

Open Call Collection OC-2019-1

Proposal Reference OC-2019-1-23680

Title: Maximising the use of L2 data across disciplines: from collection to applications

Acronym: L2data4all

Summary

Targeting the empirical core of second language (L2) learning as a foundational component of the European vision, the Action aims to support networking across well-established yet largely separate research communities – Second Language Acquisition, Learner Corpus Research and Computer-Assisted Language Learning – with the goal of optimizing the use of L2 data on the product and process of L2 learning for research, practice and society. L2 data includes all types of written, spoken and multimodal language production and interaction data by learners of the many languages spoken, taught and learned in Europe. Across disciplines, researchers collect and analyse such data to answer their own research questions, each capturing just a part of the foreign language learning puzzle. Cross-fertilization is severely hindered by a variety of other factors, including lack of shared conceptualisations, standardised representations and procedures for data processing, and limited access to L2 data.

Through its multidisciplinary network, the Action will seek to foster mutual reinforcement and provide collective interdisciplinary answers to repeated calls for (1) FAIR data and the adoption of principles of Open Science, (2) standardisation and inter-operability of data description, annotation procedures and tools, (3) the adoption of ethical principles and GDPR regulation, and (4) replication studies, triangulation of methods, and cumulative science.

The Action also aims to build a sustainable connection between researchers and practitioners by promoting synergies through collaboration with foreign language teachers, language testing agencies, and teaching material developers, actively engaging with their practical needs, which will benefit European language learners of all ages and backgrounds.

Key Expertise needed for evaluation

Languages and literature

Second language teaching and learning

Languages and literature

Databases, data mining, data curation, computational modelling

Educational sciences

Education: training, pedagogy, didactics

Languages and literature

Linguistics: formal, cognitive, functional and computational linguistics

Keywords

learner language

FAIR principles

ethical use of L2 data

study quality

L2 data for practice

TECHNICAL ANNEX

1 S&T EXCELLENCE

1.1 SOUNDNESS OF THE CHALLENGE

1.1.1 DESCRIPTION OF THE STATE-OF-THE-ART

A range of stakeholders are directly concerned with the collection, analysis and use of **second language** (L2) data, i.e., written, spoken and multimodal language productions by foreign learners of the many languages spoken in Europe (e.g., in the form of academic essays, official language exams, emails, exchanges on social media, interviews). Stakeholders include language teachers, second language acquisition researchers, language testers, companies that produce language teaching/learning materials, and developers of Computer-Assisted Language Learning (CALL) systems and other applications such as spell/grammar checkers and writing assistants. Interaction between the groups has traditionally been very limited due to different types of data being considered, different needs and interests, and the different expertise required to work in different areas. As a result, Second Language Acquisition (SLA), Foreign Language Teaching (FLT), Learner Corpus Research (LCR), Language Testing and CALL have developed as distinct domains of enquiry, with their own journals, conferences and international associations, to the point that it has even become a challenge to read and understand each other's publications. At the same time, unfortunately, this has severely limited (1) the ability to characterize second language learning processes and products with precision and ecological validity, each field being concerned with and capturing just a part of the language learning puzzle; and (2) the impact and integration of research insights on the practice of teaching, learning and assessment.

Cross-fertilization of the different fields is currently seriously hindered by a variety of factors, including:

- (1) Lack of a common terminology, standardised representations and procedures for data processing within and across fields, which restricts study comparability, replication and meta-analyses, and by extension the understanding of the foreign language learning phenomena investigated (e.g., development of vocabulary, grammar, pronunciation, and communicative competence). For example, given the diversity of the transcription guidelines (to transcribe oral data into machine-readable searchable text) and annotation schemes (e.g., to code errors) employed, studies are hardly comparable and it is therefore difficult to evaluate the degree of progress in the field.
- (2) Limited access to L2 data despite official recognition of the potential benefits of data sharing and dissemination across Horizon 2020 initiatives through the European Open Science goals. Studies investigating barriers to data sharing in psychology research have identified researchers' lack of training as a key constraint, alongside insufficient curricular resources related to ethical, moral and legal issues inherent in data sharing (Houtkoope, Cambers, MacLeod, Bishop, Nichols & Wagenmakers 2018); it is likely that L2 researchers face similar challenges. Further barriers to data dissemination arise as the increased availability of personal data underscores the need to ensure researcher and stakeholder trust that the highest levels of ethical procedures have been maintained throughout the research data lifecycle. The cross disciplinary nature of L2 research means that L2 data is generated in an ever increasing number of contexts, and underpins increasingly diverse applications ranging from policy support for bilingual education, development of software and teaching materials, and voice recognition technology, among many others. The ways in which ethical principles in L2 research are realised in practice differ as a function of discipline, research methodology, and institutional and national context (DeCosta 2015). Complicating matters further, complex ethical issues have had little explicit consideration from within key fields engaged in L2 research (Thomas & Pettitt 2016). Finally, the implementation of the General Data Protection Regulation (GDPR) in May 2018 has introduced important changes to regulatory frameworks governing data protection. Shifts in accountability from data protection authorities to data controllers and data processors, combined with a lack of clear procedural guidance, and flexibility in national definitions of key terminology, have generated a risk that researchers and their institutions "err on the side of caution" and choose not to share their research data (Nature 557, 467 (2018) doi: 10.1038/d41586-018-05220-y), despite initiatives put in place by journals, researchers and funders. Although the regulation provides opportunities for "bottom-up" harmonisation via Codes of Conduct, and Data Seals, these have yet to be developed within the L2 research community.
- (3) Yet-to-be-addressed challenges raised by new L2 data types such as big L2 data from online language teaching platforms and social media. The emerging fields of Learning Analytics and

