E4.5 ELSA-Training curriculum for Data Scientists

Maria Christoforaki

Institute for Biomedical Informatics, Medical Faculty and University Hospital
University of Cologne

FAIR Data Spaces Final Event
3.12.2024

FAIR Data Spaces

Universität zu Köln

Contents

- WP description
- Landscape Analysis
- Feedback from the Community
- Results
  - Description of Curriculum
  - Publications
  - Events

AP4.5 Targets-Tasks-Deliverables

- Targets
  - Foster ethics awareness of data professionals via training
  - Prepare an ELSA training curriculum to support ethical use of research data by GAIA-X innovators
  - Exploit the outcomes of AP2 and transfer this knowledge to practising data scientists.
- Tasks
  - Landscape Analysis (UzK)
  - ELSA in Data Science Workshop Series (UzK)
  - Development of ELSA Training Curriculum (UzK)
  - Outreach and Liaise with GAIA-X partners (FhG)
- Deliverables
  - Assessment of existing approaches
  - Workshop Series
  - ELSA- Training curriculum Version 1 , 2

Landscape Analysis

Guidelines, Existing courses and the profile of the Data Scientist

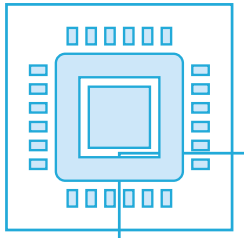
Landscape Analysis

- The guidelines issued by a variety of international organisations
- The existing ELSA training approaches, as manifested in tertiary education courses
- Construct a profile of the data scientist

Landscape Analysis-Guidelines

Visualization by Sinclair, Stéfan and Geoffrey Rockwell. *Voyant Tools*.
https://voyant-tools.org/?corpus=ad6de10ab2522553b734ca0c66036c95&panels=cirrus,reader,trends,summary,contexts

Existing ELSA Training Approaches-Teaching Scope

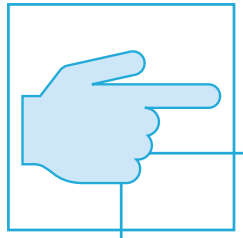


Computer Science

- Privacy
- Security
- Algorithms and inequality/justice/fairness, bias
- Transparency, accountability and responsibility
- Data validity

Professional and Business skills

- Code of Ethics from a professional body
- Codes of Conduct
- Professional Decision Processes Training
- Professional and leadership skills for computer science students, specifically in senior project classes

