

State of the technical infrastructure

Dieter Van Uytvanck

Technical Director CLARIN ERIC

dieter@clarin.eu

CLARIN Annual Conference 2019

Leipzig

2 October 2019



An overview of the technical infrastructure highlights



Short round-up 2019: Centres

- All newly (re)certified centres now in sync with the CoreTrustSeal (validity of 3 years)
- 1 new B-centre: CELR (EE)
- Re-assessment of 9 centres: ASV (DE), BBAW (DE), EKUT (DE), IDS (DE), IMS (DE), LINDAT (CZ), MPI (NL), UCPH (DK), UdS (DE)
- 7 more assessments pending
- in total 22 certified B-centres, 57 registered centres

Short round-up 2019: Federated login

- New countries in Service Provider Federation: South Africa and Cyprus
- Total number of organisations that can login: **1870**
- On average **2082 logins per month** via the central discovery service
- Major upgrade of the CLARIN Identity Provider, paving the way for future improvements



Short round-up 2019: VLO

- Two releases (4.6 and 4.7) brought
 - collapsing of similar records
 - link checking
 - better visibility of contributing centres



Short round-up 2019: Curation (1)

- The completely revamped Curation Module provides
 - a faster and more complete overview
 - insight in the metadata provided by the CLARIN centres
- It is now actively used during the centre assessments
- www.clarin.eu/curation

Curation Module 3.1.1

| | | |
|--|-----------------------|-------|
| Validator | license | 96.3% |
| | availability | 96.7% |
| | facet-coverage: 71.4% | |
| Profiles | | |
| Collections | | |
| Link Checking Statistics | | |
| Help | | |
| <h3>ResourceProxy Section</h3> | | |
| Total number of resource proxies: 6611039 | | |
| Average number of resource proxies: 454.46 | | |
| Total number of resource proxies with MIME: 6611036 | | |
| Average number of resource proxies with MIME: 454.46 | | |
| Total number of resource proxies with reference: 6611039 | | |
| Average number of resource proxies with references: 454.46 | | |
| <h3>XML Validation Section</h3> | | |
| Number of Records: 14501 | | |
| Number of valid Records: 14491 | | |
| Ratio valid Records: 99.9% | | |

Short round-up 2019: Curation (2)

- Link check results are now also integrated into the VLO

The screenshot displays a record page with tabs for 'Record details', 'Links (2)', 'Availability', 'All metadata', and 'Technical details'. The 'Links (2)' tab is active, showing a table with two entries:

| Name | Type |
|---|--------------|
| lat:1839_00_0000_0000_0009_295C_5 | landing page |
| content | audio file |

Below the table, the 'Resource details' section is expanded, showing:

- Mime type:** audio/x-wav
- Link:** <https://hdl.handle.net/1839/00-0000-0000-0009-295D-C>

An 'Automatic link checking' section is highlighted in yellow, containing a warning triangle icon and the text: **Authentication and/or special permissions may be required in order to access the resource at this location.**

Below this, the 'Result of automatic checking' is shown as: access to the resource is restricted (401 Unauthorized)

The 'Last check' timestamp is: 28 August 2019 17:37

Short round-up 2019: Content Search

- New
 - version of the protocol and search engine
 - endpoints by IvdNT and ILC4CLARIN
- Currently there are **15 centres** with an endpoint
 - 9 have at least one version 1 endpoint
 - 9 have at least one version 2 endpoint



Short round-up 2019: Virtual Collection Registry

- New version with
 - Improved user interface
 - Better connectivity to the switchboard



Short round-up 2019: Switchboard

- Major overhaul of the back-end lead to a more robust and stable version
- Updated integration into the EUDAT cloud storage platform (B2DROP),
- Integration into the Parthenos VRE (D4Science) completed
- Please share your experience!
 - beta-switchboard.clarin.eu
 - switchboard@clarin.eu



Short round-up 2019: Back-end

- The average uptime for the period September 2018 to August 2019 for 9 central services is 99.84% (99.91% last year)
- Deployment of a central log server
 - Fluentd, Kibana, Elastic Search
- Improvements to the geographically redundant setup.
- Continuous code-based security scans.
- Upgraded Monitoring setup (Icinga2)



Search...



