

Knowledge Sharing & Uptake by researchers

Franciska de Jong

f.m.g.dejong@uu.nl

CLARIN for Newcomers

23 January 2020

Utrecht, Netherlands



Strategic planning 2018-2020

Four priority areas

- Uptake by researchers
- Technical infrastructure
- Knowledge Sharing Infrastructure
- Sustainability

Webpages

- <https://www.clarin.eu/content/vision-and-strategy>
- Where we stand
- Where we want to go

Building the Knowledge Sharing Infrastructure at the national level

Franciska de Jong

f.m.g.dejong@uu.nl



KSI topics covered

- KSI online documentation
- KSI instruments
- Recent KSI workshops
- Plans for 2020
- Local KSI tasks

Where to find the KSI documentation?

- Introductory text to Knowledge Sharing Infrastructure
 - <https://www.clarin.eu/node/4211>
- Overview of K(knowledge)-Centers
 - <https://www.clarin.eu/content/knowledge-centres>
- KSI-committee
 - <https://www.clarin.eu/governance/knowledge-sharing-infrastructure-committee>
- Annual Conference
 - <https://www.clarin.eu/content/clarin-annual-conference>
- Call for workshop funding
 - <https://www.clarin.eu/content/call-clarin-workshop-proposals-2-types-2020>
- Call for mobility grants
 - <https://www.clarin.eu/content/clarin-mobility-grants>
- Newcomers pages
 - <https://www.clarin.eu/content/information-potential-new-members>

Video channels: YouTube and VideoLectures



- <http://videolectures.net/clarin/>
- Archive of recorded lectures
 - 16 events
 - 159 videos



[CLARIN@YouTube](#)

Other video materials,
e.g. interviews, webinars

CLARIN Seminar 2018 - Szeged

CLARIN Seminar on Speech and Language Technology Tools, Szeged 2018

released under terms of: Creative Commons Attribution (CC-BY)

The seminar on Speech and Language Technology Tools is organized by HunCLARIN, Juhász Gyula Faculty of Pedagogy at the University of Szeged and the Hungarian Association of Applied Linguists and Language.

The aim of the seminar is to provide researchers, teachers and students working in the humanities and social sciences a broad view of the corpora and state-of-the-art software in speech and language processing developed for Hungarian, mostly as part of the activities of HunCLARIN.

This seminar is supported by CLARIN ERIC.

The seminar took place at the **Juhász Gyula Faculty of Education, University of Szeged, Hungary on Friday, 19 October, 2018.**

Categories

- Top » Data Science
- Top » Humanities » Linguistics » Lexicography
- Top » Humanities » Languages
- Top » Social Sciences » Education

[syn] 2 views, 41:03 Introduction to the world of corpora and linguistic databases Veronika Vincze	[syn] 1 view, 40:33 Corpus query Bálint Sass	[syn] 27:16 Historical corpora for Hungarian Eszter Simon	[syn] 25:57 Introduction to e-magyar Iván Mittelholz	[syn] 19:11 The possibilities of MAXQDA content analysis software Valéria Juhász
[syn] 32:56 Child language corpora and resources Anna Babarczy	[syn] 1 view, 29:19 The big data challenge in Humanities: Introduction of some research tools in digital humanities Róbert Péter			

Annual Conference

- Organized since 2012
- Co-production by Local team and CLARIN Office
- NCF acts as PC (Programme Committee)
- Funded participation:
 - Members: 5 +1 PhD funded participants
 - Observers: 3 funded participants
 - Extra funded places for accepted papers (1 author per paper)

Consider to host the event!