Educational Data Mining, which in principle are designed to focus on this type of big data, have so far focused on the teaching and learning of mathematics, programming, and specific, richly modeled scientific content domains – avoiding engagement with second language learning, which is understandable considering that language learning/teaching is traditionally regarded as an “ill-defined domain” in the related Intelligent Tutoring Systems world. In order to exploit such large-scale learner corpora, we need tools to automatically analyse them. With modern computational power, one can readily identify and retrieve linguistic patterns based on surface forms (e.g., a particular sequence of words). Much of learner corpus research, therefore, has focused on lexical dimensions of language and the higher-level aspects that are signalled by lexical items (e.g., discourse markers indicating cohesion of text). SLA researchers, however, are often interested in more abstract features such as verbal morphemes (e.g., past tense *-ed*) or relative clauses, or in their aggregated indices in the form of complexity measures such as the number of dependent clauses per sentence. In order to extract those features from corpora automatically, annotation using Natural Language Processing (NLP) tools at various levels (e.g., word, syntax) is essential. There are now relatively easy-to-use NLP tools (e.g., taggers, parsers) freely available, as well as graphical user-interfaces that support calculating various complexity metrics for English. However, like other types of ‘non-standard’ languages such as spoken language or Internet language, learner languages have their own characteristics which may adversely influence the accuracy of those tools, and appropriately assessing the validity of such tools on learner language, let alone adjusting them, requires expertise in computational linguistics. To complicate matters further, tools for L2 data analysis are still sorely missing for the many other languages learnt as an L2 in Europe. It is, therefore, necessary to collaborate with NLP researchers in order to take the full potential of automatic language processing tools for a wide variety of languages while maintaining the validity of the annotation (Meurers & Dickinson, 2017).

- (4) Various degrees of engagement with the methodological turn and principles of Open Science. Recently, fundamental issues in scientific reasoning have been vividly debated in a number of scholarly disciplines with strong quantitative traditions (see, e.g., Open Science Collaboration & Kappes, 2014). Relevant publications – referring to a paradigm labelled the “replication crisis” – have explored major shortcomings of current research and publication practice, which may inhibit the ultimate goal of scholarly study: progress in knowledge construction. These publications explicitly aim to make researchers aware of problems relating, for instance, to lack of scientific rigour in study design, dangers of convenience samples, inadequate use and interpretation of statistical methods, selective reporting of results, lack of methodological transparency, unavailability of original data, etc. Even though work in disciplines dealing with L2 data is also increasingly of a quantitative nature (Khany & Tazik, 2019; Paquot & Plonsky, 2017) and inevitably affected by related shortcomings (Al-Hoorie & Vitta, 2018), a comparable discourse is only starting to emerge in the respective communities at large (cf. the IRIS database project, <https://iris-database.org>). Some aspects of the replication crisis are being tackled within the “methodological turn” in quantitative studies, but the transparency of methodological choices and the application of principles of Open Science (OS) are still at a rather low level. In addition, there is a lack of methodological communication between sub-disciplines that share the interest in L2 data, and comparable methodological attention appears to be much less present in qualitative and mixed methods work.