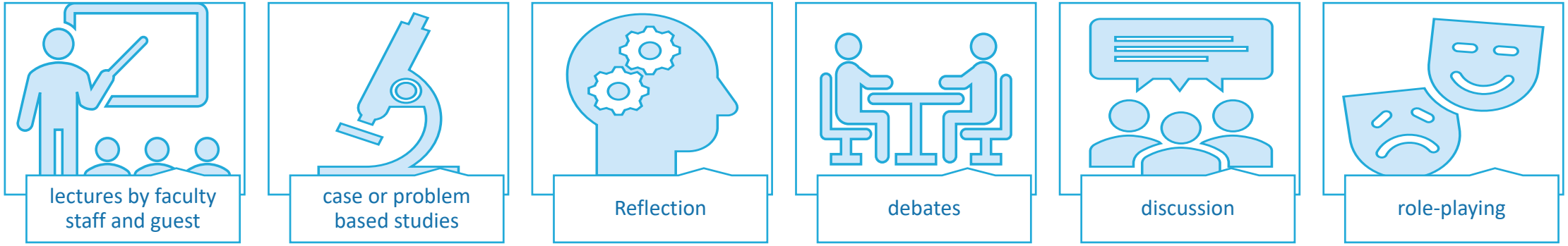


Ethical and Legal topics

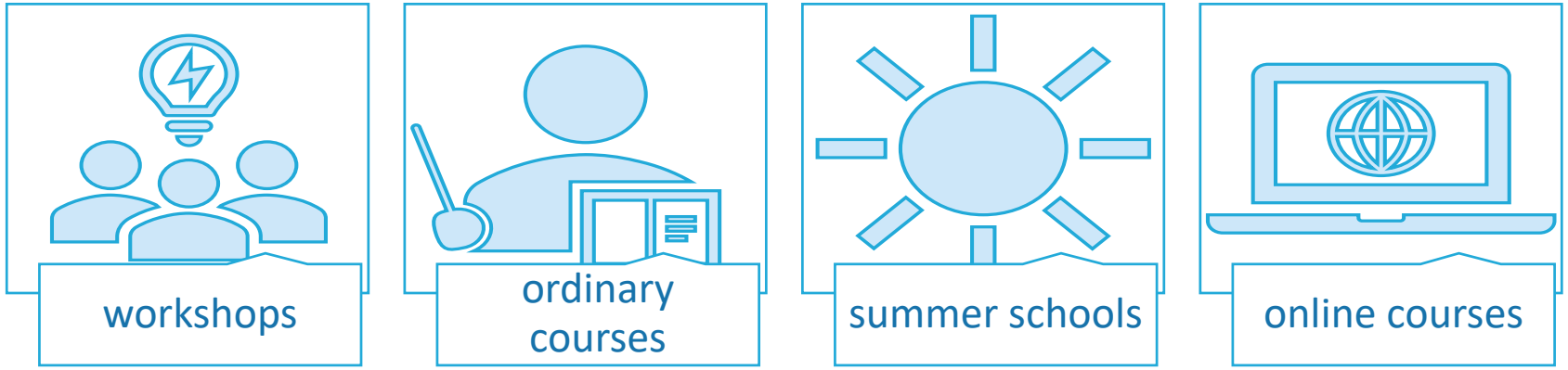
- Philosophical frameworks and moral theories
- Legal frameworks
- Human rights, risk and liabilities
- Privacy and civil liberties
- Work and labour
- Social responsibility
- Social context and stakeholders
- Ethics washing and environmental impact

Existing ELSA Training Approaches- Teaching Methods and Course Format

Methods



Format



The profile of the Data Scientist

Now:

- Mostly with tertiary education degrees
- Computer science, engineering or other relevant degrees
- Few institutions offer ELSA courses for Data Scientists, and even fewer employers demand ELSA education/experience

In the Future:

- Heterogeneous education level
- Job transitioning from other sectors
- Data scientists will not form a homogeneous group, which is the target group of today's ethics curricula

Feedback by the Community

Workshops and Surveys

Workshop Series First, Curriculum Version, and Survey Feedback

- 1st ELSA Workshop 1.6.2022
- 2nd ELSA Workshop 14.6.23
- ELSA Training Curriculum for Data Scientists Version 1.0 31.8.2023
- Feedback Survey March 2024
- Other smaller workshops and presentations 2022-2023

Feedback

- Subjects: The most popular subject is Data protection, followed by Basic legal concepts, Transparency and explicability, and Stakeholder identification
- Format of training: Workshop series on specific topics (2-3 days per topic)
- Level of knowledge offered and Target audience: depends on the format and duration
- Provider: Academic Institutions as independent programs addressed to all that are interested

