HZB BESSY II Light Source

A step towards open data and open Science

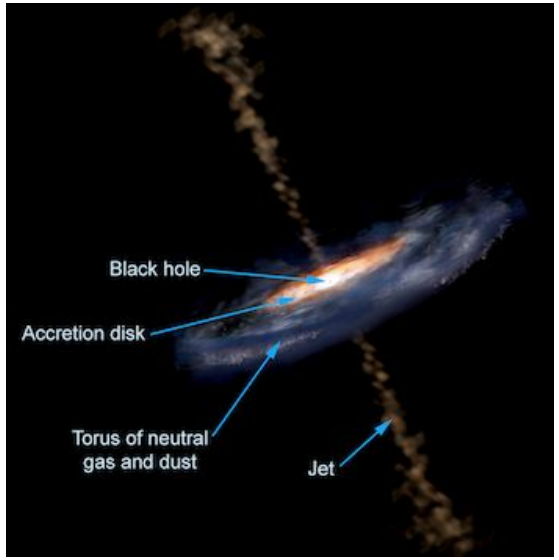
Introduction to FAIR principles

Sonal. R. Patel
06.08.2024

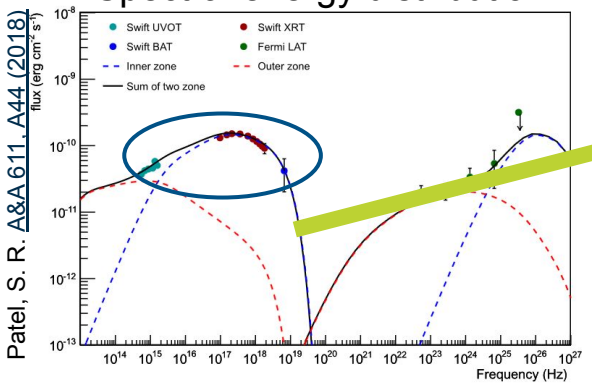
FAIR principles

About me

Ph. D
Working with open data



Spectral energy distribution



Patel, S. R. A&A 611, A44 (2018)

Post-doctoral
Making open software and data

MAGIC Major Atmospheric Gamma Imaging Cerenkov Telescopes

H.E.S.S.

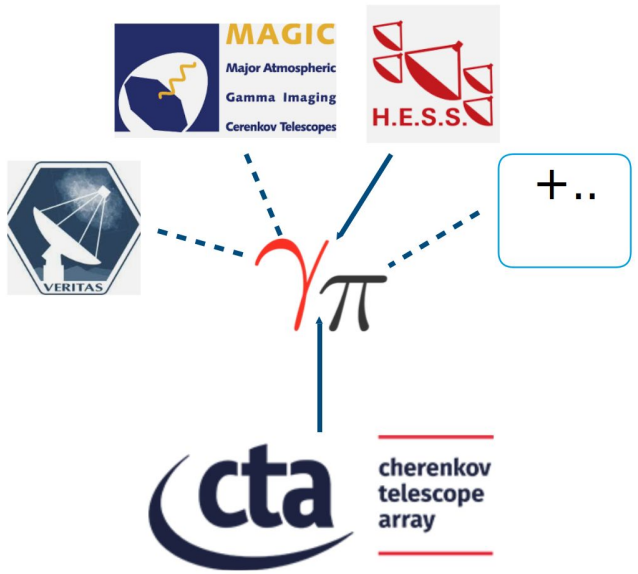
VERITAS

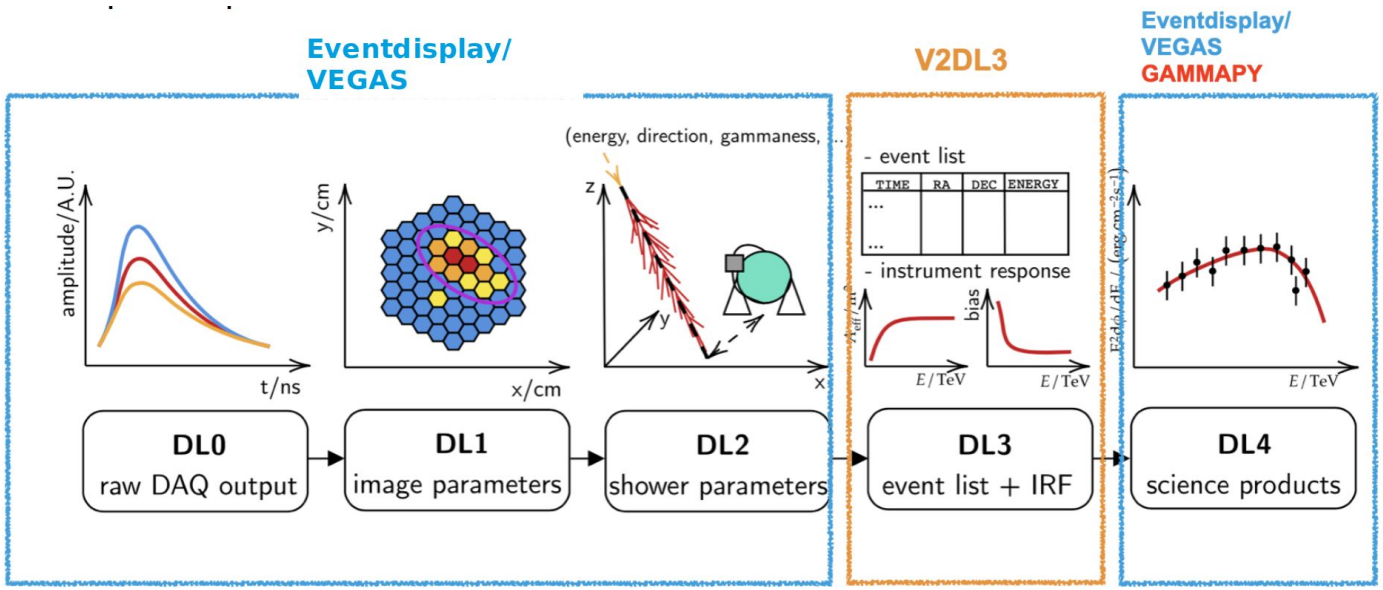
+..

cta cherenkov telescope array

Data Steward
Producing FAIR data

Motivation





Credit: Nigro C. et al. (Universe 2021, 7, 374)

- Long-term maintenance and Support

- GADF: Gamma-ray AstroData Format. The *Data formats for gamma-ray astronomy* is a **community-driven initiative** for the **definition of a common and open** high-level data format for gamma-ray instruments

Introduction

- Need:
  - Why?
  - Where?
  - What?
  - When?
  - Who?
  - How?