1.1.2 DESCRIPTION OF THE CHALLENGE (MAIN AIM)

The main aim of this Action (**Challenge 1**) is to bring together, bridge gaps, foster networking, and establish best practices across well-established but separate communities that share foreign/second language learner data as their first source of evidence to enhance L2 data impact on research, practice and policy. In the long term, such a multi-source and multi-perspective view of the process and product of second language learning will not just provide a broader empirical basis for more comprehensive and robust theories, but will arguably also make the research outcomes more relevant and applicable for stakeholders engaged with teaching and learning practice. In Language Testing with the psychometric properties and statistical methods refined to a high degree, an increasing focus on ecological validity and a keen awareness of the “Validity Models in Transition” (the title of the keynote symposium at the 2019 conference by the European Association for Language Testing and Assessment) is starting to lead the language testing field to look for more cross-disciplinary connections that promise to provide authentic data on the process and product of language learning. In learner corpus research, the collection and multi-level annotation of corpus data is being refined to a degree that makes this type of evidence increasingly relevant for addressing long-standing research questions in SLA where researchers are still mostly working with fairly small-scale empirical studies that often lack generalisability. At the same time, more sustained collaboration of corpus researchers with teachers and learners is needed for collecting longitudinal data (i.e. data for the same learner across a given

period of time), which is essential for broadening the scope of research questions that can be answered using corpus data. In a related vein, the range of language learning tasks that it would be relevant to collect data in order to obtain a representative view of language learning is something that is actively researched in Task-based Language Teaching and Learning – but this substantiated rich view on tasks has so far only had limited impact on the tasks used to collect large learner corpora in the field of Learner Corpus Research.

The second aim (**Challenge 2**) of the Action is to provide interdisciplinary answers to the many challenges identified under 1.1.1. by (1) working towards the development of standards and best practices in the collection of metadata, data annotation, research ethics, data protection, open science, and study quality; (2) tackling challenges posed by the increasing availability of diverse L2 data and novel ways of collecting these in the digital era (e.g., interoperability, variety of data formats), and (3) making existing and future L2 data (of various sorts) maximally accessible and re-usable for as wide a variety of stakeholders as possible (researchers, language teachers, textbook and other pedagogical material developers, policy makers, engineers, language assessors, speech therapists, etc.).

More generally, the Action will seek to promote collaboration and provide training around FAIR research principles and open science, which are yet to be developed in the many communities that will make up the Network. By creating interdisciplinary synergies around L2 data for many languages in Europe, the Action will also help answering long term EU language objectives, from increasing linguistic and cultural diversity and multilingualism to linguistic/social/economic/cultural integration/inclusion/participation (Saville & Gutierrez, 2016).

1.2 PROGRESS BEYOND THE STATE-OF-THE-ART

1.2.1 APPROACH TO THE CHALLENGE AND PROGRESS BEYOND THE STATE-OF-THE-ART

The challenges identified under Section 1.1.2 will mainly be addressed through network-building among researchers and stakeholders who collect, annotate, analyse or evaluate learner data in a wide variety of languages, as well as through interdisciplinary knowledge creation in the form of collaborative development of recommendations and best practices for a variety of purposes (e.g., data collection, data transcription, data annotation) across fields (see Sections 1.2.2 and 4). To this end, COST instruments such as Working Groups Meetings and Short Term Scientific Missions will be particularly useful (cf. 3.2.1). The Action will also build on a variety of existing efforts at national, European and international level (cf. 2.1.1).

Cross-fertilisation being hindered by discipline-specific training, the Action will also put strong focus on transfer of knowledge and organize Training Schools (with a particular focus on Early Career Investigators and Inclusiveness Target Countries) as well as promote joint PhD supervision by experts from all involved fields, thus educating the next generation of researchers to bridge the gap between fields. At master level, the Action will also facilitate international student mobility via Erasmus+ student exchanges and internship placements, as well as the creation of interdisciplinary double degrees (cf. 3.2).

1.2.2 OBJECTIVES

1.2.2.1 Research Coordination Objectives

Objective 1: Develop a common theoretical and methodological framework to answer Challenge 2, including the definition of a shared terminology and a joint set of minimal standards to enhance the quality and reusability of L2 data (WG1), the development of common ethical/legal procedures in conducting L2 research at all stages (WG2), the support of NLP development for other languages than English (WG3) and the promotion of study quality and OS practices (WG4). Indicators of completion e.g., for Objective 1 will be the successful completion of each WG's related tasks (see Section 4.1.2).

Objective 2: Promote innovative ways of L2 data collection (online, in the classroom, etc.) and analysis (based on the current state of the art in Natural Language Processing, Learning Analytics and Educational Data Mining) and pilot their implementation for a minimum of 5 European languages other than English (cf. D3.2.1). Indicators of completion for Objective 2 will be the successful completion of Tasks 3.2, 3.3 and 5.3.

Objective 3: Bridge gaps between researchers and practitioners by promoting synergies through active discussion with major stakeholders (language testers, teachers, policy makers, language teaching

material developers), and engagement with their practical needs. Indicators of completion for Objective 3 will be the successful completion of Tasks 5.2 and 5.3. Success will also be measurable in the number of non-academic stakeholders who participate actively in the Working Groups.