ELSA Curriculum for Data Scientists

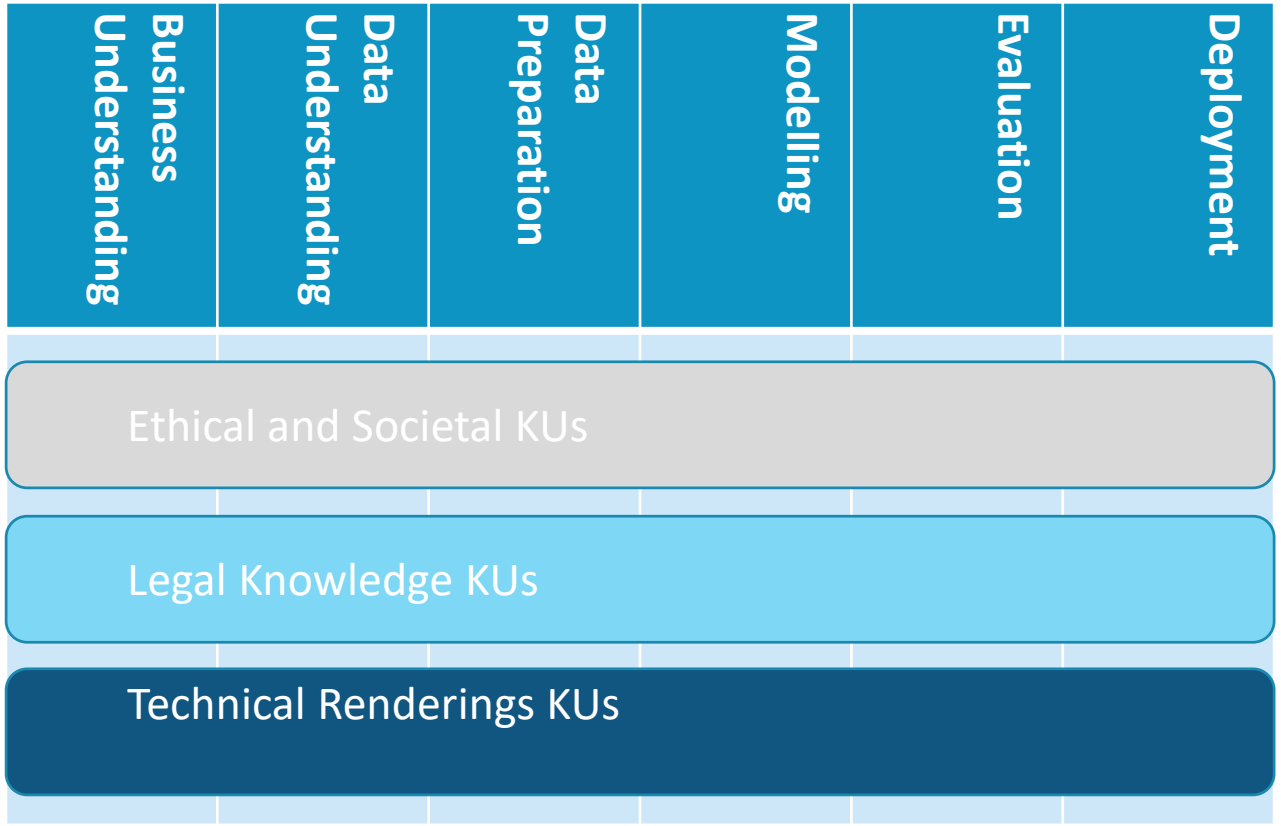
Final Version

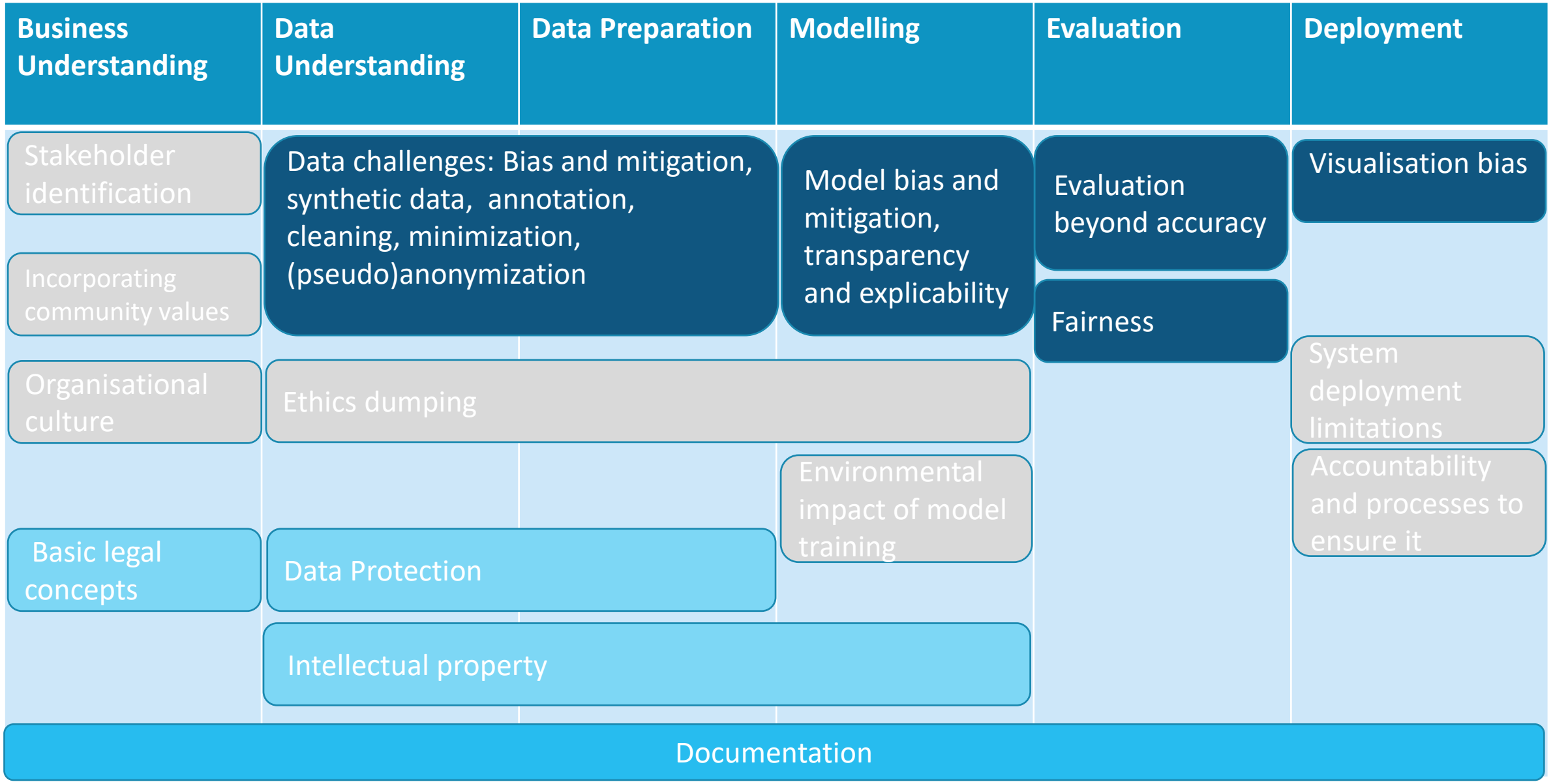
Curriculum Vision and Objectives

- Awareness of Ethical Legal and Societal Aspects of Data Science
- Establish a common language with the relevant domain experts in order to cooperate to find appropriate solutions
- Incorporate ELSA in the data science workflow as an integral part of the Data Science Project Lifecycle

Curriculum Structure

- Vertical: Modules that correspond to the CRISP-DM phases
- Horizontal: Knowledge Units (KUs) for each Module:
  - Ethical and Societal KUs
  - Legal Knowledge KUs
  - Technical Renderings KUs