CLARIN-DARIAH Digital Humanities Course Registry

- The Digital Humanities Course Registry has been further developed; improved interface developed.
- Increasingly recognised as a platform with added value for various sub-communities of the global DH field; further internationalisation is on its way.
- The ADHO SIG for Pedagogy has endorsed the DH Course Registry and it is listed in the ESFRI Roadmap 2018.
- Agreement has been reached with DARIAH on the joint maintenance of the DH Course Registry for the coming three years.
- Number of entries has increased.
- <https://dhcr.clarin-dariah.eu/>



DIGITAL HUMANITIES
COURSE REGISTRY

Recent CE workshops

- Data protection (December 2018)
 - focus on the impact of the GDPR regulation on the field of language resources and tools
 - in collaboration with the CLIC committee and invited legal experts
 - follow-up workshop currently investigated
- The use of CLARIN in academic curricula (November 2019)
 - focus on the experience and best practices with the integration of CLARIN content into university programmes
 - contribution to a pre-workshop survey by all participants
- Twin Talks: Understanding Collaboration in DH
 - workshop organised in the context of the 4th Digital Humanities in the Nordic Countries Conference (DHN 2019) in Copenhagen
 - focus the dynamics of collaboration dynamics in cross-disciplinary projects
 - follow-up events planned

Funding for KSI events

Call for mobility grants

- Short visits
- Aimed at exchange of expertise between nodes
- <https://www.clarin.eu/content/clarin-mobility-grants>



Call for workshop funding

- Cross-country collaboration
- Possibility for follow-up development work
- <https://www.clarin.eu/content/call-clarin-workshop-proposals-2-types-2020>

KSI - Activities planned for 2020 (highlights)

CLARIN Annual Conference

- PhD poster session
- PhD counseling

Other KSI instruments

- Organisation of a series of meetings aimed at re-articulation the objectives of the CLARIN KSI; the topics will include the revisit of the KSI instruments; invites may include K-centre managers on the experiences and issues with running a K-centre
- Further promotion of the mobility grants (10 grants expected in 2020)
- Further collaboration with DARIAH-EU on the course registry
- Further development of the CLARIN website and populating it with content; a.o. increased visibility of best practices for GDPR-related issues and Data Management Planning
- Development of a strategy aimed at increasing the volume and measurability of publication output as illustration of the support for scientific excellence by CLARIN; this may include the establishment of an award for a 'best paper' and other instruments that could stimulate including references to the use of the CLARIN RI.

Local Knowledge Sharing tasks

- Set up one or more K-centres (Tour de K-CLARIN is coming!)
- Plan workshops, training, outreach (calls for grant applications!)
- Plan trips abroad for the exchange expertise



- Appoint a coordinator for the DH course registry
- Plan workshops between countries (workshop call)
- Develop showcases, training materials, webinars, etc.
- Promote integration of CLARIN use in university curricula
- Lessons learnt: https://office.clarin.eu/v/CE-2016-0879-CLARINPLUS-D5_3-3-Sharing-knowledge-infrastructure.pdf

UI @ CLARIN ERIC



Darja Fišer
Director of User Involvement CLARIN ERIC
darja.fiser@ff.uni-lj.si



UI topics covered

- UI online documentation
- National UI Representatives
- UI Instruments
- Funding for UI events
- UI and Uptake in 2019-2020

Where to find the UI documentation?

- List of national UI representatives
 - <https://www.clarin.eu/governance/user-involvement-committee>
- UI dashboard with the archive, calls, documents
 - <https://www.clarin.eu/group-page/user-involvement-group-dashboard>
- Mailing list
 - userinvolvement@lists.clarin.eu
- CLARIN for researchers
 - <https://www.clarin.eu/content/clarin-for-researchers>
- Tour de CLARIN
 - <https://www.clarin.eu/tour-de-clarin>
- CLARIN Resource Families
 - <https://www.clarin.eu/resource-families>
- Funding for UI events
 - <https://www.clarin.eu/news/call-clarin-user-involvement-event-proposals-2019>
- Videlectures
 - <http://videolectures.net/clarin/>