Introduction

Introduction

Introduction

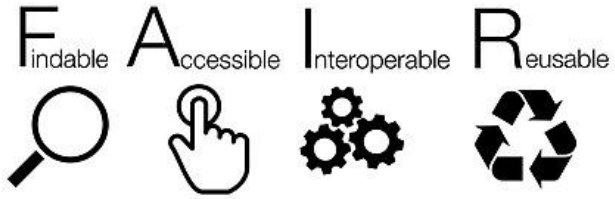
Introduction

Introduction: FAIR

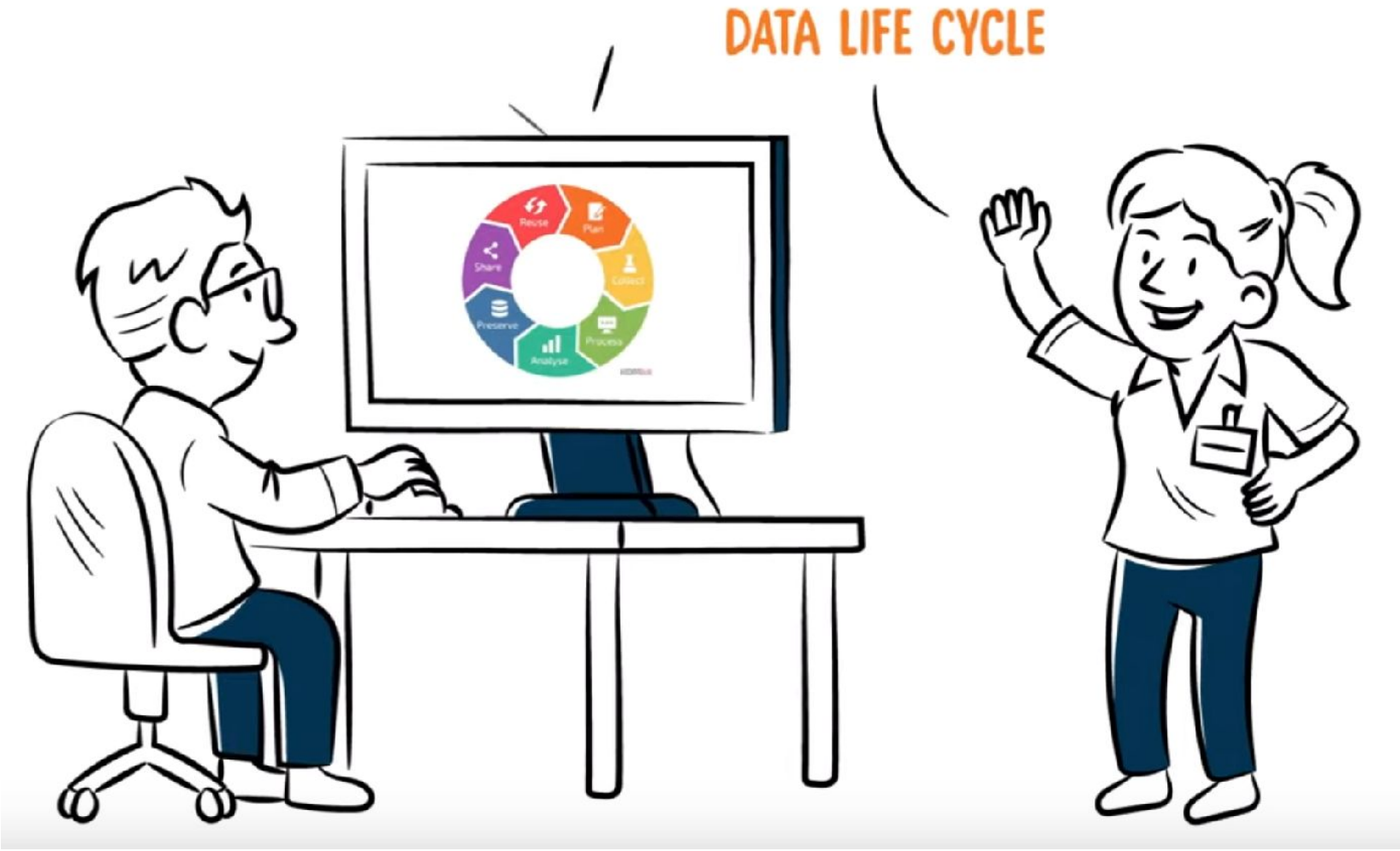
Introduction: FAIR

- **GO FAIR initiative**: The acronym and principles defined for the first time (Mark D. Wikinson et al. 2016, March 15) by **Force11**

- **Conclusion by the council:** Transition towards an open science system (2016, May 26)
  - The Council of the European Union emphasises that “the opportunities for the optimal reuse of research data can only be realised if data are consistent with the FAIR principles (findable, accessible, interoperable and re-usable) within a secure and trustworthy environment”
- **European Data Strategy** (2020, February 19):
  - A way to implement interoperability.