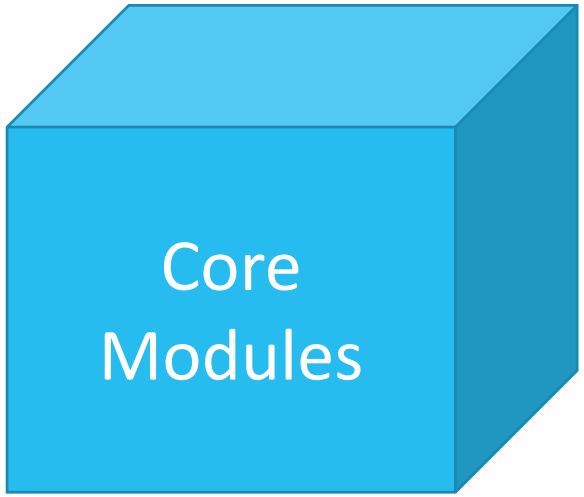




Implementation Strategies

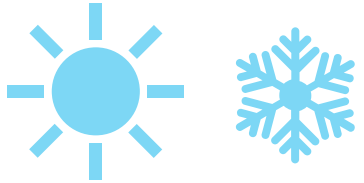
- The Curriculum, as presented, assumes a generic audience with little knowledge of ELSA - not always the case
- Implementation strategies
  - Introductory course assuming little or no knowledge of ELSA
  - Selection of modules that fit the roles of the target audience
  - Adaptation of the content according to the application domain of the target audience
  - A combination of the previous two categories

Core Modules



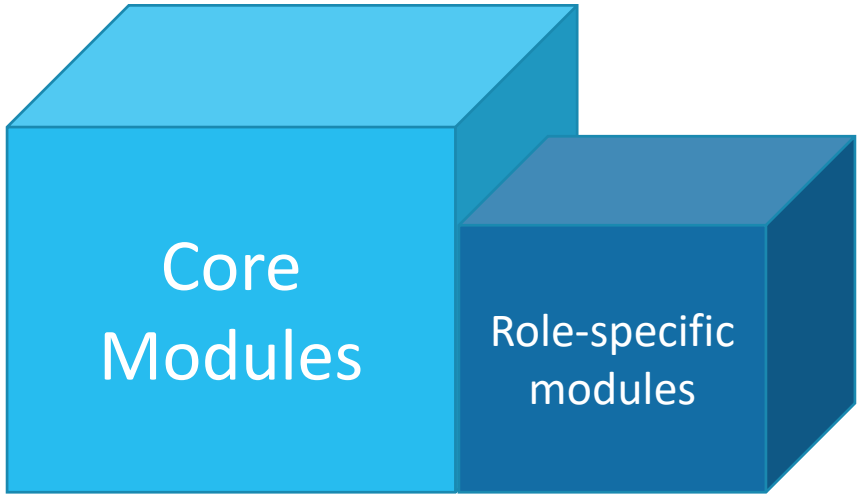
Community values
Data protection
Bias and Fairness
Documentation