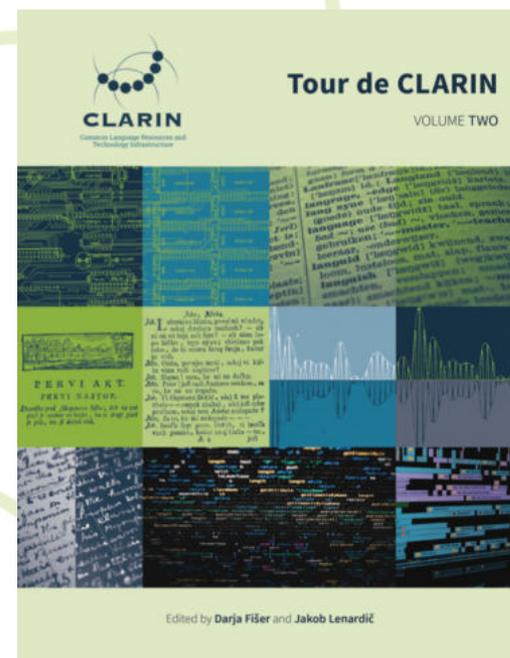
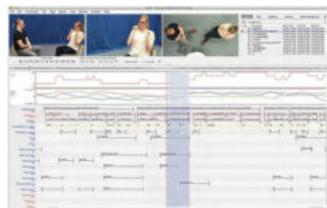
User Involvement Committee

- Tasks
 - facilitating the UI initiatives conducted by CLARIN ERIC
 - promoting the UI financing instruments within their consortium and networks of researchers (mobility grants, support grants for UI events, etc.)
 - promoting UI participation at the relevant national and international UI events
 - facilitating the UI surveys conducted by CLARIN ERIC
- Coordination
 - Chair: Darja Fišer
 - Vice-chair: Maria Gavriilidou (clarin:el)



- Aims
 - increase the visibility of the national consortia and K-centres
 - reveal the richness of the CLARIN landscape
 - display the full range of activities throughout the network

- Highlights
 - consortium
 - resource
 - tool
 - UI event
 - researcher



Aims

Evaluation and improvement of the state of the infrastructure
user-friendly overview of available resources in the CLARIN
infrastructure for SSH & HLT

- Corpora
 - historical corpora
 - Literary corpora
 - L2 learner corpora
 - spoken corpora
 - manually annotated
 - CMC corpora
 - newspaper corpora
 - parallel corpora
 - parliamentary corpora
- Lexical Resources
 - Lexica
 - Dictionaries
 - Conceptual Resources
 - Glossaries
 - Worldlists
- Tools
 - For normalization
 - For name entity recognition



Francesca Frontini



Maciej Maryl



Toine Pieters

- **Scientific events**

- Poster & demo by FF at conference “**Intelligence mécanique**”, 11–12 June 2019, Montpellier
- Lecture by MM at workshop **NEP4DISSENT COST Action**, 31 May 2019, Tirana
- Seminar by MM at **Summer Training School in Digital History**, 15-23 July 2019, Budapest
- Talk by TP at workshop **Time Machine for digital heritage and humanities**, 29 August 2019, Avila
- Talk by TP at conference **HERA Public Spaces: Culture and Integration in Europe**, 11–12 September 2019, Gdańsk
- Talk by TP at conference **International Pharmaceutical Federation**, 24–28 September 2019, Abu Dhabi

- **Projects, networks, institutions, research infrastructures**

- Contribution by FF at **Atelier Condorcet** on geo-parsing historical texts, 14 May 2019, Paris
- Contribution by FF at meeting of **corpus linguistics labs** on creation of South France network, 22 May 2019, Montpellier
- Talk by MM at a joint workshop on **literary bibliographies** by the Institute of Czech Literature and the Institute of Literary Research, 17 October 2019, Prague

- **Media dissemination**

- Interview for **ILIS Digital Libraries & Information Sciences blog** with FF, 11 July 2019

Funding for UI events

- **CLARIN Master Classes**

- a session at a major event in Digital Humanities and/or Social Sciences
- CLARIN experts recommend the approaches /tools/resources/training models that have been adopted or developed in CLARIN

- **CLARIN Courses at Summer Schools**

- a course/tutorial/workshop on a CLARIN-related topic at a prominent DH and/or SS summer school

- **CLARIN Training Events**

- a workshop/seminar for researchers on new/key language tools and resources
- a seminar for teachers addressing topics not sufficiently covered in the curricula
- a seminar addressing an under-serviced community (e.g. diaspora, minorities, special needs)
- a refreshment course for language professionals (e.g. translators, journalists)

UI events in 2019

UI events organized by CLARIN ERIC

- 5 workshops

UI events (co)funded by CLARIN ERIC

- 2 Summer schools
- 1 Hackaton

Invited contributions at 3rd party events

- 1 Workshop

Contributions by CLARIN Ambassadors

- 7 Regular talks
- 3 Promotions within projects, institutions, Ris
- 1 Media dissemination