Data Life Cycle

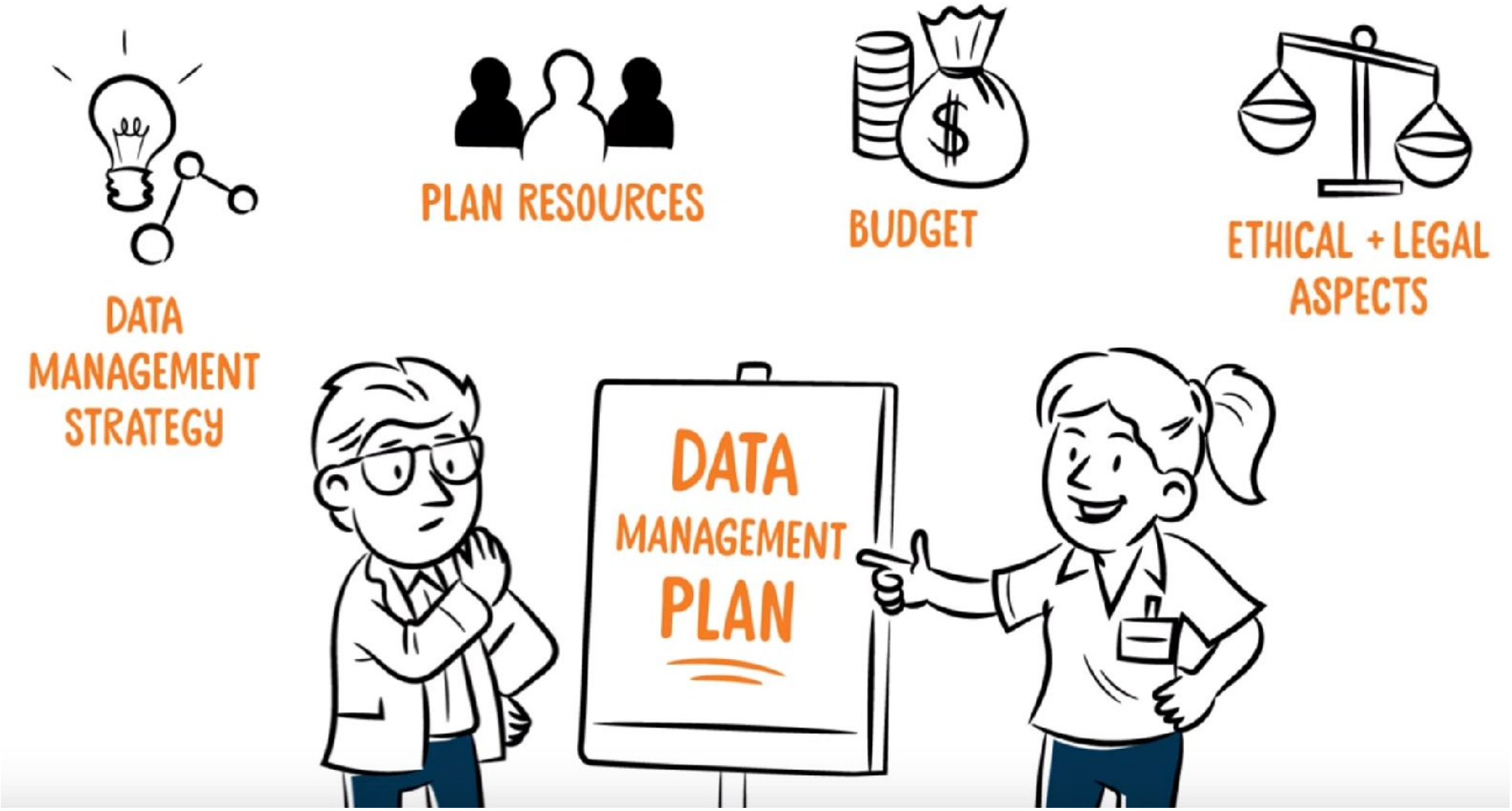


FAIR

DATA LIFE CYCLE
DATA PLANNING
Plan
Collect
Process
Analyse
Preserve
Share
Reuse

FAIR

Data Life Cycle: Plan



FAIR

DATA LIFE CYCLE
DATA COLLECTION
Reuse
Plan
Share
Collect
Preserve
Process
Analyse

FAIR

Data Life Cycle: Collection

Data Life Cycle: Collection

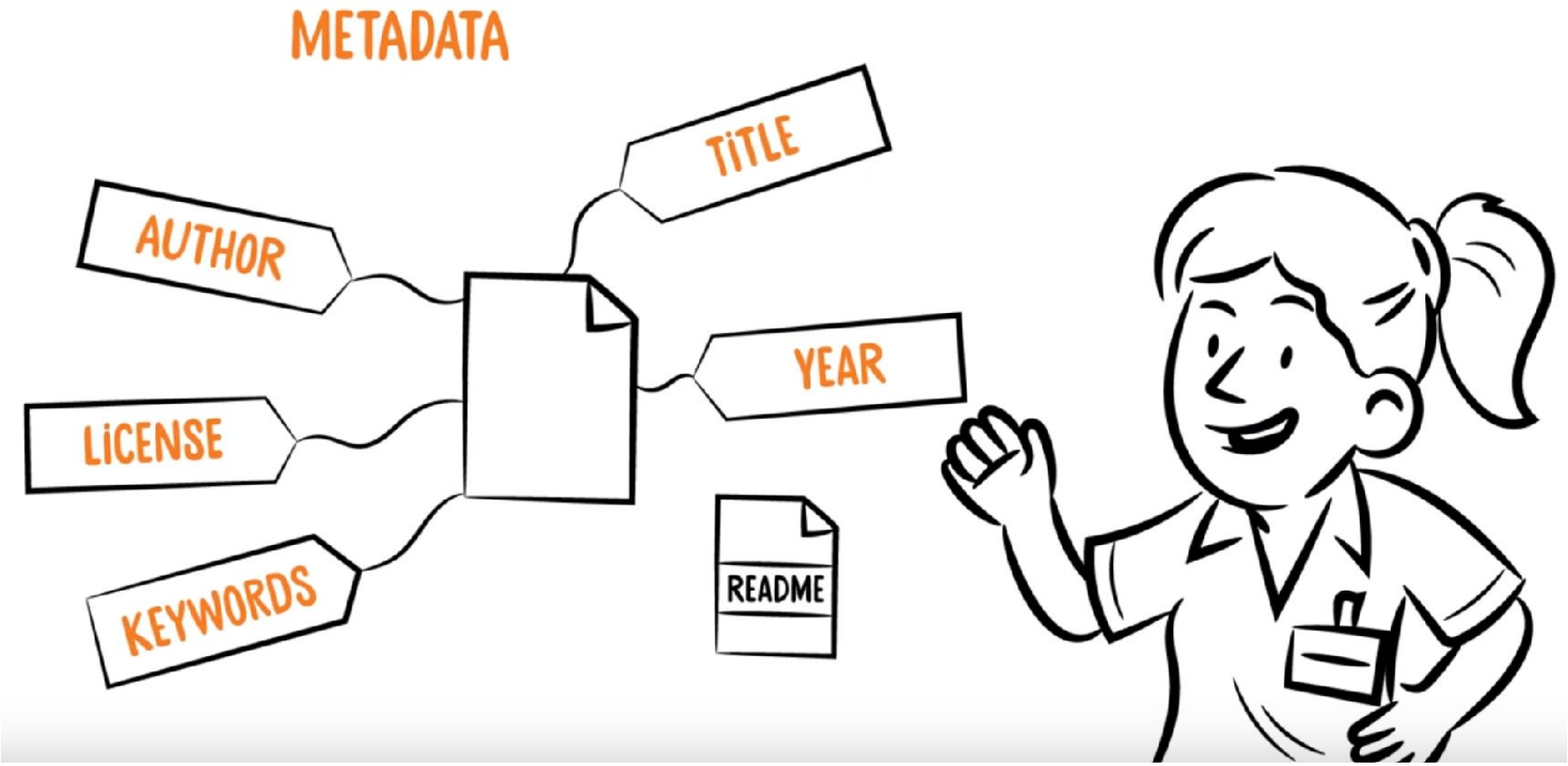
FAIR

Data Life Cycle: Collection

FAIR

Data Life Cycle: Collection

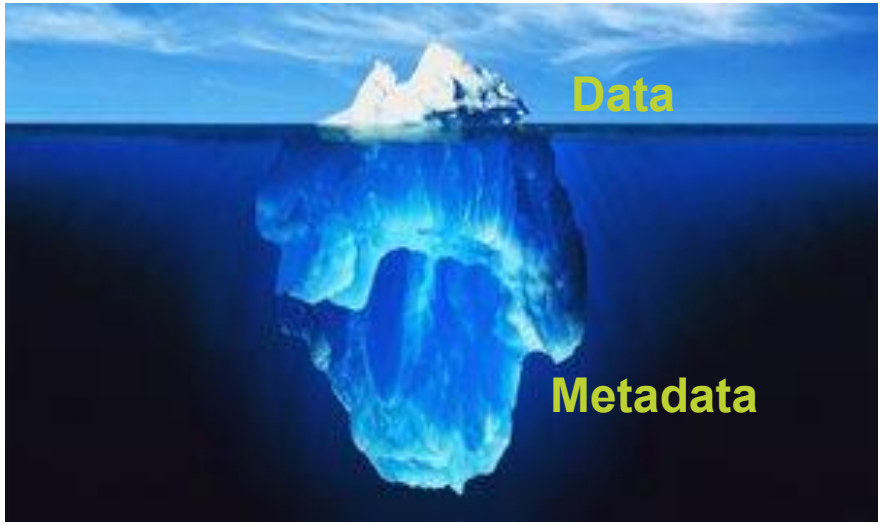
- Metadata: A Love note to the future (Jason Scott)



