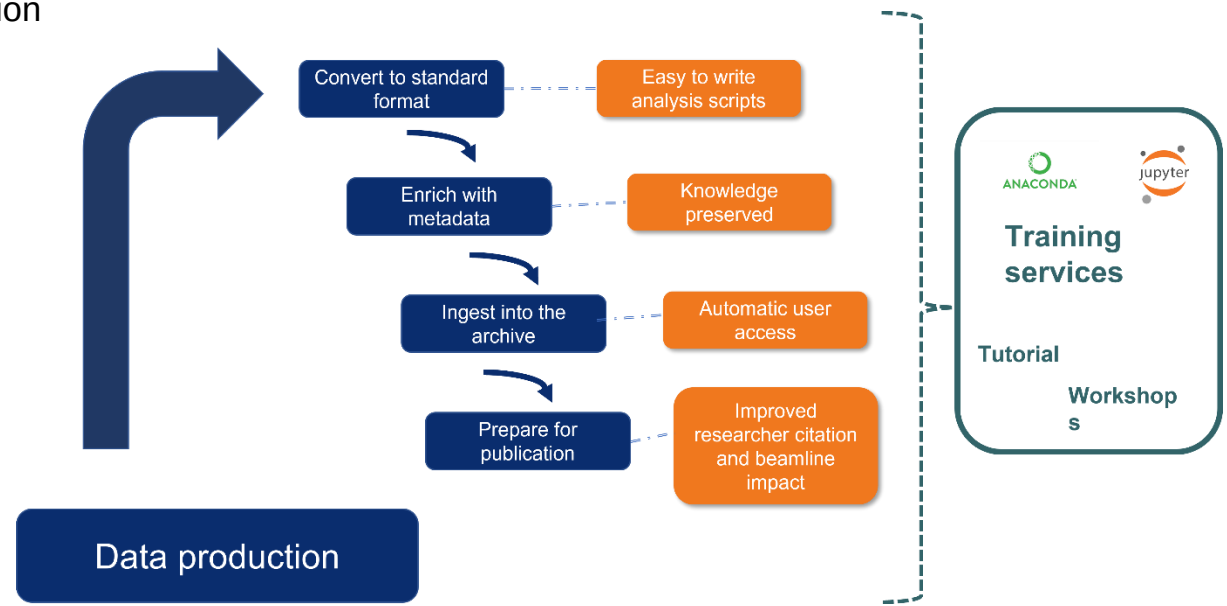
- Learning targets
- Shaping of the content to the participants knowledge
- Time frame-agenda
- Hands on session
- ROI

Being able to evaluate and improve

Pilot course: the subject and the aim

Trigger the implementation of FAIR data curation practices:

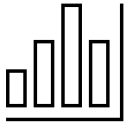
- Refer to published FAIR implementation Guidelines
- Presentation of use cases
- tools
- Implementation strategies



[1]

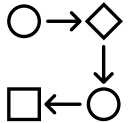
Planning the curriculum as an element of a training program:

1. Identify the target group



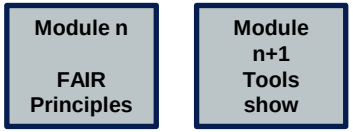
Survey and interviews to collect discipline and technical knowledge and training needs

2. Define learning goals and evaluation strategies



Focuses on the competences and trigger of actions and awareness

3. Decide the container: format, structure and duration



workshop, seminar, tutorial

3 days, 1h, 1h to be repeated for updates

jupyter lab, gitlab, virtual environment, institutional web services

Modularity, tools provision, reusability and portability

4. Identify the methods and the group participation

Didactic structure and methods

Training cycle: Implementation and test

Design

- **Didactic structure** as training course pillar : for a **learning goals** oriented course design
- Training **tailored** to the researchers needs: targeted audience
- **Modular** structures and multiformat FAIR offer
- **Tutorial** as wrapper of the developed tools
- Training **review** by peers ex-ante and ex-post

