Shared Development Roadmap 2022-2023

Introduction

Clariah

Purpose

- Bring all the work from CLARIAH Core & Plus & WPs together in the CLARIAH Infrastructure
- From CLARIAH Infrastructure “**as a concept**” towards a CLARIAH Infrastructure as a “**tangible product**” for scholars to use
  - Either *DIRECTLY*
  - Or *INDIRECTLY* providing key components that are needed for scholarly applications

Clariah

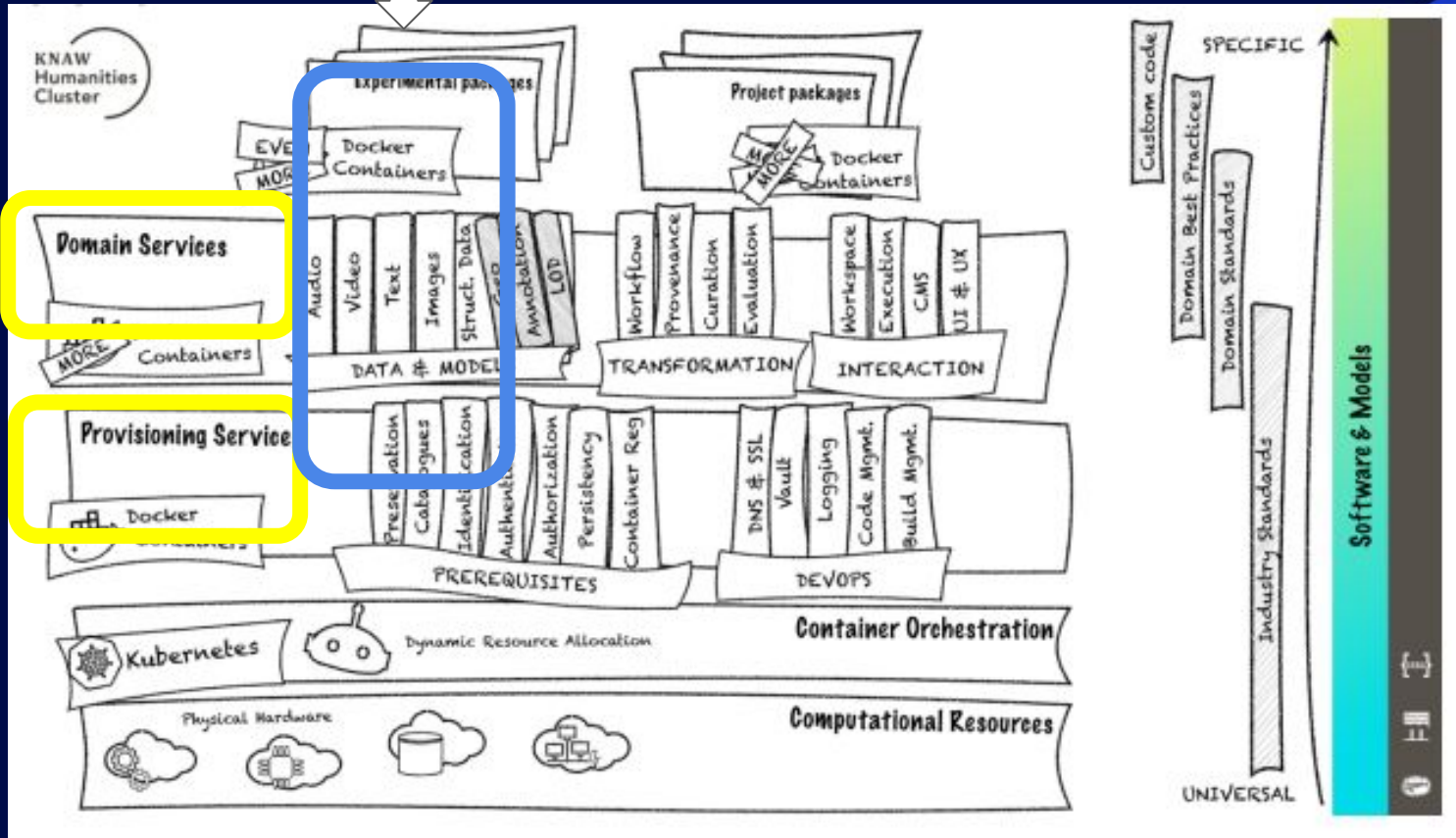
History

- 2021 Q3:
  - List of all CLARIAH tools and resources
  - List of practical use cases from researchers
- 2021 Q4:
  - Discussions on most important building blocks that CLARIAH should take up in last 2 years
  - Iterations of “CLARIAH Shared Development Roadmap” (e.g., “hay-days”)