National UI Activities (1/2)

2019

229 UI activities recorded

- 37 more than in 2017
- max: Germany 53
- > 10: Estonia, Austria, Sweden, South Africa, Finland, Poland Denmark
- ≥ 1 : all members & observers except UK, Croatia, Norway & Portugal
- duration: 1-day events (68%)
- attendance: up to 350 participants, over 7,400 in total

2018

192 UI activities recorded

- 32 more than in 2017
- max: Germany 37
- > 10: the Netherlands, Hungary, Poland, Estonia, Finland
- ≥ 1 : all members & observers except UK & Croatia
- duration: 1-day events (61%)
- attendance: up to 1000 participants, over 10,000 in total

National UI Activities (2/2)

- Types
 - most frequent: teaching activities (34%) & organizations of UI events (33%)
 - most consortia: teaching activities (16 out of 20)
- Organization
 - by consortia 42% > more than last year
 - external initiatives 46% > more than last year
- Scope
 - 79% national character > the same as last year
 - 47 events involved participants from multiple countries
- Trends in Disciplines
 - **Linguistics** (107) > in 2019 twice as many as in 2018
 - Humanities (78) > in 2019 a quarter less
 - **Social Sciences** (35) > in 2019 10 more
 - **Literary studies** (28) > in 2019 among top-ranking disciplines for the first time
 - Computer Science (25) > in 2019 6 less than last year
 - History (24) > in 2019 7 more than last year
 - **Political science and sociology** (12) > in 2019 among top-ranking disciplines for the first time
- Language
 - English (44), South African (26), German (23), Polish (19), Swedish (15)

CE UI activities in 2019

- Establishing new connections
 - CLARIN Ambassadors appointed ([link](#))
 - through COST Actions (e.g. literary studies)
 - through the SSHOC project (social sciences)
- Existing user communities
 - Support on how to use CLARIN
 - Support on how to contribute to CLARIN
- Highlights:
 - Tour de CLARIN for K-centres
 - CLARIN Resource Families - additional focus: tools
 - Text analysis tools beyond concordancers
 - Workshop on the use of CLARIN in university curricula



Uptake - Activities planned for 2020 (highlights)

Enhancing Uptake-Oriented Outreach

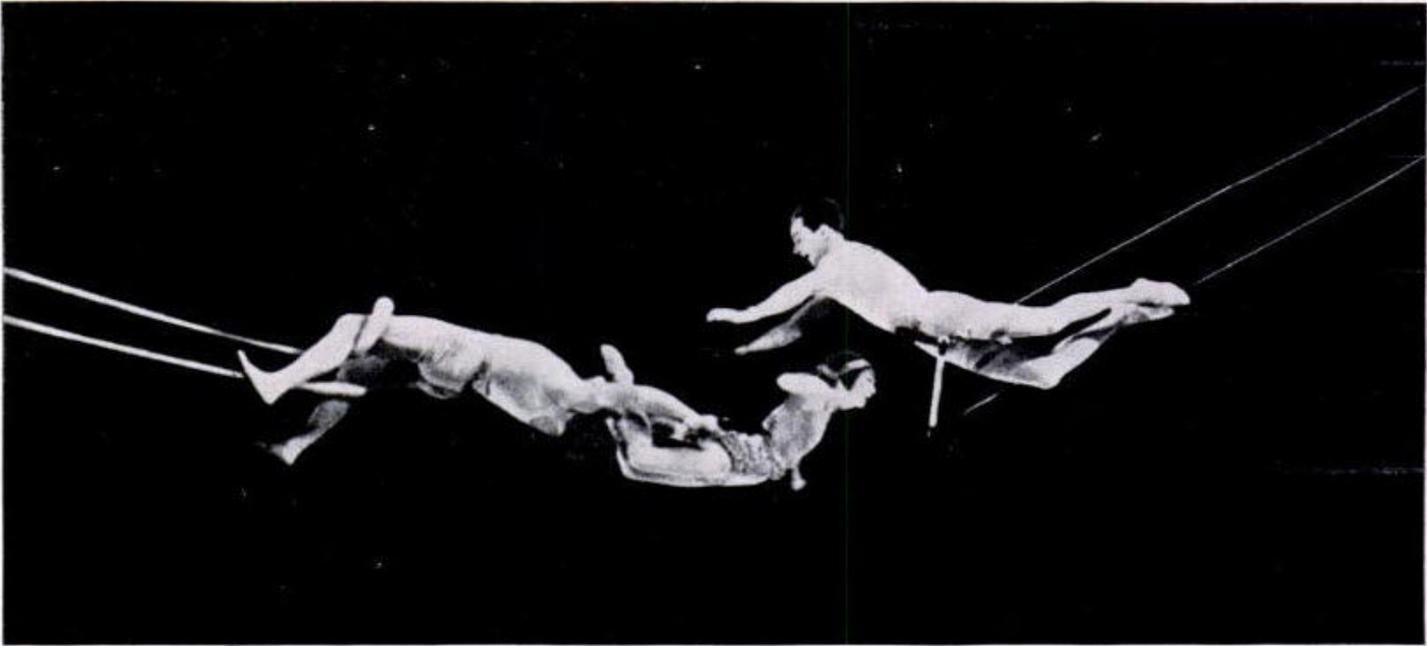
- Promote integration of CLARIN use in university curricula and seminars offered by research libraries (in collaboration with LIBER).
- Appoint 2 more CLARIN Ambassadors
- Continue the outreach activities and support for events
- Production of dissemination materials:
Tour de CLARIN for K-centres, instructional videos