1-2 weeks



Summer/Winder school

Role-specific modules

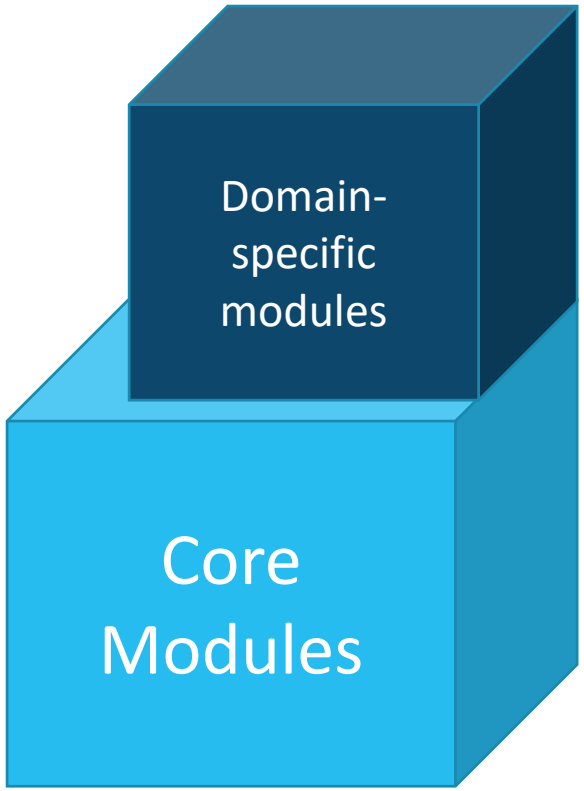


Deep dives into specific knowledge areas

2-3 days

Workshops

Domain-specific modules

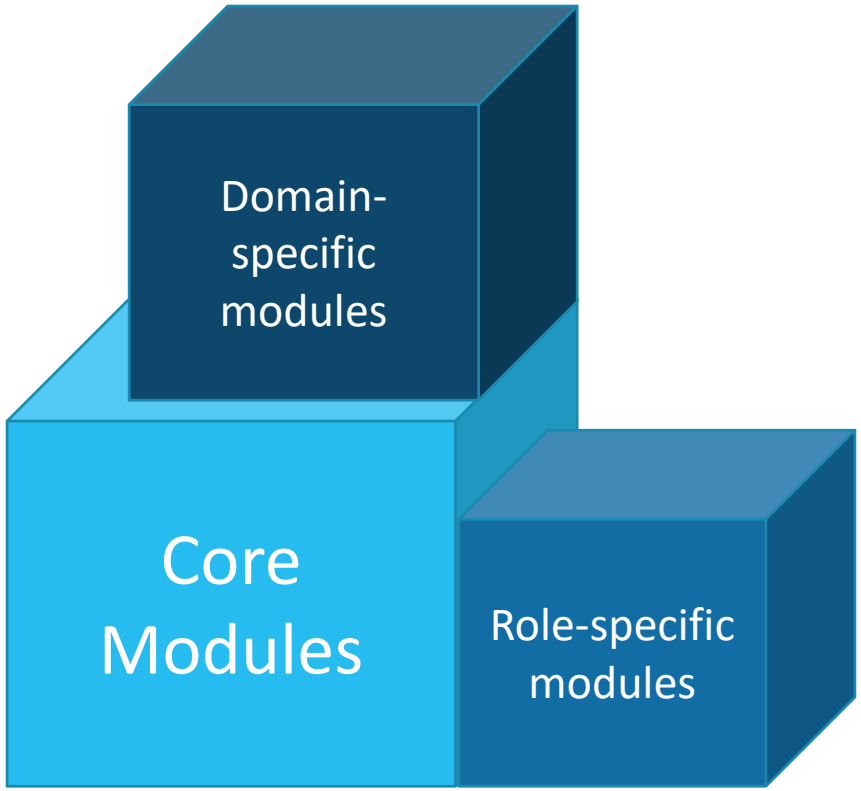


adaptation of the generic modules to the specific application domain

2-3 days

Workshops

Combination of Role- and Domain- specific modules



adaptation of the generic modules to the specific roles and application domain

2-3 days

Workshops

The FAIR Data Spaces Demonstrators as Case Studies

The FAIR Data Spaces Demonstrators

- Biodiversity (GeoEngine)
- Research data quality assurance
- Cross-platform data analytics(PHT)
- Input from
  - WP2 (“Rechtliche und ethische Rahmenbedingungen”)
    - “E2.2.1:Bericht Vertragsrechtliche Überprüfung und Compliance”, a.o.
  - WP4 Deliverables and Publications

Biodiversity

- Geo Engine:
  - supports interactive data analyses for geodata (such as vector and raster data)
  - allows data scientists to focus on the actual data analyses rather than data preparation
- Not aimed to be used with personal data, but it can exemplify concerns on the collection and processing of personal data
- Ethical concerns: the use of geolocation wrt bias and stigmatisation of populations and regions

Data Validation and Quality Assurance

- Performs automated quality control and data assurance within a commonly available or easily provided environment.
- Can be used to explain the concepts copyright, trade secret and patent
- FAIR vs FRAND principles