Search ...

Dashboard

Problems **39**

Overview

Tactical Overview

Hosts

Services

Hostgroups

Servicegroups

Contactgroups

Contacts

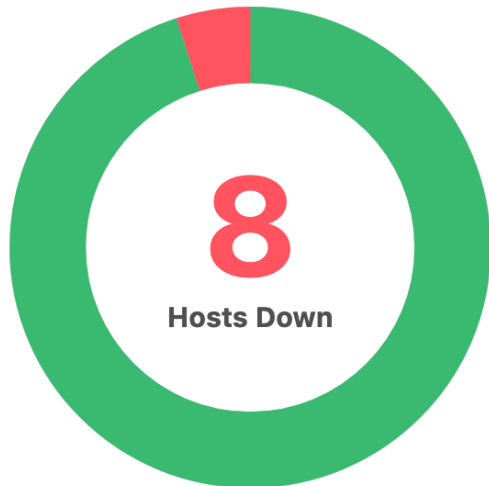
Comments

Downtimes

History

Maps

Host Summary



153 Up

8 Down

Service Summary



229 Ok

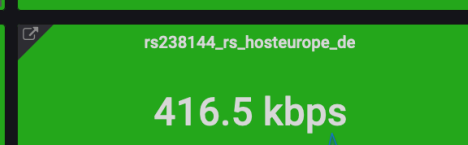
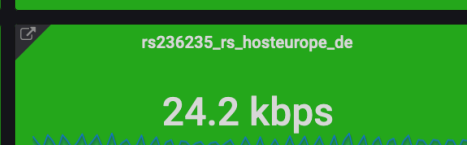
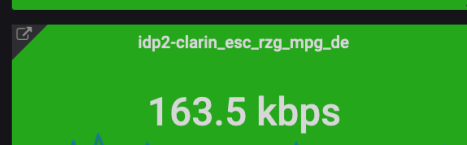
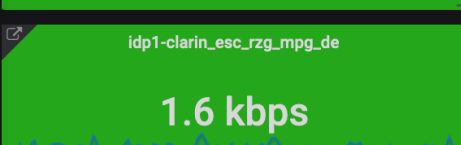
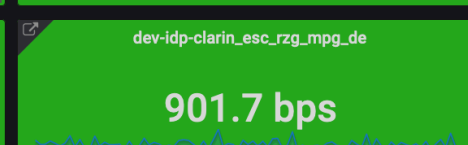
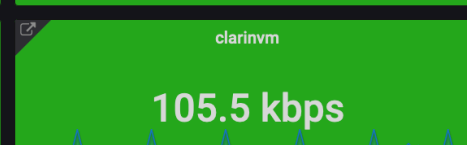
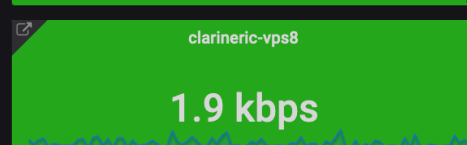
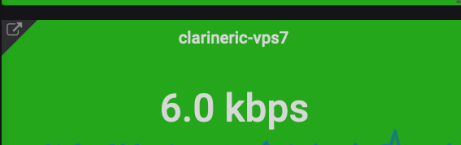
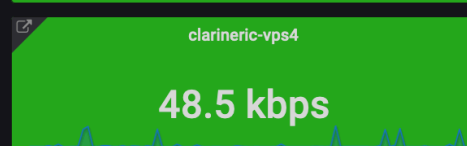
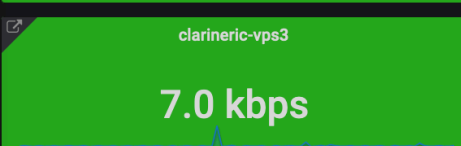
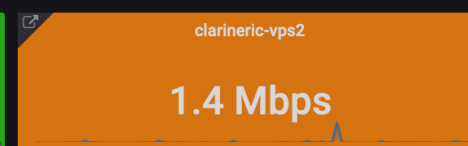
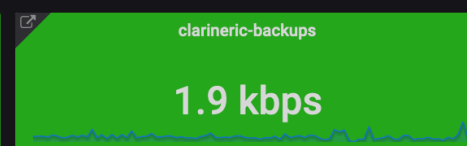
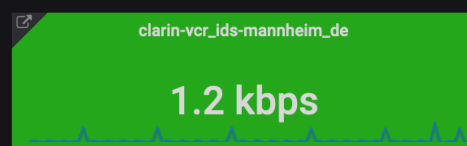
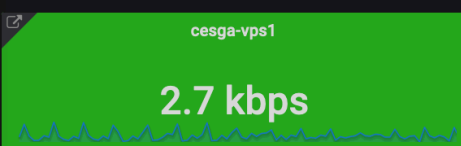
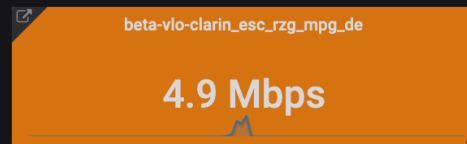
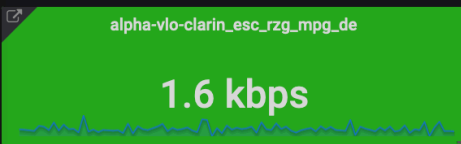
1 Warning (Handled)

