

# CLARIAH Heden & Toekomst: Shared Development Roadmap

# Huidige Situatie

**Doel:** We bouwen een **shared common infrastructure** (for the humanities)

Maar...

- ▶ In hoeverre kunnen we daar nu van spreken?
- ▶ **Realiteit:** meerdere onafhankelijke infrastructuren
- ▶ Wat betekent het om deel uit te maken van de CLARIAH infrastructuur / CLaaS ?
  - ▶ Welke eisen stelt CLARIAH?

## Huidige situatie (2)

- ▶ Gebrek aan interoperabiliteit, ook *binnen* WP3
- ▶ Gebrek aan duidelijk vastgelegde afspraken over interoperabiliteit (specificaties/requirements).
- ▶ Te weinig samenwerking tussen WP's
- ▶ Moeilijk overzicht te krijgen, veel legacy binnen WP3 en vanuit CLARIN
- ▶ Dubbel/conflicterend werk (ook binnen WP3)

## Gewenste situatie

- ▶ **Harmonisatie:** Meer interoperabiliteit tussen tools en services
  - ▶ Meer samenwerking, minder dubbel werk
- ▶ Meer **transparantie**; zicht op wat er gaande is en de status -> **CLARIAH PLUS work plan**
  - ▶ WP3 Task Descriptions -> **CLARIAH PLUS task descriptions**
  - ▶ Welke use cases we bedienen -> **CLARIAH PLUS use cases**
- ▶ **Richtlijnen:** Duidelijke vastgelegde (en gecommuniceerde!) afspraken

## Gewenste situatie: Duidelijke richtlijnen

- ▶ Technische richtlijnen:
  - ▶ Deelname aan de infrastructuur: **CLARIAH Infrastructure Requirements**
  - ▶ Software ontwikkeling, kwaliteit & duurzaamheid: **CLARIAH Software Requirements**
  - ▶ Verdere interoperabiliteit? CLARIAH Interoperability Requirements?
- ▶ Documentatie: **CLARIAH Documentation Guidelines**
- ▶ Welke services we willen bieden en hoe we daarin samenwerken: **CLARIAH Shared Development Roadmap**
- ▶ Gemeenschappelijk instappunt voor gebruikers: **Ineo**

# Interest Groups

- ▶ Text, AV, Annotation, LoD, DevOps, UI/UX, Workflows, ad-hoc . . . .
- ▶ Mensen bij elkaar brengen rondom gemeenschappelijke thema's ipv werkpakketten
- ▶ Loopt erg stroef (weinig animo/tijd)
- ▶ Technical Advisory Committee

# Shared Development Roadmap: Introductie

Overkoepelende doelstellingen:

- ▶ Concreet maken wat CLARIAH PLUS gaat opleveren in 2022-2023
- ▶ Richting bepalen voor eventueel vervolgproject
- ▶ Focus op generieke infrastructuur en cross-WP samenwerking
- ▶ Input voor board om budget op aan te passen

# Shared Development Roadmap: Introductie (2)

Wat is de Shared Development Roadmap (SDR)?

- ▶ Een overzicht van generieke **CLARIAH services** geformuleerd vanuit een gebruikersperspectief (high-level)
  - ▶ (bestaande danwel potentieel toekomstige)
- ▶ Een **CLARIAH service** maakt een 'workflow' van een onderzoeker mogelijk (analyse, datatransformatie, presentatie)
  - ▶ Service is hier geen technisch begrip!
  - ▶ Generiek geformuleerd vanuit het perspectief van een onderzoeker: **user stories**
  - ▶ Een CLARIAH service kan meerdere *implementaties* hebben
  - ▶ Een implementatie bestaat meestal uit meerdere software- en data-componenten (stand-off beschreven)
  - ▶ We richten ons op **minimaal vatbare** en **generieke/pervasieve** CLARIAH services en componenten



# Shared Development Roadmap: Doelstellingen

- ▶ **Harmonisatie** van verschillende oplossingen binnen CLARIAH; welke zijn volwassen en hebben potentie? Welke kunnen afvallen?
- ▶ **Planning**; Wat mist nog? Waar willen we heen in de toekomst?
- ▶ **Transparantie**; 'the big picture'
- ▶ **Interoperabiliteit en hergebruik** van software/data bevorderen
- ▶ **Evaluatie** van CLARIAH services en componenten:
  - ▶ Technology Readiness Level (TRL)
  - ▶ Compatibility Level (CL)
  - ▶ Stakeholder Readiness level (SRL)
  - ▶ Data Readiness Level (DRL)

