

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 ?

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 **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**
- ▶ Duidelijke vastgelegde (en gecommuniceerde!) **afspraken**

Gewenste situatie: Duidelijk vastgelegde afspraken

- ▶ 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**
- ▶ Instappunt voor gebruikers -> **Ineo**

Interest Groups

- ▶ Text, Annotation, LoD, DevOps, UI/UX, Workflows
- ▶ 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)
 - ▶ Een 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?
- ▶ **Transparantie**; 'the big picture'
- ▶ **Interoperabiliteit en hergebruik** van software/data bevorderen
- ▶ **Evaluatie** van 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	Idea - Unproven, untested and largely unformulated concept	Planning (pre-alpha)
1	Initial Research - Basic (scholarly) needs observed and reported	
2	Concept Formulated - Initial technology/application has been concept formulated	
3	Proof of Concept - 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.	PoC (alpha)
4	Validated PoC - 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	Early Prototype - Technology validated in target setting (e.g. with potential end-users)	Experimental (beta)
6	Late Prototype - Technology demonstrated in target setting, end-users adopt it for testing purposes.	
7	Release Candidate - Technology ready enough and in initial use by end-users in intended scholarly environments. Further validation in progress.	
8	Complete - Technology complete and qualified, released for all end-users in scholarly environments.	Production (stable)
9	Proven - Technology complete and proven in practice by real users.	

Shared Development Roadmap: CL

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

Figure 1: CL

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, GreTeL, PaQu)
- ▶ Manuele linguïstische annotatie: (FLAT)
- ▶ Automatische linguïstische verrijking: (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
- ▶ Audioacquisitie in surveys: (SPAQ)
- ▶ Relation Registry

Links, feedback & samenwerking

- ▶ CLARIAH Shared Development Roadmap:
https://docs.google.com/document/d/1dCTK5w9jJRKlQuQ9t_xl7Yb
- ▶ CLARIAH Infrastructure & Software Requirements:
<https://github.com/CLARIAH/IG-DevOps/issues/4>
- ▶ CLARIAH Werkplan:
https://docs.google.com/spreadsheets/d/1WTbtA20vpKz5Oo_EnDYe