Run

- **Pilot workshop** on 22-23th Nov' 21
- ~15 researchers from a measurement station where tools have been installed
- Slide plus "test the tools" in Jupyter
- Tools demo

Evaluate

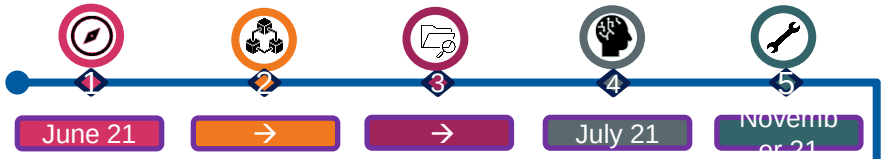
- **Categories/Dimensions and indicators** defined for the evaluation
- Self assessment, participants satisfaction and external review by peers
- FAIRness, originality of the product Trainer problem solving capabilities are also used as quality indicator

Future steps

- Improve, integrate, variate the training offer

Into the Pillars of the training implementation

1.Plan and Project Management



Identify the training needs of the research and project fellows

Develop a didactic structure to serve a permanent training

Identify the available training services and material.

Start a course design to better integrate the external with the internal training services

Propose the developed course to the community

2. Plan a learning Unit

Work units: time in hours	Assignments	Didactic approaches/ methods and strategies	Structure and scheduling of the work	Tools and environment
Day 1- Part1 3-hours	FAIR principles: good practices for data curation and data distribution	Problem posing Problem solving Learning processes: Interpretation Elaboration	Structure: module 1 -direct teaching; -guided confrontation	datacite material Internet Teaching platform Slides VM zoom environment online in presence

3. Modules

Module 1

FAIR Principles and PaN data life cyclcle

Module 2

Preserve and sharing

Module 3

FAIR implementation at HZB

4. Evaluation Rubric

Criteria + indicators …

Building the training Pilot

01

1. Identify target group : on going collaboration

02

2. Interview for identifying the knowledge level and technical background

03

3. Learning Units definition: Training offer in form and duration scaled to the target group necessities

04

4. Abstract was distributed, agenda was defined with staff scientists

05

5. Hands on session planned to trigger the trainee participation and the tools provision

Plan a Learning/Teaching Unit

1. Learning unit approach adapted for the training environment

2. Focuses on the disciplines related competences

4. Modularity implemented to favour the interoperability and reuse of the Pilot model

5. The development of the training content started from Hands on sessions and tools show

6. Review at the end of each unit

Reference: european guidelines for permanent training of adults

Assignment: FAIR principles and its implementation	Learning targets/disciplinary competences
FAIR Data curation guidelines	Getting familiar with FAIR principles and being aware of the implementation strategies
Metadata structures, PIDs, Licences	Being able to navigate through the metadata standards fields, being able to distinguish between different PIDs and their use. Awareness on the terms of use of the data
Tools for metadata conversions formats	Being able to convert from table to xml and json, viceversa and knowing the context in which this manipulation might be necessary
Assignment: FAIR data curation at Hub Matter- HZB	Learning targets/ disciplinary competences
Data curation workflow at Hub Matter and HZB	Getting familiar with FAIR principles implementation at HuB matter – HZB
“ “	Being able to use the institutional tools for data curation

Didactic plan of a Learning Unit

1. Teaching didactic plan approach adapted for the Hub environment and trainee chara.

2. Focuses on the competences

4. Modularity implemented to favour the interoperability and reuse of the Pilot model

5. **Identification of methodologies for the environment and the forma and participation level of the trainee group**