  - Approval by the CLARIAH Board
- 2022 Q1+2:
  - START WORK ON 7 PROJECTS

Clariah

CLARIAH Shared Core Services

data



data

Use cases

As a scholar, I want to have an overview of datasets providing me with information from collection metadata, including the dataset's **distribution** (e.g. SPARQL endpoint, full text search API endpoint, RDF or CSV data dump file), the **organization** that publishes the dataset and the **license** under which it is published

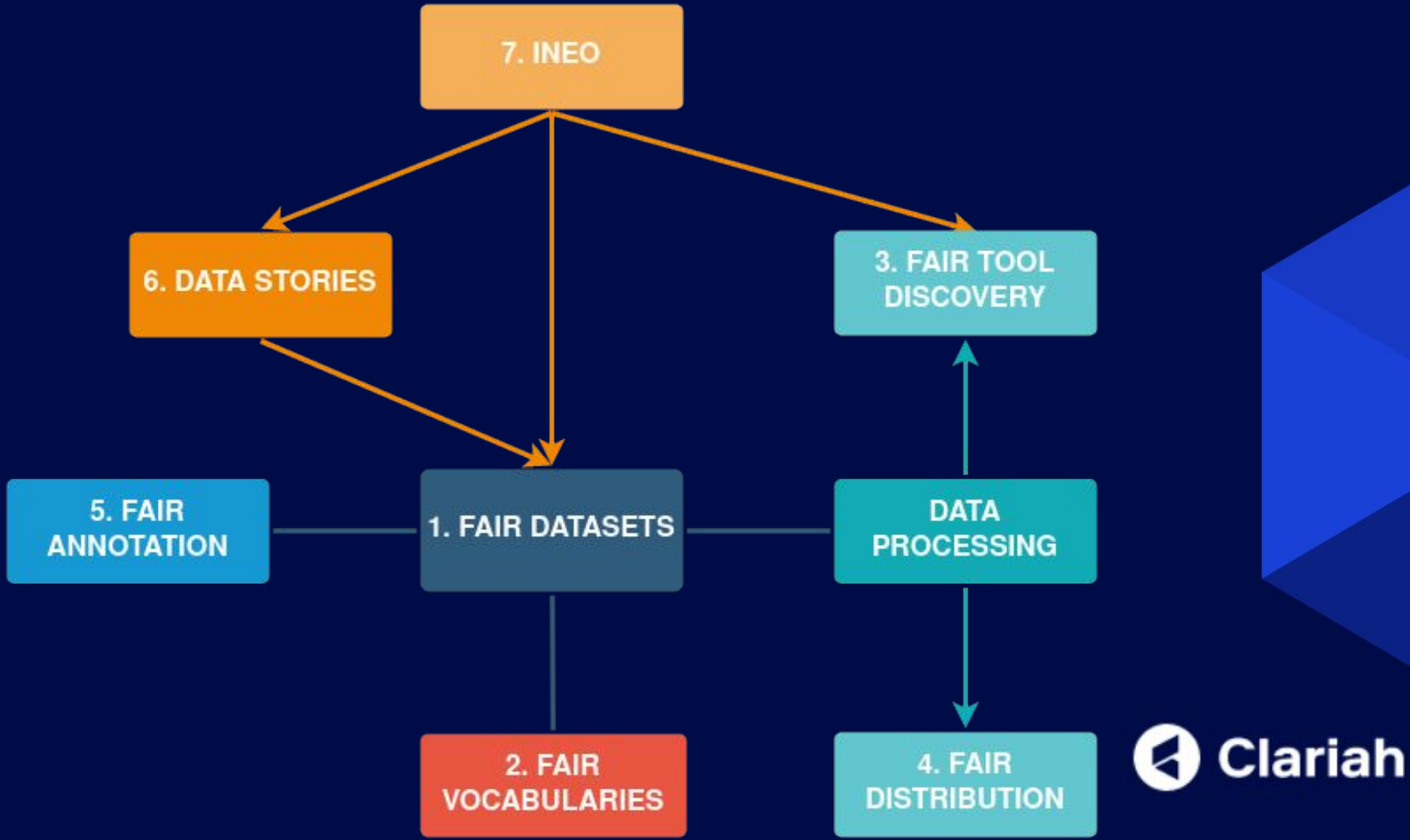
in order to select interesting data sets of my research and access them either via domain portals or central services, by downloading the content myself, or by visiting an organisation.