25 Warning

5 Critical (Handled)

31 Critical

Network Utilization

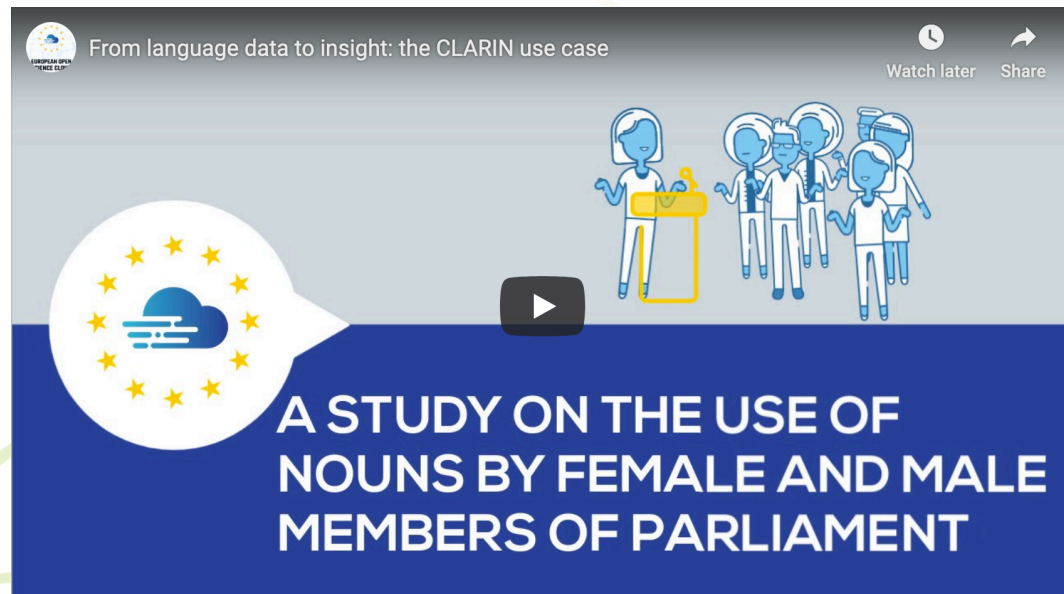


Bridging CLARIN and the rest of the universe



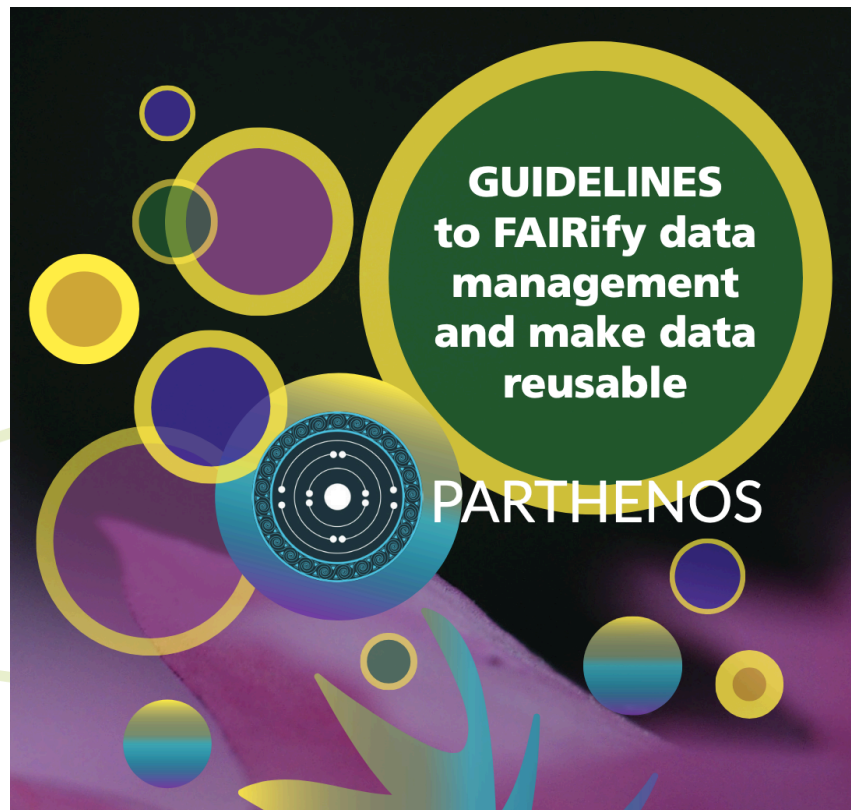
European Open Science Cloud (EOSC)