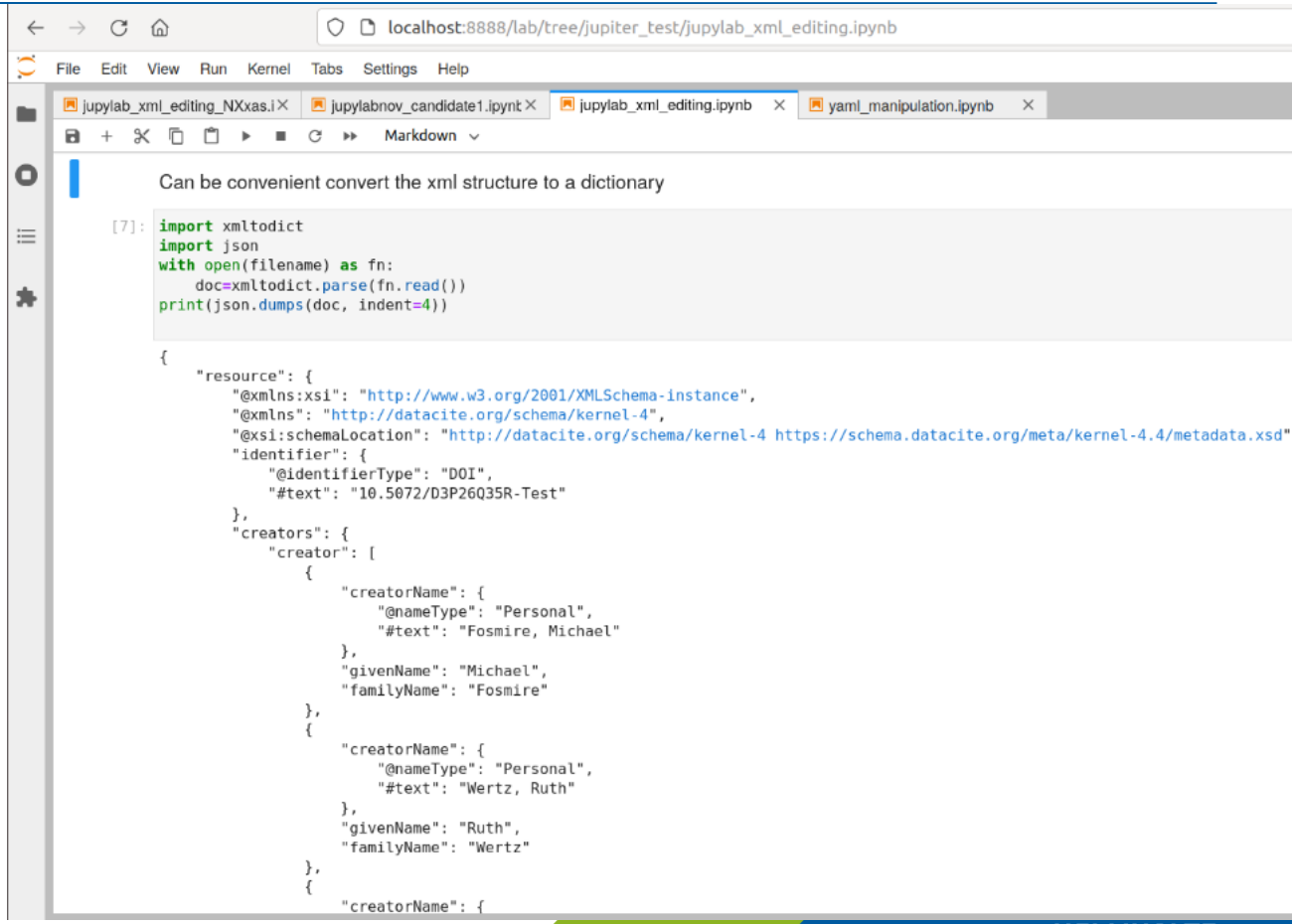
Work block: time in hours	Assignment*	Didactic methods and strategies	Structure and organization	Tools and environment
Day 1- Part1 20 min	Presentation of the FAIR principles And reference to the HMC initiative	Direct teaching Problem solving Investigation To Interprete and elaborate the information	Slide presentation and discussion	Tools Slides Zoom environment online
Day 1- Part1 30 min	Presentation of use cases for Scientific data curation	Investigation, interaction participation	Slide presentation and discussion	Respect previously listed tools
Day 1 -Part2 50 min	Metadata standards: generic and discipline dependent + Hands-on session	Problem solving approach Is chosen for this “How to” session	Slide presentation and Guided diiscussion is followed by the hands on session with a jupyter notebook This helps to make a break on the direct teaching and discuss use cases	Jupyter lab, python functions, metadata schemas
“ ”	Presentation of PIDs. Which FAIR aspect they enable, different PIDs services Licences overview of terms of use of data	Direct teaching	discussion using material collection listed in the slides	slides
Day 1 – Part 2 30 min	Open discussion	The information is elaborated on the short term and implementation issues are discussed. Guided discussion	Conclusion of the day, open discussion for remarks and questions that we could not have during the other time slots	
Day 2 Part 1	Presentation of data	Direct teaching+ problem posing	Introducing the services	Slides, web services, jupyter