FAIR

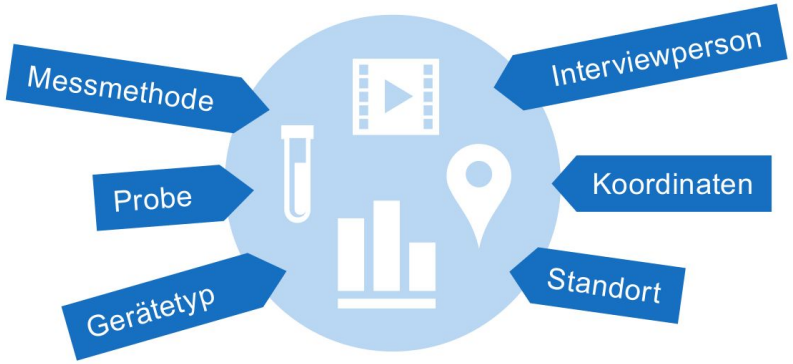
Data Life Cycle: Collection

Metadata 🎯



Bibliographic or administrative data:

- Author
- Title
- year
- Format

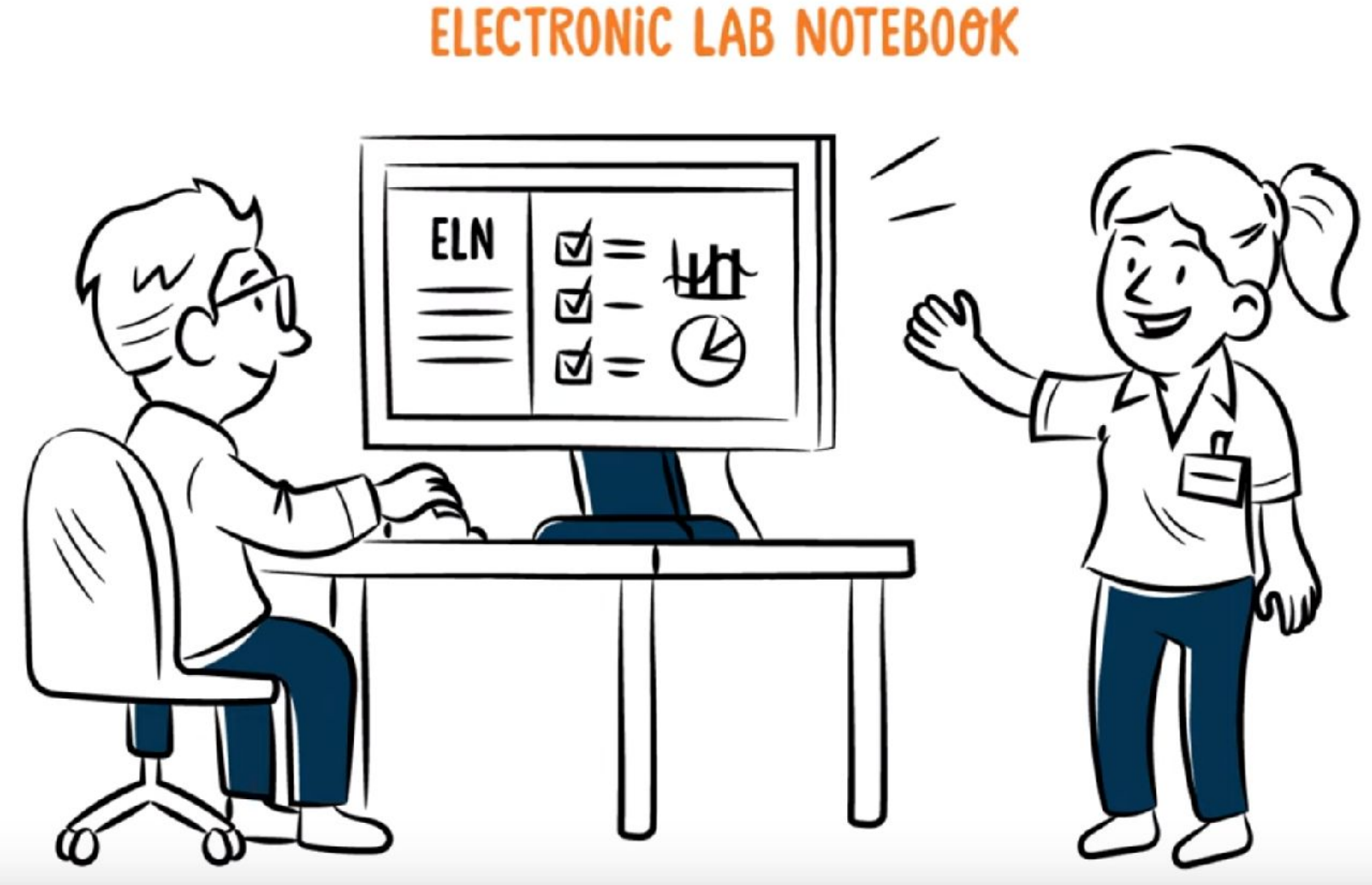


Content-descriptive or technical data:

- Measurement method
- Probe
- Device type
- Location
- Coordinates
- Interview person