Enhancing the Findability and Usability of Services

- For enhancing the interoperability and visibility of CLARIN tools and resources, the focus in 2020 will not only include parliamentary records but political texts at large, thus including also political speeches, party manifestos etc. Mature and manageable sets of resources and tools are available -> a testing ground to develop a systematic, comprehensive and strategic model for future Families of Resources activities.
- Related scientific events:
 - workshop on political data (e.g. LREC)
 - SSHOC events

Uptake by researchers

Technical Infrastructure



"The Passing Leap" was the most spectacular display of the Codona's perfect co-ordination. Swinging 60 ft. above the arena,

Vera would catapult from Lala's hands to Alfredo's trapeze just as her husband was soaring over her into his brother's hands.

Knowledge Sharing Infrastructure

Contributions by
Darja Fišer
Elisa Gorgaini
Jakob Lenardič

see you @
www.clarin.eu
or
f.m.g.dejong@uu.nl



Background information: H2020 project SSHOC in relation to EOSC

Franciska de Jong

f.m.g.dejong@uu.nl

CLARIN for Newcomers

23 January 2020

Utrecht, Netherlands





Horizon 2020
European Union Funding
for Research & Innovation

Type of action & funding:
Research and Innovation action
(INFRAEOSC-04-2018)

Partners: 47
(20 beneficiaries + 27 LTPs)

SSH ESFRI Landmarks and Projects
& international SSH data infrastructures

Project budget:
€ 14,455,594.08

Duration: 40 months
(January 2019 – 30 April 2022)

Project website:
www.SSHOpenCloud.eu



Objectives:

- creating the social sciences and humanities (**SSH**) part of European Open Science Cloud (**EOSC**)
- maximising **re-use** through **Open Science** and **FAIR** principles (standards, common catalogue, access control, semantic techniques, training)
- interconnecting existing and new infrastructures (clustered cloud infrastructure)
- establishing appropriate **governance model** for SSH-EOSC

SSHOC Partners



*E-RIHS is not a legal partner in the SSHOC project but we connect to the E-RIHS community through the Institutum Archaeologicum Germanicum.