Hands on session with Jupyter notebooks

1. Provide tools and environment for testing them

A virtual environment for local testing is going to be distributed

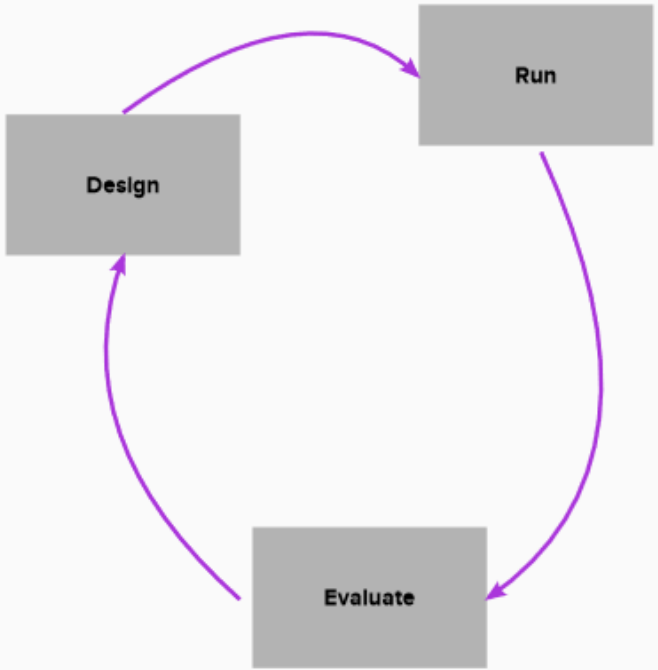
What can we add here ?