- See the demonstration by Maciej Ogródniczuk
- www.clarin.eu/eosc



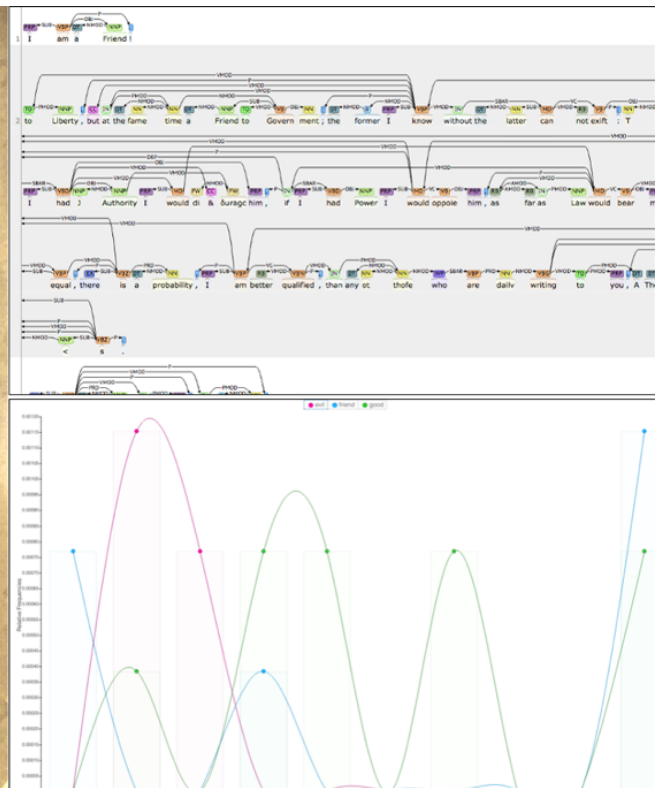
