

LREC 2020 Workshop
Language Resources and Evaluation Conference
11–16 May 2020

**Workshop about Language Resources for the SSH Cloud
(LR4SSHOC)**

PROCEEDINGS

Editors: Daan Broeder, Maria Eskevich, Monica Monachini

Proceedings of the LREC 2020 Workshop about Language Resources for the SSH Cloud (LR4SSHOC)

Edited by: Daan Broeder, Maria Eskevich, and Monica Monachini

Acknowledgements: This workshop is supported through funding from European Union's Horizon 2020 research and innovation programme (under grant agreement No 823782 for the SSHOC project) and the CLARIN infrastructure.

Workshop web site: <https://www.clarin.eu/LR4SSHOC>

ISBN: 979-10-95546-43-6

EAN: 9791095546436

For more information:

European Language Resources Association (ELRA)

9 rue des Cordelières

75013, Paris

France

<http://www.elra.info>

Email: lrec@elda.org

© European Language Resources Association (ELRA)

These workshop proceedings are licensed under a Creative Commons
Attribution-NonCommercial¹4.0 International License

Introduction

In recent decades the development of language resources (LR) and language technologies (LT) and their management have reached the level of maturity that allows their usage to be expanded beyond the borders of traditional linguistic disciplines and implementations. Although admittedly many challenges remain such as easier localisation, configuration and deployment of LT services for non LT experts.

At the same time, the broad domain of social sciences and humanities (SSH) research despite having a diverse set of domain specific methods and practices, can benefit from LT infrastructure approaches and research results, to extract information from natural language content. These LT methods and practices can be adopted by the already existing SSH research infrastructures that support their domain specific work. And In the larger context, on the European landscape level, the European Open Science Cloud (EOSC), currently under development, will be created to also facilitate the cross domain use of data and technologies. To support this strategy, a set of EU thematic cluster projects collects common and specific requests from kindred fields and domains to ensure a smooth transition and collaboration in order to reach the goal of bringing data, tools and services into the common cloud. The Social Sciences and Humanities Open Cloud (SSHOC) project is the SSH thematic cluster project aiming to create the SSH part of the EOSC.

This workshop was envisaged to focus on the goals and aims of realising the SSHOC part of the EOSC, where SSH data, language processing tools, and services are made available, adjusted and accessible for users across SSH domain. It provides a forum to discuss common requirements, challenges and opportunities for developing, enhancing, integrating tools and services for managing and processing SSH research data. Such SSH scenarios based implementations of currently existing language tools and services demonstrate their multidisciplinary usability and stimulate further multidisciplinary collaboration across the various subfields of SSH and beyond, which will increase the potential for societal impact.

The workshop introduces the SSHOC project and its ambitions while also including its embedding in the EOSC for dialogue with the LREC community. On the one hand, such discussion between SSH data-practitioners, infrastructure and LT experts will strengthen and support the SSH community connection with the LREC landscape and initiatives. On the other hand, the much increased interest in and availability of cloud type infrastructure approaches represent an opportunity for the language technologies to support the field of social sciences and humanities on a large scale following the F.A.I.R. principles, thus supporting its replicability and reproducibility.

For this workshop we have ask for contributions aimed at aligning and integrating services and infrastructure from the Social Sciences, Humanities and Cultural Heritage with one another and with the now emerging European Open Science Cloud, that is being built for sharing and optimising research data and services in a sustainable way. It is especially interesting to see the examples of such infrastructure components that (can) play a role in this, but also at use-cases of cross-domain use of SSH services.

This workshop aimed at gathering together academics, industrial researchers, digital language resources and technology providers, software developers, but also, and in particular, SSH representatives in order to identify the current capacity and the difficulties in creating and sustaining an infrastructure for SSH domain.

The accepted papers address the following topics:

- Research infrastructure components
- Use cases of text and data mining for SSH-driven tasks

- FAIRness of sensitive language data
- Challenges for language technologies in EOSC
- SSH future and governance in the EOSC world
- EOSC and business models for language data, tools and services

The workshop programme is composed of 4 invited papers and 5 peer-reviewed papers. We would like to thank the reviewers for their careful and constructive reviews which have contributed to the quality of the event.

Due to the COVID-19 pandemic, the main conference is postponed, and consequently, the LR4SSHOC workshop is postponed till later possibility to gather the audience of authors and invited speakers.

D. Broeder, M. Eskevich, M. Monachini, M. Kleemola, N. Larrousse

May 2020

Organizers:

Daan Broeder, CLARIN ERIC
Maria Eskevich, CLARIN ERIC
Monica Monachini, Institute of Computational Linguistics - CNR
Mari Kleemola, TAU, CESSDA-FI
Nicolas Larrousse, Huma-num, DARIAH-FR, CLARIN-FR

Program Committee:

Daan Broeder, CLARIN ERIC, The Netherlands
Maria Eskevich, CLARIN ERIC, The Netherlands
Vasso Kalaitzi, LIBER, The Netherlands
Mari Kleemola, TAU, CESSDA-FI
Nicolas Larrousse, Huma-num, DARIAH-FR, CLARIN-FR, France
Monica Monachini, Institute of Computational Linguistics - CNR
Jan Odijk, Utrecht University, The Netherlands

Invited Contributors:

Daan Broeder, Maria Eskevich, CLARIN ERIC
Donatella Castelli, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Consiglio Nazionale delle Ricerche
Christopher Cieri, University of Pennsylvania, Linguistic Data Consortium
Henk van den Heuvel, CLS/CLST, Radboud University
Monica Monachini, Institute of Computational Linguistics - CNR

Workshop Program

Peer-reviewed papers

- Store Scientific Workflows Data in SSHOC Repository*
Cesare Concordia, Carlo Meghini and Filippo Benedetti 1
- Social Sciences and Humanities Pathway Towards the European Open Science Cloud*
Francesca Di Donato, Monica Monachini, Maria Eskevich, Stefanie Pohle, Yoann Moranville and Suzanne Dumouchel 5
- From the attic to the cloud: mobilization of endangered language resources with linked data*
Sebastian Nordhoff 10
- Verbal Aggression as an Indicator of Xenophobic Attitudes in Greek Twitter during and after the Financial Crisis*
Maria Pontiki, Maria Gavriilidou, Dimitris Gkoumas and Stelios Piperidis 19
- Mining Wages in Nineteenth-Century Job Advertisements. The Application of Language Resources and Language Technology to study Economic and Social Inequality*
Ruben Ros, Marieke van Erp, Auke Rijpma and Richard Zijdeman 27

Invited position papers

- LR4SSHOC: The Future of Language Resources in the Context of the Social Sciences and Humanities Open Cloud*
Daan Broeder, Maria Eskevich and Monica Monachini 33
- EOSC as a game-changer in the Social Sciences and Humanities research activities*
Donatella Castelli 37
- Stretching Disciplinary Boundaries in Language Resource Development and Use: a Linguistic Data Consortium Position Paper* 39
Christopher Cieri
- Crossing the SSH Bridge with Interview Data* 42
Henk van den Heuvel