Evaluation

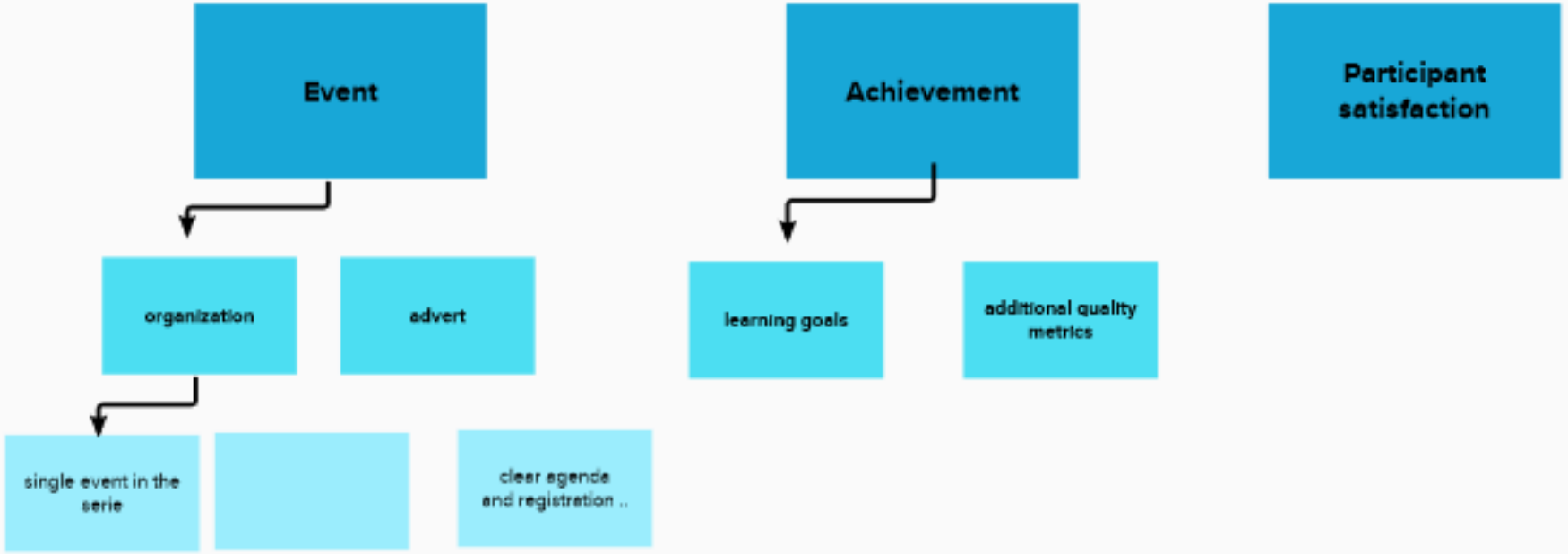
1 **Why evaluate**

Use knowledge to improve the training. ex ante and ex post (systematic) evaluation to be fed in the training design and a grid to track the time dependency



Evaluation

2 What evaluate



Evaluation

Achievement of the learning goals and additional quality metrics

1. Awareness, competences, impact and actions
2. Coherence and relevance of the training
3. Efficiency and effectiveness
4. Implementation of FAIR
5. Technical support
6. Identification of target group and scaling and adapting of the training offer

……..

Methods

1. Rubric: categories, indicators, scores
2. Experts as evaluators. Briefing and observation
3. Benchmark on the efficiency and effectiveness