# Shared Development Roadmap: TRL

TRL	Description	Stage
0	<b>Idea</b> - Unproven, untested and largely unformulated concept	<b>Planning (pre-alpha)</b>
1	<b>Initial Research</b> - Basic (scholarly) needs observed and reported	
2	<b>Concept Formulated</b> - Initial technology/application has been concept formulated	
3	<b>Proof of Concept</b> - Initial Proof-of-concept of key functionality . Concept presented for initial feedback from scholarly users. Not yet validated and not suitable for end-users yet.	<b>PoC (alpha)</b>
4	<b>Validated PoC</b> - Validated Proof-of-concept of key functionality. Technology validated in its own experimental setting (e.g. the lab). Not mature enough for end-users yet.	
5	<b>Early Prototype</b> - Technology validated in target setting (e.g. with potential end-users)	<b>Experimental (beta)</b>
6	<b>Late Prototype</b> - Technology demonstrated in target setting, end-users adopt it for testing purposes.	
7	<b>Release Candidate</b> - Technology ready enough and in initial use by end-users in intended scholarly environments. Further validation in progress.	
8	<b>Complete</b> - Technology complete and qualified, released for all end-users in scholarly environments.	<b>Production (stable)</b>
9	<b>Proven</b> - Technology complete and proven in practice by real users.	

# Shared Development Roadmap: CL

CL	Description
<b>A</b>	<b>Excellent</b> - Technology adheres to as-good-as all posited infrastructure and software requirements.
<b>B</b>	<b>Good</b> - Technology adheres well to the requirements, there only some minor lapses
<b>C</b>	<b>Adequate</b> - Technology adheres to a sufficient amount of requirements, but some major ones are lacking.
<b>D</b>	<b>Lacking</b> - There are too many major requirements that are not met
<b>E</b>	<b>Bad</b> - Many requirements are not met.
<b>F</b>	<b>Unacceptable</b> - Technology violates or is completely dismissive of most requirements. It can not possibly be accepted without drastic changes.

Figure 1: Compatibility Level

# Shared Development Roadmap: SRL

We use the **Stakeholder Readiness Level (SRL)**, a measure that defines the user readiness of a new service to be used by scholars. This measure can be used for example to prioritize development using criteria such as:

- ▶ **Value:** the added value of the service for scholars (1-10)
- ▶ **Support/Commitment:** the enthusiasm in the community to adopt the service (1-10)
- ▶ **Cost:** costs for development but also cost involved for using (1-10)
- ▶ **Adaptability:** the level of adaptability in existing work processes (1-10)
- ▶ **Risks:** an assessment of the risks and their manageability that are involved in using the service (1-10)

# CLARIAH Services vanuit WP3 (1)

- ▶ OCR/HTR: *PICCL/TICCL*
- ▶ Corpus Search: Text & Annotations: *Blacklab, corpus-frontend, AutoSearch, OpenSoNaR, GreTeL, PaQu*
- ▶ Manuele linguïstische annotatie: *FLAT*
- ▶ Automatische linguïstische verrijking (NLP): *Frog, Alpino, UD-Pipe Frysk, DeepFrog*
  - ▶ voor Nederlands
  - ▶ verrijking historisch Nederlands

## CLARIAH Services vanuit WP3 (2)

- ▶ Lexicon service:
  - ▶ Dialectwoordenboeken: *WLD, WBD, WGD, WALD*
- ▶ Dataconversie: *Piereling, OpenConvert*
- ▶ Spraakherkenning:
  - ▶ Voor Nederlandse dialecten: *ASTA*
  - ▶ Vanuit Stichting Openspraak: *Nederlandse ASR (oral history)*
- ▶ Audio-acquisitie in surveys: *SPAQ*
- ▶ FAIR Vocabularies
- ▶ Linguistic

# Oproep

Oproep aan alle WP3 deelnemers:

- ▶ Denk/werk mee aan de software/infrastructure requirements
- ▶ Denk/werk mee aan de shared development roadmap

# Links

- ▶ CLARIAH Shared Development Roadmap
- ▶ CLARIAH Infrastructure & Software Requirements
- ▶ CLARIAH Werkplan