FAIR

Data Life Cycle: Collection



FAIR

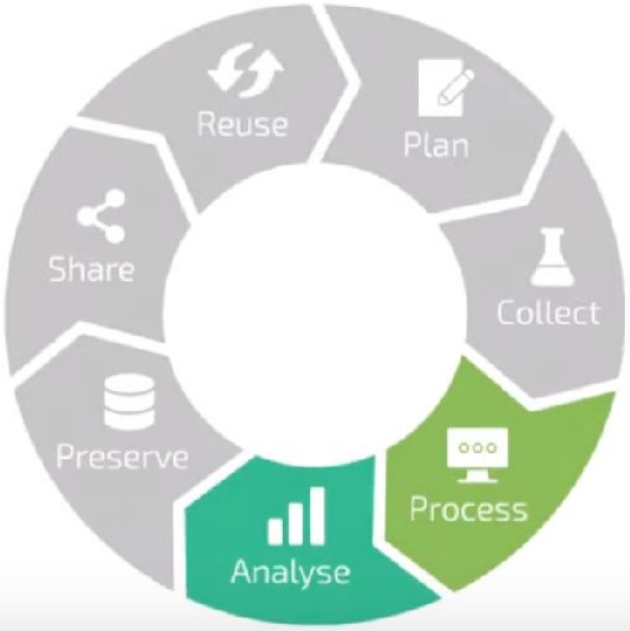
Data Life Cycle: Collection

FAIR

Data Life Cycle: Collection

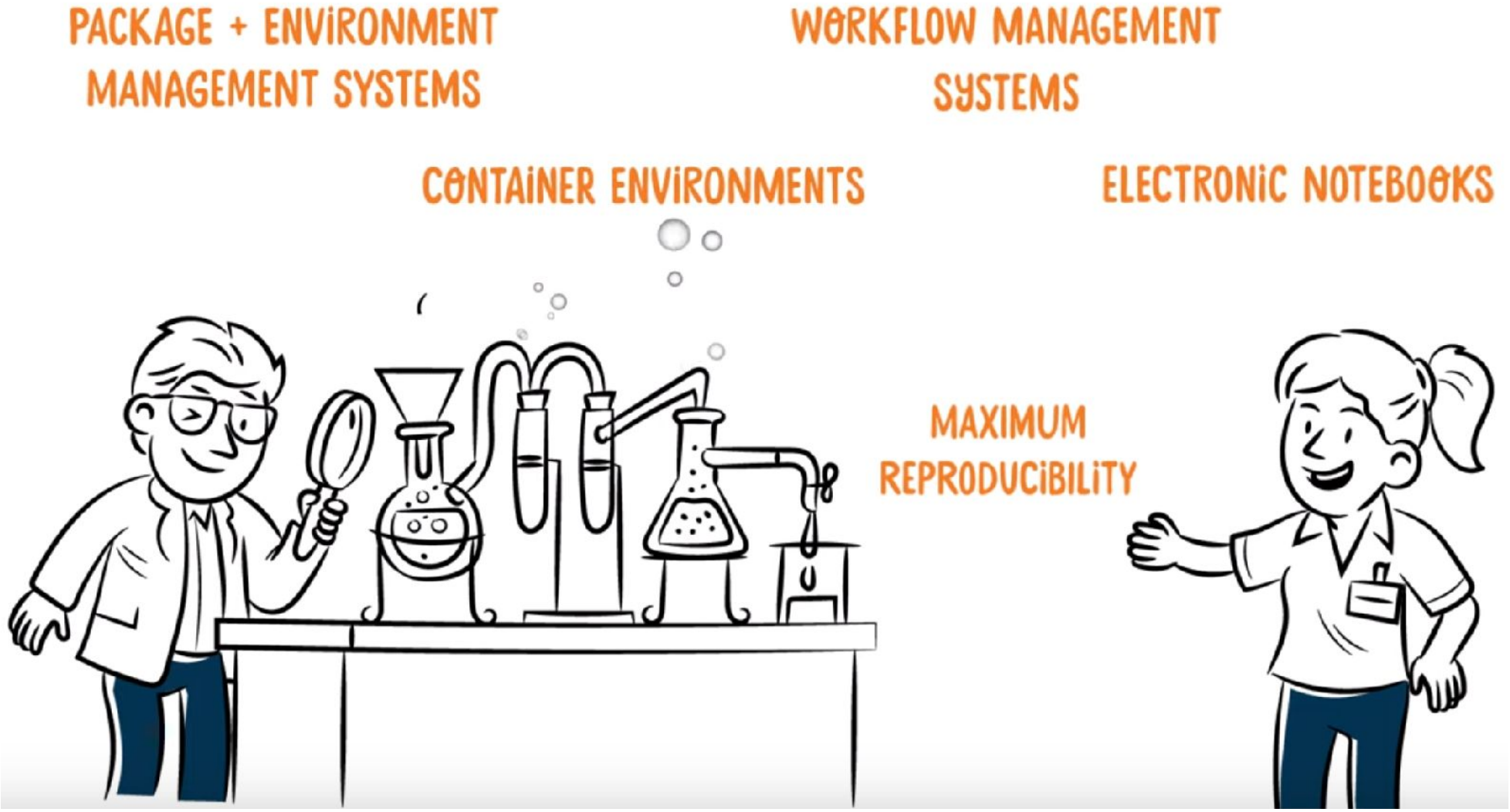
DATA LIFE CYCLE

DATA PROCESSING AND ANALYSIS



FAIR

Data Life Cycle: Process and Analyze



FAIR

Data Life Cycle: Preserve

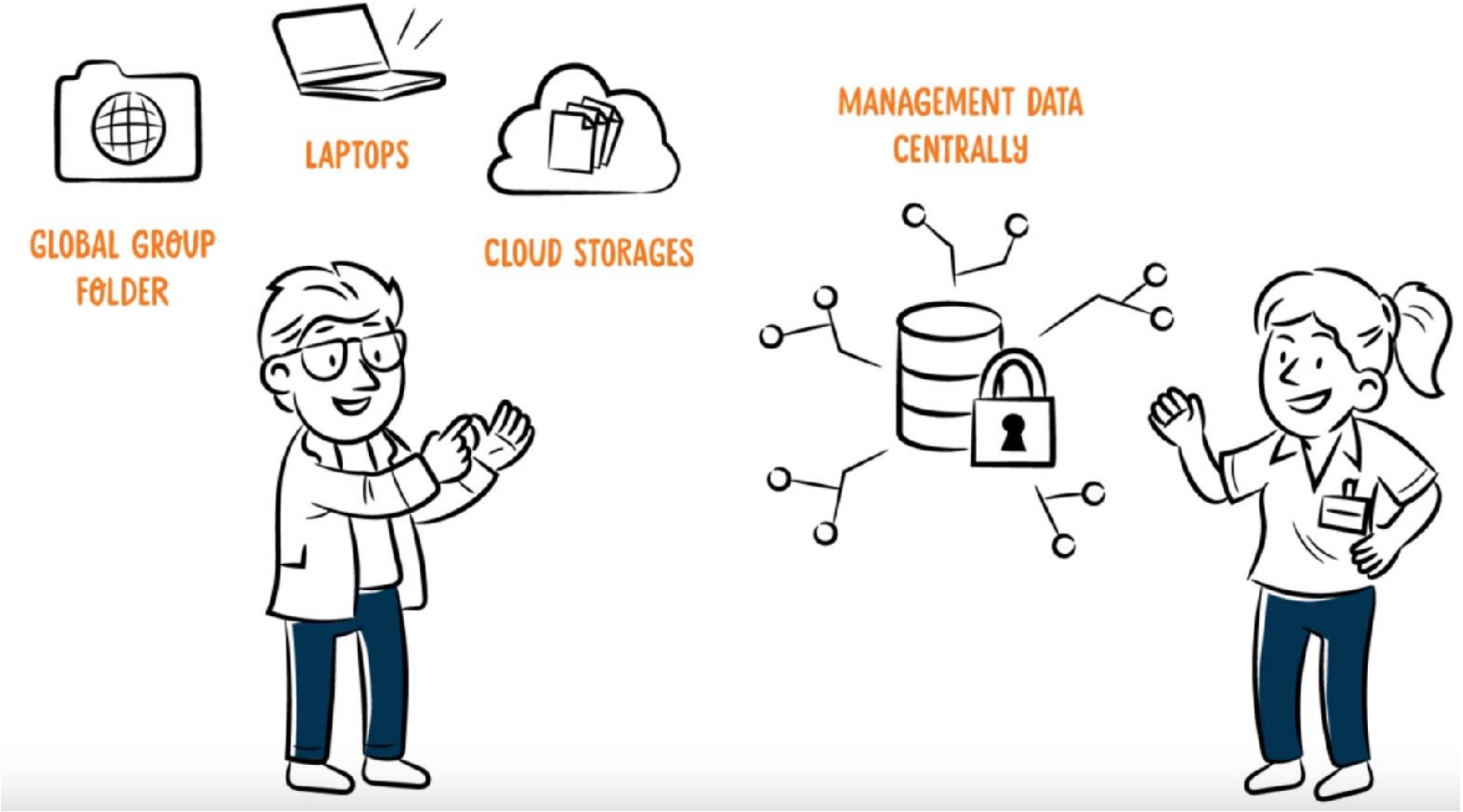