Evaluation

3 **How evaluate**

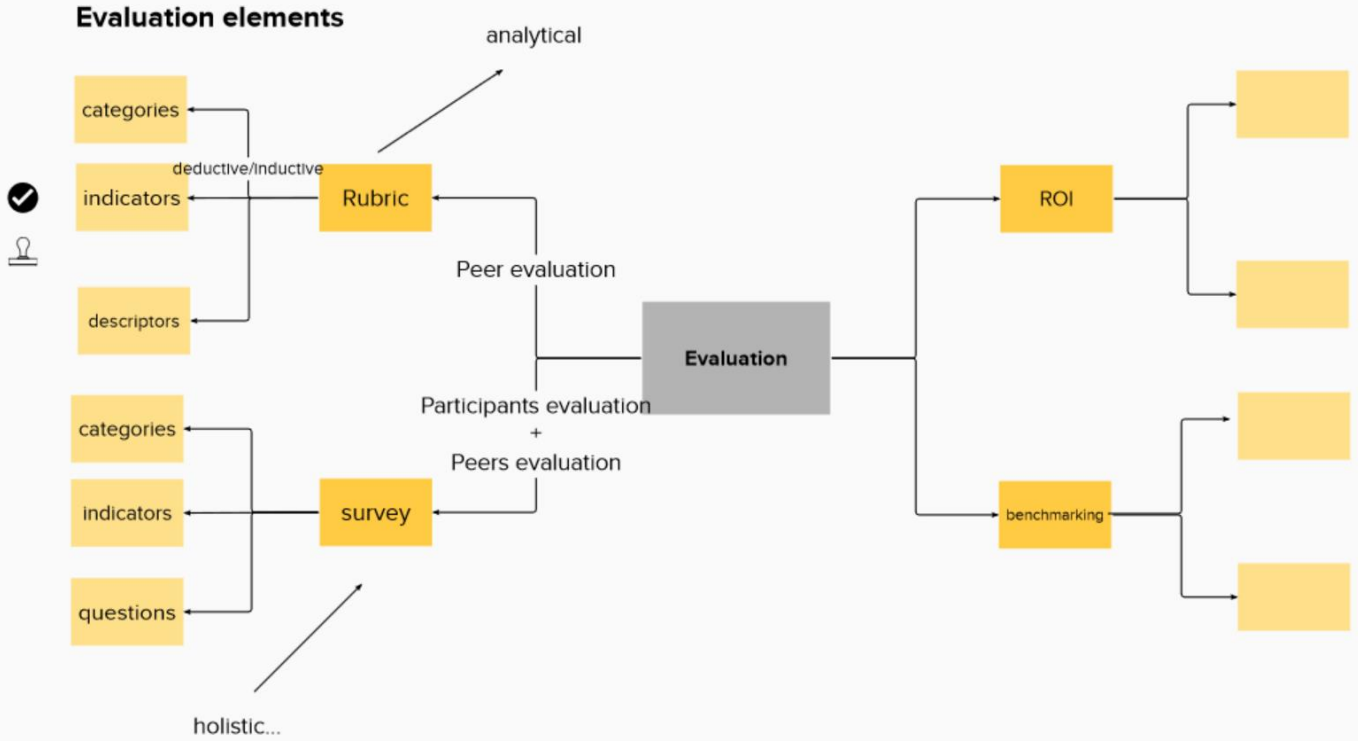
Define approach and tools to evaluate a training offer.

Indicators need to be:

Robust

valide

reliable



4 Identify Categories to evaluate

Some definitions:
Dimensions /categories: Characteristics of the event we want to measure

criteria: which aspect to measure

indicators: how to depict what we want to measure and how to measure the target achievement

scores and anchor points: grid representation of the measures
descriptors: detailed explanation of the indicators

example
Dimension: Organization
criteria: efficiency in the event organization
indicators: workshop as single event in a project
descriptors: the event was clearly embedded in a series with a project behind

Goal achivement	FAIR	Portability	Originality
Learning objective	material reusability	platform independence	not collecitve work

1. Categories
2. Indicators: univocal, precise, measurable
3. Indicators needs to be validated→pilot course
4. The definition of the categories was done considering the critical elements of the training offer. The definition of the indicators followed the SMART approach: Strategic, Measurable, achievable, Relevant and Time related

identified the stakeholders

Evaluation : First definition of Criteria and Indicators-- inductive

Criteria/classes	Indicators/Scores 1	2	3	4	5
Clearness and coherence	poor **material,** chaotic **structure** Training goals not indicated	No overview of the training structure but modularity is there. Not all the material is organized per training areas		The training aim has been clearly stated at the beginning of the course but not all the material was coherent to the scope	The didactic/training unit presents a reusable modular structure a learning goal is indicated and all material is coherent to it
Completeness and redundancy	The content was **redundant** and not completely **relevant** to the scope of the training			The hands-on session is coherent with the course content and learning targets	The presented material was relevant to the learning objectives and training goals
FAIR	the data provenance is not indicated the material is not on line accessible	The data provenance is clear, the material is not completely reusable due to licence problem and portability	Modularity structure of the training product favour the reuse but no guidelines on the hands on session is provided	The learning material is made available in an open-access data publication repository	The learning material is provided in a accessible repository. A setup file is provided to run the provided tools. The toolbox slides+jupnot are OS independent
Goal achievement	The training does not promote the engagement of the participants in the data curation.	The training offer consist of slides and doesn not offer a conctrete possibility to implement the data curation		The design of the course has been accurate and offer a rich hands on session that facilitate the competences acquisition	The implementation of FAIR data curation has been fostered and tools have been provided for the direct implementation of the acquired knowledge and competences
Product originality	The training is a material collection of accessible sources	The knowledge and prerequisites are not taken in account into the course structuring		The design of the training offer has been accurated but the slides offered are information collection of already available material	