Clariah

Main Topics & FAIR

1. Data
2. Vocabularies (“interoperable”)
3. Tools
4. Annotation
5. Computational approach (“data stories”)
6. Infrastructure (“technical needs”)
7. Community portal

Clariah


7. INEO
6. DATA STORIES
3. FAIR TOOL DISCOVERY
5. FAIR ANNOTATION
1. FAIR DATASETS
DATA PROCESSING
2. FAIR VOCABULARIES
4. FAIR DISTRIBUTION
Clariah

Main Topics & FAIR

1. **Data** Registry: make data findable, accessible via domain portals
2. **Vocabulary** Registry: interoperability, recommendation, alignment
3. **Tool** Registry, incorporate info on TRL, SRL
4. Portable **annotation** to save collections and annotations cross-platforms
5. Data Stories → Computation, Visualization & Storytelling
6. Distribution & Deployment → Collaborate on infrastructure level
7. INEO as Community portal

Clariah

1. FAIR Datasets (Data Registry)

EXAMPLE

Coordinator: Menzo Windhouwer & Femmy Admiraal
Work packages involved: WP3, WP4, WP5, WP6 & WP2
Github projects link: https://github.com/orgs/CLARIAH/projects/2/

- **Make datasets *findable* (metadata) via a Data Registry (linked to INEO)**
- **Make datasets *accessible* (resource) via domain portals (Media Suite, Linguistic Corpus Search, Nederlab, etc.)**
- **Stimulate *interoperability* and *reusability* via requirements for the registry using a “star model” (Data Readiness Level)**

1. FAIR Datasets (Data Registry)

EXAMPLE

User story:

As a scholar, I want to have an overview of datasets providing me with information from collection metadata, including the dataset's distribution (e.g. SPARQL endpoint, full text search API endpoint, RDF or CSV data dump file) where the dataset is distributed, the organization that publishes the dataset and the license under which it is published, **in order to** select interesting data sets of my research and access them either via domain portals or central services, by downloading the content myself, or by visiting an organisation.

New:

- **Automatic Harvesting Service (DCAT, schema.org, NDE Register, partner endpoints, PID inventory)**
- **Search interface with high-level integration in INEO**
- **Additional WP effort on dataset level!**
- **CDO + IG Data Curation**

Legacy

- **Use existing knowledge & code (VLO, NDE register, ODISSEI data registry, EOSC)**
- **Existing metadata curation activities (WP3, WP5, MPI, RU, DANS)**

Organisation of work 2022-2023

- Shared Development Roadmap is on GitHub:

  → https://github.com/CLARIAH/clariah-plus

- Cross-WP Collaboration
- Infrastructure integration coordinated by WP2 (Thomas)
- 7 coordinators of SDR projects
- Monthly onsite planning/discussion sessions (Tech Days)
- Quarterly demo sessions to all CLARIAH on work in progress
- Testing with user community!!

Clariah