Cross-Platform FAIR Data Analysis

- PHT(Personal Health Train) operating on Health care data
- Employs federated learning
  - ensures privacy and data governance by keeping the cloud-based data spaces separate from hospital data silos only accessed by local clients
- Can be used to exemplify
  - the concept of personal and sensitive data, an example of privacy by design architecture
  - the issues posed by opaqueness regarding the data, e.g. bias

Publications and Events

Publications

- Papers:
  - Christoforaki, Maria, and Oya Beyan. 'AI Ethics—A Bird's Eye View'. *AI Ethics—A Bird's Eye View* 12, no. 9 (20 April 2022): 4130. https://doi.org/10.3390/app12094130.
  - Christoforaki, Maria, and Oya Deniz Beyan. 'Towards an ELSA Curriculum for Data Scientists'. *AI* 5, no. 2 (11 April 2024): 504–15. https://doi.org/10.3390/ai5020025.
- Reports
  - Christoforaki, Maria. 'ELSA Training for Data Scientists-Describing the Landscape'. FAIR-DS Project Deliverable. Cologne: UzK, 31 December 2021. https://doi.org/10.5281/zenodo.7233569.
  - Christoforaki, Maria. 'ELSA Training Curriculum for Data Scientists - Version 1.0'. FAIR-DS Project Deliverable. Cologne: UzK, 31 August 2023. https://zenodo.org/record/8318726.
- Presentations
  - Christoforaki, Maria. 'Data Scientists and ELSA:Describing the Landscape'. Presented at the 1st FAIR Data Spaces ELSA Workshop, 1 June 2022. https://doi.org/10.5281/zenodo.6637759.
  - Christoforaki, Maria. 'Towards an ELSA Curriculum for Data Scientists'. Presented at the 1st FAIR Data Spaces ELSA Workshop, 1 June 2022. https://doi.org/10.5281/zenodo.6637778.
  - Christoforaki, Maria. 'A First Approach Towards an ELSA Curriculum for Data Scientists'. Presented at the 2nd FAIR Data Spaces ELSA Workshop, 16 June 2023. https://doi.org/10.5281/zenodo.8046261.

Events

- Workshop Organization
  - 1st FAIR Data Spaces ELSA Workshop, "*Towards creating an ELSA Curriculum for Data Scientists*", Wednesday, 1 June 2022, online, https://www.nfdi.de/1-elsa-workshop/
  - 2nd FAIR Data Spaces ELSA Workshop, "*A first approach to an ELSA Curriculum for Data Scientists- The FAIR Data Spaces Project as a Use Case*", Wednesday, 14 June 2023, online, https://www.nfdi.de/2-elsa-workshop/
  - videos and slides for both Workshops are published in the FAIR Data Spaces Community in Zenodo: https://zenodo.org/communities/fair-ds
- Participation in Workshops
  - Christoforaki, Maria, "An ELSA curriculum for Data Scientists", BI-K, MeDIC Retreat, 14.9.22, Winningen
  - ELSA Training –Feedback from the Community Survey, Horizon, Europe Project Shift-Hub Open Workshop, 13.6.2023
- Talks
  - Christoforaki, Maria, "Towards an ELSA curriculum for Data Scientists-A first Approach", Data Driven Medical Research (DDMR) Seminar Series, 20.4.2023, Cologne
  - Christoforaki, Maria, "Towards an ELSA curriculum for Data Scientists", Research Data Alliance (RDA) Plenary 21, 23 – 26 October 2023, Salzburg, Austria

Thank you for your interest!

Contact: Maria Christoforaki, maria.christoforaki@uk-koeln.de

Stay in touch:

www.nfdi.de/fair-data-spaces

@FAIRDataSpaces

#FAIRDataSpaces

Community newsletter: https://www.nfdi.de/newsletter-abo/

FAIR Data Spaces Community in Zenodo:

https://zenodo.org/communities/fair-ds

FAIR Data Spaces