More on the evaluation rubric

Criteria/classes	Indicators/Scores 1	2	3	4	5
Trainer role: Technical and disciplinary competences	The trainer showed poor competencies in handle basic technical issues and his/her	…	…	…	Strong problem solving capabilities of the trainer,
Organizational	The poor organization competences affected the advertisement of the event and the participants registration procedures	The course design was not aligned to the training a scope There is no report available on the modality of course preparation	….	the course was characterized by a very well balanced breaks and exercises sessions	The trainer adapted the course to fit the audience expectation and showed excellent organization skills
Communication aspects	….	….	The trainer showed good communication competences speaking with confidence, the language was not enough precise	The trainer show good communication competences presenting in a straight forward manner and technical language	The trainer used technical and precise language preserving the understanding of a non specialist

Evaluation: validation of the indicators

- **Peers committee**: experts in didactic, project management, data manager

- Method: **measure the validity** but also the reliability and robustness.
- Go for a dense array of indicators and not for a sparse one

- Note the **indicators** can be also a **proxy** measure

- Integrate the indicators with complementary approaches

- Indicators might be converted to **degree of achievement** instead of scores

- Indicators might define the survey structure

Definitions:
Validity: to which degree this measure depict the information we want to measure,
Reliability (this measure tells us of the relative category information without bias or free of errors),
Robustness (independent on the context- not sure we want this)

How to validate ?

- Identify possible dependencies between the indicator,
- Define different measures and evaluate how stable is the assessment by replacing the indicators.

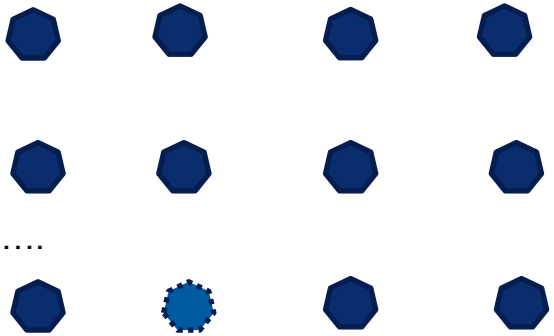
- Minimum set of system representation or grid search and minimum misfit
- Committee of experts
- Test the matrices and adjust

$ind_{i,i}$, $ind_{i,j}$

$ind_{j,i}$, $ind_{j,j}$

……

$ind_{z,i,\ldots..}$



References and Further reading

1. https://ec.europa.eu/info/sites/default/files/file_import/better-regulation-toolbox-47_en_0.pdf
2. https://www.euroguidance.eu/survey-on-update-of-the-european-guidelines-for-validating-non-formal-and-informal-learning
3. Mc Tighe J., Ferrara S. (1996). Performance-based Assessment in the Classroom: A Planning framework, in Blum R. E., Arter J. A. (eds), A Handbook for student performance assessment in an era of restructuring, Alexandia: Association for Supervision and Curriculum Development
4. European qualification framework and national guidelines (QCER)
5. Teaching unit form modified after https://www.icsassuolo1centroest.edu.it
6. Guidelines for the development of a didactic unit Italian Ministry of Education, 2021. https://www.miur.gov.it/documents/20182/0/ALL.+A++*Linee*Guida*DDI*.pdf/f0eeb0b4-bb7e-1d8e-4809-a359a8a7512f
7. European guidelines for learning objectives : https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018H0604(01) guidelines of the European Commission on key competences for lifelong learning

Slide 6 : sketch modified after verbaltovisual.com

Thanks.

luigia.cristiano@helmholtz-berlin.de