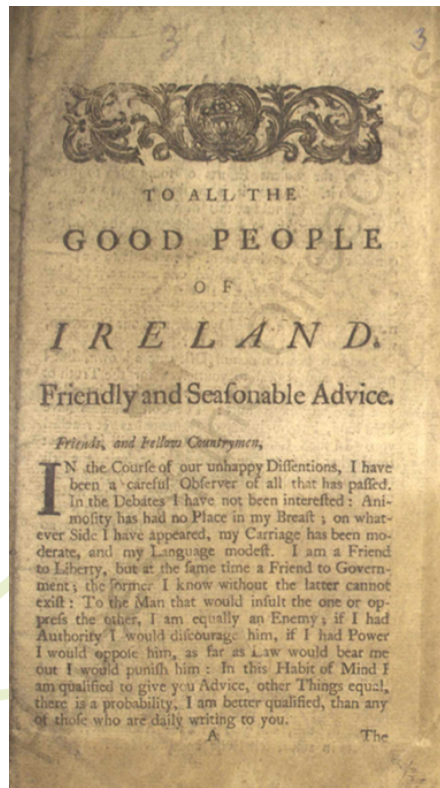
FAIR

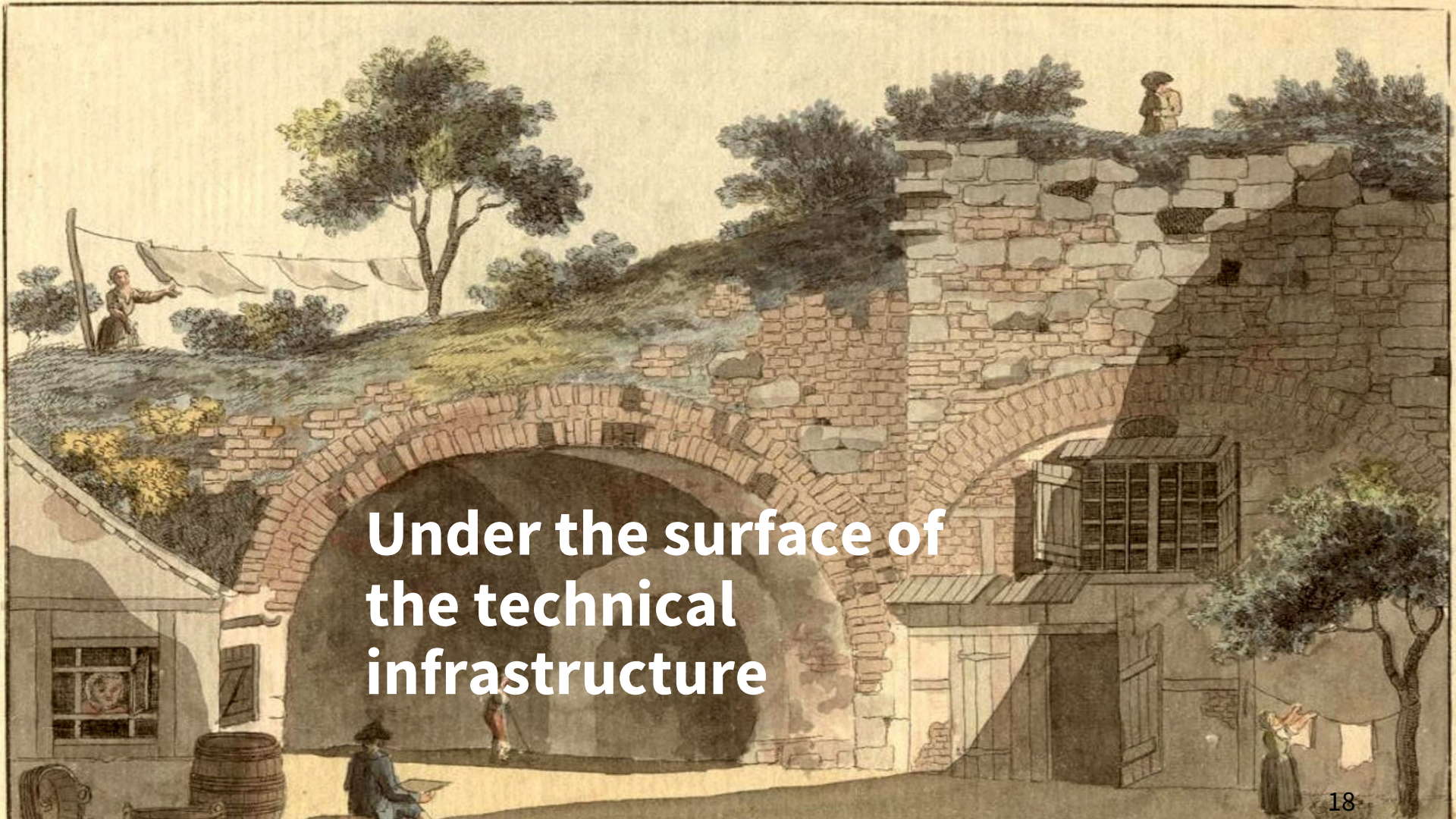
- FAIR is in CLARIN's DNA
- www.clarin.eu/fair