DATA SECURITY AND STORAGE

FAIR

Data Life Cycle: Preserve

FAIR

Data Life Cycle: Preserve

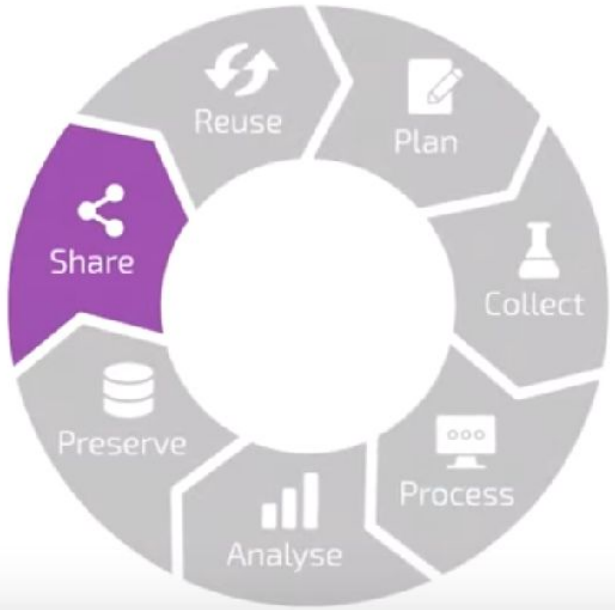


FAIR

Data Life Cycle: Preserve

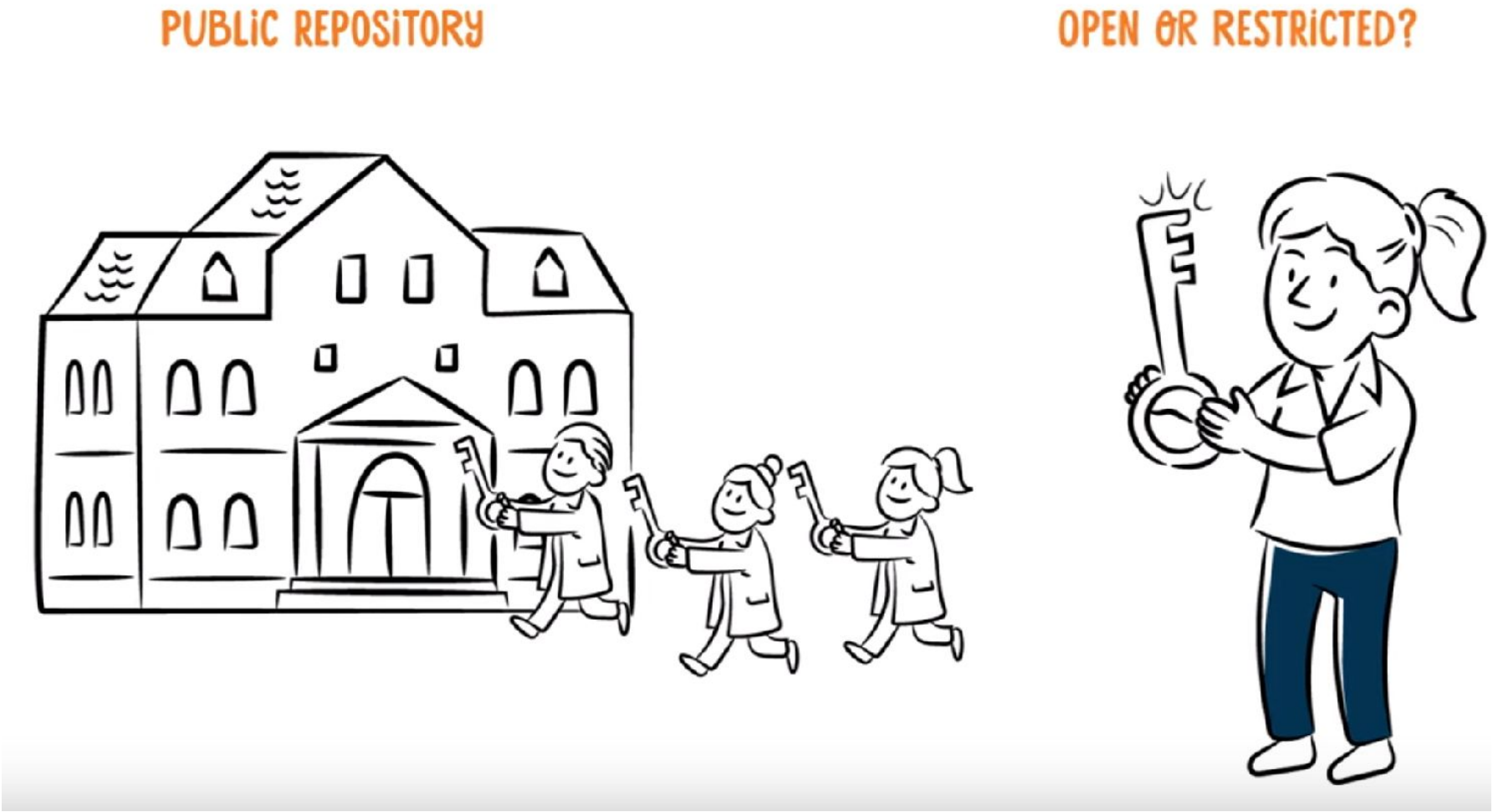
DATA LIFE CYCLE

DATA SHARING AND PUBLICATION



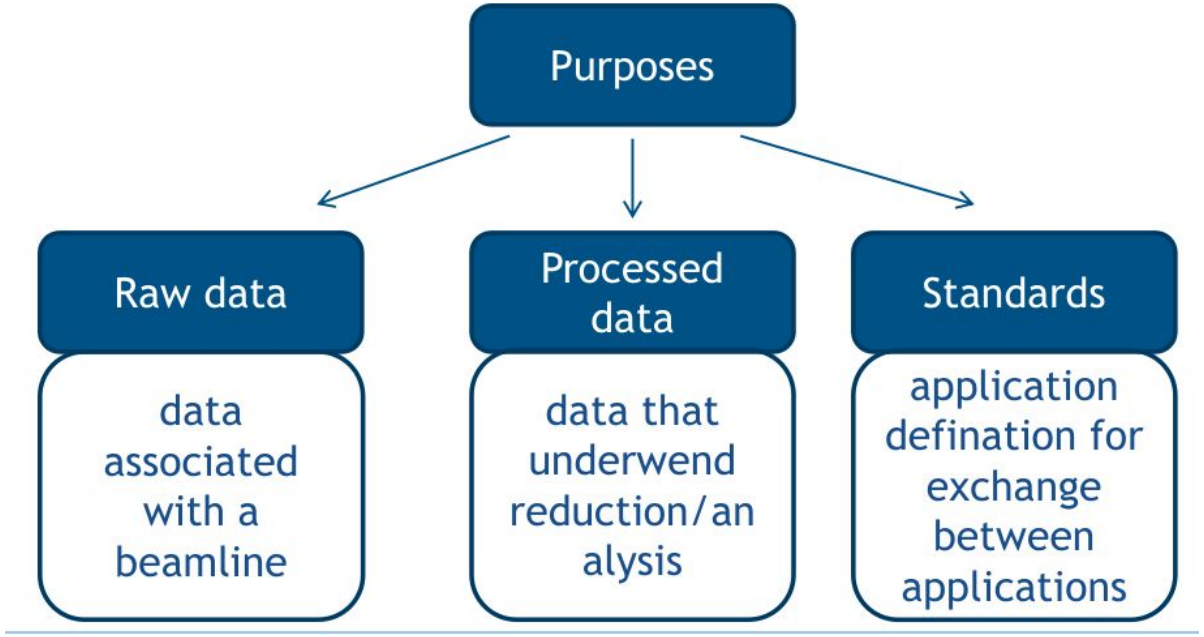
FAIR

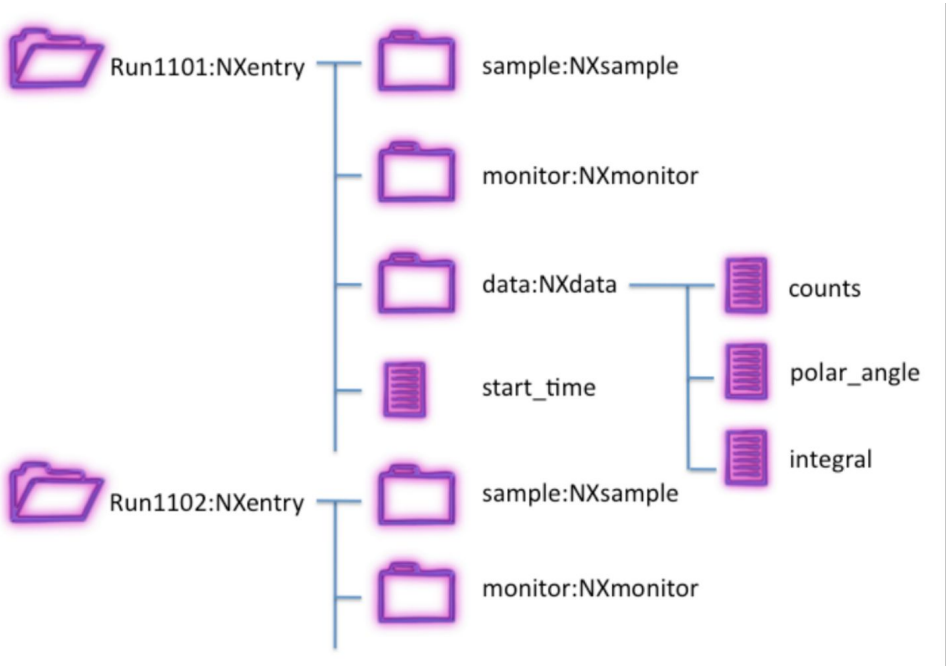
Data Life Cycle: Share