Expected impact



The Social Sciences and Humanities are seamlessly integrated in the European Open Science Cloud



Availability of an EU-wide, easy-to-use SSH Open Marketplace, where tools and data are openly accessible



EU-wide availability of high quality "cloud ready" SSH tools and high quality SSH data



EU-wide availability of trusted and secure access mechanisms for SSH data, conforming to EU legal requirements



State of the art Research Infrastructure in several pilot domains advanced through dedicated SSH data pilots cluster projects



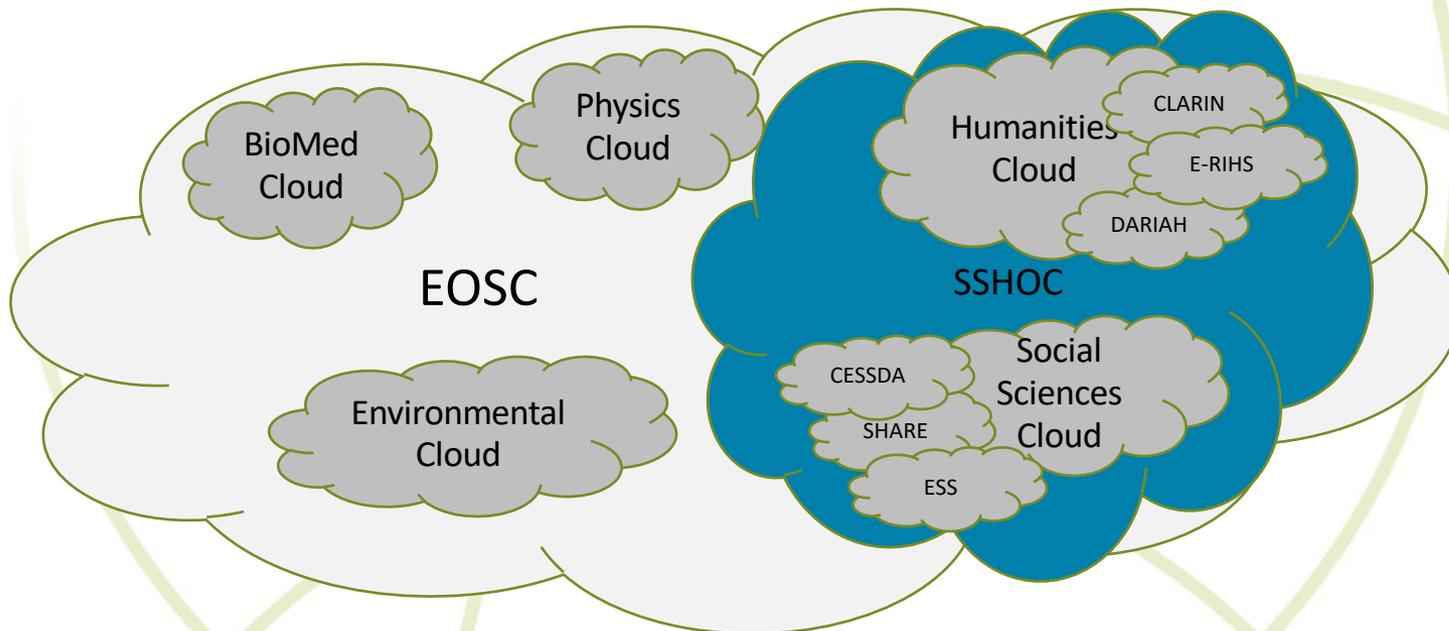
Maximising reuse through Open Science and FAIR principles (standards, common catalogue, access control, semantic techniques, training)

SSHOC Stakeholders and offering



EOSC as Cloud of Clouds

- 'Cloud' is a label for distributed infrastructure, with shareable resources, optimized access for researchers and cost efficiency.
- Many options for organizing the current research infrastructure landscape where much has been created already: services, infrastructures, collaboration



Clusters in the EOSC landscape

- All of different size, structure and scope
- Different levels of EOSC and e-Infra involvement: for SSHOC partners:
 - CLARIN ERIC and some DARIAH partners are involved in EOSC-hub
 - other partners were involved in the EOSC pilots, and
 - SSHOC/CESSDA participation in EOSC-hub strategy board
 - SSH represented in EOSC executive board
- Still most pronounced are differences with non-thematic e-Infrastructure organizations
 - that have been active long time and
 - have excellent ties to the national research infra facilities while
 - clusters have their roots in research communities and disciplinary practices.
- Challenge: how to manage the diversity and the need for alignment of strategies
 - at the disciplinary level
 - at the national level
 - at the European level
 - at the level of e-Infrastructure organizations