Europeana

- CLARIN and Europeana make discovery and processing quick and easy for 135,000 cultural heritage objects
- clarin.eu/europeana

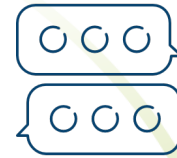




**Under the surface of
the technical
infrastructure**

Resource families links

- A first round of 1040 link checks for the 8 resource families in June
- In June, around 20 broken links (2%)
- About 10 different people were contacted to follow-up
- These conversations gave some good insight in the dynamics behind a simple link



Broken links – technical causes

- 503 Service Unavailable (temporary maintenance)
 - Still needs a check later on
- Real technical issues were found
 - “our handle server is running but cannot open connections to the database”
 - “firewall misconfiguration of our handle server”
 - unreachable DOI: “machine was running out of space; fixed now”

Broken links – other causes

- URL changes & DNS failures
 - “corpus project page got a new URL, that one should be used”
 - “data now available at new subdomain”
 - “... told me that the server has been discontinued, data still available at our repository”
- Observations:
 - Often a URL change reflects an organisational change – it cannot be avoided
 - Interaction with resource providers lead also (unexpectedly) to improvements of the description (fixes, additions, ...)

Link checking: some first conclusions

- Automated link checking is of utmost importance
- Following-up requires time, but this is a good investment since improvements often come with a broader impact:
 - Fixes of otherwise “hidden” technical issues
 - Better metadata, reflecting e.g. a new structure of a dataset
- This is a theme that is also relevant on a broader scale (think EOSC)

Thank you for your attention!

... and credits for the following CLARIN superheroes:

- The assessment committee:
 - Lene, Cyprian, Daan, Jozef, Riccardo, Tomas
- The CLARIN ERIC central developer team:
 - André, Hendrik, Menzo, Michał, Twan, Willem
- Developers and contributors from the national consortia:
 - Alexander, Can, Claus, Dirk, Emanuel, Leif-Jöran, Matej, Nathanael, Oliver, Tariq, Thomas, Wolfgang
- The taskforces:
 - AAI, CMDI & Curation, FCS, PIDs
- And all the others who are contributing to the construction & operation of the infrastructure!