A step towards open science at BESSY II

- Metadata Standard:
  - **NeXus: A common data exchange format for neutorn, X-ray, and muon science**
  - **It is being developed as an international standard**





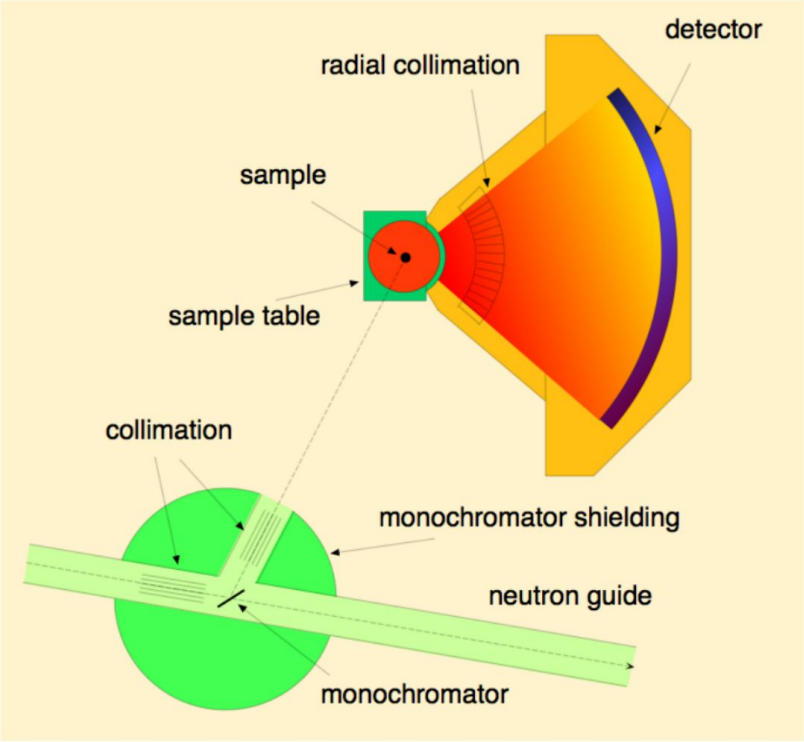
NeXus: What is it?

- Definitions: v2022.07 on 02.08.2022 (12 Release so far, the first one on 15.12.2015)

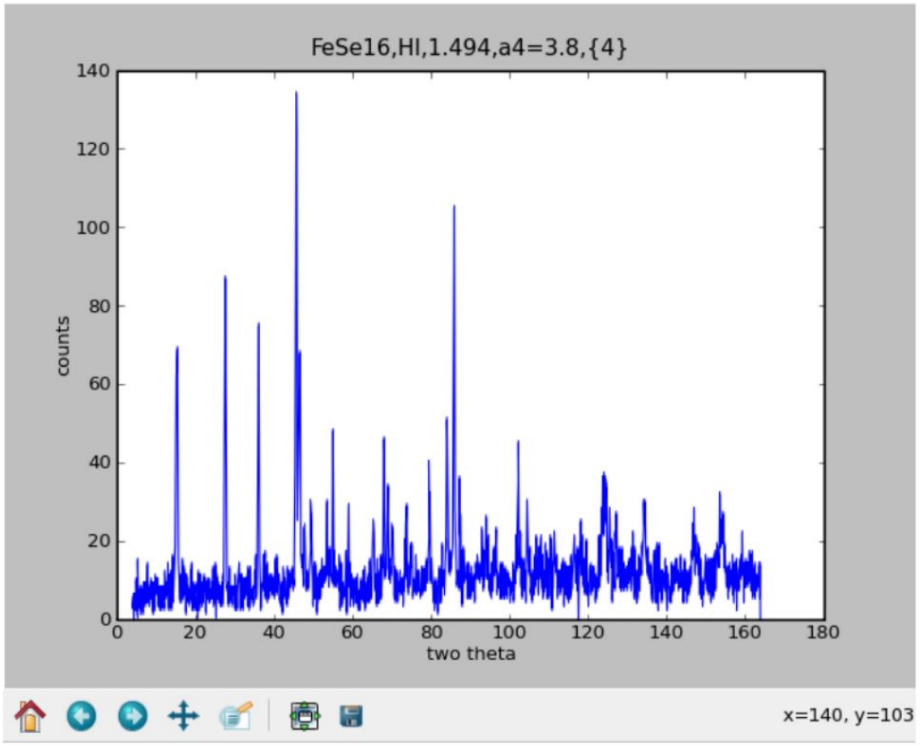
  - **Base Classes:**
    - NXmirror, NXmonochromator, NXslit, NXaperture, etc.
  - **Application Definations:**
    - A contract between a data provider (beamline control system) and a data consume
    - It makes the information is certain to be available in a data file
  - **Contributed Definations:**
    - Proposed definitions from community

How do we do it?

- WONI: **WO**nderful **N**ew **I**nstrument



The (fictional) WONI example powder diffractometer



Example Powder Diffraction Plot from (fictional) WONI at HYNES

Application Defination: Steps

Basic structure of NeXus data file

```
entry:NXentry
  NXdata
  NXinstrument
     monochromator:Nxcrystal
        wavelength
        d_spacing
        rotation_angle
        reflection
        type
  NXmonitor
  NXsample
```

- **Application Defination: An inerface for your data file**
- **Steps:**
  1. **Think**! hard about data
  2. **Map** the required fields into the NeXus hierarchy
  3. **Describe** this map in a NXDL file
  4. **Standardize** the defination through comminication with NIAC

Standardize application definition

```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="nxdlformat.xsl" ?>
<!--
# NeXus - Neutron and X-ray Common Data Format
#
# Copyright (C) 2008-2022 NeXus International Advisory Committee (NIAC)
#
# This library is free software; you can redistribute it and/or
# modify it under the terms of the GNU Lesser General Public
# License as published by the Free Software Foundation; either
# version 3 of the License, or (at your option) any later version.
#
# This library is distributed in the hope that it will be useful,
# but WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the GNU
# Lesser General Public License for more details.
#
# You should have received a copy of the GNU Lesser General Public
# License along with this library; if not, write to the Free Software
# Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA  02111-1307  USA
#
# For further information, see http://www.nexusformat.org
-->
<definition name="NXmonopd" extends="NXobject" type="group"
    category="application"
    xmlns="http://definition.nexusformat.org/nxdl/3.1"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://definition.nexusformat.org/nxdl/3.1 ../nxdl.xsd"
    >
    <symbols>
        <doc>
            The symbol(s) listed here will be used below to coordinate datasets with the sam
        </doc>
        <symbol name="i">
            <doc>i is the number of wavelengths</doc>
        </symbol>
        <symbol name="nDet">
            <doc>Number of detectors</doc>
        </symbol>
    </symbols>
    <doc>
```

- Step 3: Application definition
  - **WONI application definition (here)**

Contributing to NeXus standards

- Contribution to standards through github issues and pull requests
- You are Welcome!!!

37 Open ✓ 417 Closed | Author | Label | Projects | Milestones | Reviews | Assignee | Sort

- **Xrd 2thta fix** — #1399 by RubelMozumder was closed 2 weeks ago • Review required (1)
- **2 parse max iv proposal** ✓ — #1397 by woutdenolf was closed on Jun 30 • Review required (1)
- **NXazint contribution** — #1394 by woutdenolf was closed on Jun 28 • Draft
- **Small change to NXtransformations** ✓ — #1391 by lukaspie was closed on Jun 3 • Draft
- **Use enumerations for absorption edge and emission lines** — #1390 by mretegan was closed on May 29 • Review required
- **Add CSS tweak to hide summary when a details section is opened** ✓ — #1382 by PeterC-DLS was merged on May 3 • Approved (3)
- **Description of `NXindirecttof`** ✓ — #1378 by sona-patel was merged on Apr 5 • Approved (1)
- **Re-usable e** — #1379 opened on Apr 15 by woutdenolf (1)
- **type of `NXinsertion_device`** — #1376 opened on Apr 3 by sona-patel (5)
- **consistent NeXus namespace in xml files** — #1374 opened on Mar 13 by woutdenolf
- **NXdetector and NXdetector_channel define redundant namespace prefixes** — #1371 opened on Mar 9 by woutdenolf (1) (3)

FAIR Indicators and Tools

- **FAIR Indicators**
  - **F-UJI**
  - **FAIR-Aware**

- **Tools**
  - **Virtual Observatory**
    - IVOA is an organisation that debates and agrees the technical standards that are needed to make the VO possible
  - **HELOPORT**:
    - makes the entire life cycle of a project at the HZDR findable, accessible, interoperable and reusable
  - **NOMAD**:
    - lets you manage and share your materials science data in a way that makes it truly useful to you, your group, and the community

Acknowledgement:
Data Stewardship Wizard
Bluesky Team

Backup

Findable

- (Meta)data are assigned a globally unique and persistent identifier
- Data are described with rich metadata
- Metadata clearly and explicitly include in the identifier of the data it describes
- (Meta)data are registered or indexed in a searchable resource

Accessible

- (Meta)data are retrievable by their identifier using a standardized protocol
- The protocol is open, free and universal
- The protocol allows for authentication and authorization, as needed
- Metadata are accessible, even when the data are no longer available

Interoperable

- (Meta)data use a formal, accessible, shared and broadly applicable language
- (Meta)data use vocabularies that follow FAIR principles
- (Meta)data include qualified references to other (meta)data

Reusable

- (Meta)data are richly described with a plurality of accurate and relevant attributes
- (Meta)data are released with a clear and accessible data usage licence
- (Meta)data are associated with a detailed provenance
- (Meta)data meet domain-relevant community standards

Abstract: Introduction to FAIR data principle

The FAIR Data Principles—Findable, Accessible, Interoperable, and Reusable—provide a framework to enhance the usability and accessibility of data to maximize its value and impact in research and various applications. Adopting FAIR principles offers numerous benefits, including **improved data management**, **increased efficiency in research**, and **greater collaboration across disciplines**. Implementing these principles requires collaboration among researchers, data stewards, and institutions to develop and maintain the necessary infrastructures and practices. The FAIR principles are **not just theoretical guidelines but practical criteria** that, when followed, significantly contribute to the advancement of open science and the realization of data's full potential in driving discovery. This presentation discusses the initiatives undertaken to generate FAIR data at BESSY II.