Objective 4: Provide opportunities for collaborative work by Action members, more particularly ECIs and members from ITCs, leading to presentations at conferences and scientific publications; and coordinate activities that encourage young researchers to establish links with industry and more senior academics. Indicators of completion for Objective 4 will be the number of completed STSMs, delivered conference presentations and publications supported by the COST Action.

Objective 5: Disseminate the results of the Action to stakeholders. This objective will be pursued through the Dissemination Plan (see Section 3.2.2) and its indicator of completion will be the successful creation of communication channels, the organization of events (training schools, workshops) and the effective dissemination of content through such communication channels and events (D7.1-3).

1.2.2.2 Capacity-building Objectives

Objective 1: Build an interdisciplinary core community of active stakeholders to foster interdisciplinary interpretation of the outlined challenges. This will be implemented through an outreach plan (Sect. 3.2.2), through the WGs (Sect. 4) and through innovative dissemination and exploitation activities devised to foster collaborations, with special focus on ECIs (Early Career Investigators) and ITCs (Inclusiveness Target Countries) (cf. Section 3.2.1).

Objective 2: Drive scientific progress by liaising extensively with (non-)government organisations, industrial stakeholders and other end-users of L2 data and derived materials. This will be implemented through all the WGs that will seek to include representatives of the different types of institutions.

The indicator of completion for Objectives 1 and 2 will be the release of reports describing, among other things, the balancing between member characteristics at Action and WG level (cf. D6.2).

Objective 3: Foster new funded initiatives to complement and practically incorporate the ideas generated by the networking stakeholders with respect to the shared research agenda. This objective aims to foster new initiatives complementing the Action and will be pursued by using the Action's achievements to plan parallel or follow-up projects. Its indicator of completion will be the successful preparation of several proposals to different regional, national, transnational (e.g., Open Research Area for France, UK and Germany) funding schemes as well as a Horizon 2020 proposal (D7.5-6).

2 NETWORKING EXCELLENCE

2.1 ADDED VALUE OF NETWORKING IN S&T EXCELLENCE

2.1.1 ADDED VALUE IN RELATION TO EXISTING EFFORTS AT EUROPEAN AND/OR INTERNATIONAL LEVEL

Research excellence can be achieved only if responsible, rigorous and innovative research practices are followed by researchers through all stages, from theorizing, data collection and processing, to sharing and dissemination. In this process, it is of paramount importance to develop an interdisciplinary network that will facilitate **knowledge exchange** by serving as a platform for centrally collecting and disseminating relevant information among researchers, teachers, test providers, publishers and policy makers. The Action will also serve to **promote innovative approaches** to L2 data collection, analysis and usage, and foster the development of a **joint research agenda** for L2-data-based research and practices.

Today, L2 data are largely collected individually or by small, local, or national initiatives, following in-house guidelines for data transcription/annotation, and they are usually not shared on open-access platforms (see CHILDES for an exception). As noted under Section 1.1.1, this raises several issues related to interoperability, reproducibility of the analyses, and reuse of data. In this context, there is clearly a need for the different research and practice communities represented in this Network to engage more with Open Access practices by providing structured and standardised L2 data that could be shared, enriched, and re-used, thereby answering the European Commission's call for more interoperable, reusable, and Open Access data.

Our proposed Action on L2 data brings together researchers who look at L2 data from different perspectives (corpora/experiments/linguistic theory/teaching practice), and who use different research

methods (quantitative/qualitative/mixed) to achieve their goals. This Action is thus much wider in scope than related discipline-specific initiatives geared towards enhancing the quality of L2 studies, such as IRIS - *A digital repository of instruments and materials for research into second languages* (through material and knowledge sharing), or *SLA for all? Reproducing SLA research in non-academic samples* (through emphasis on sampling and replication). At the same time, this Action is also much more focused than a variety of more general, large-scale European initiatives such as the Open Science Framework (OSF) ALLEA e-Humanities, or Social Sciences and Humanities Open Cloud (SSHOC). Despite representing important milestones, and probably due to their intended general usability, the applicability of such initiatives and direct benefits to language learning research and L2 data have not yet reached a threshold of wider acceptance.

To meet its objectives, the Action will build upon, and actively engage with, existing European infrastructures that focus on providing access to digital data, like CLARIN, Meta-Net, Meta-Share, or (USA-based) SLABank. We will make it a priority to make sure that such infrastructures can develop further nurturing on our insights and findings. CLARIN, for example, has already expressed interest in collaboration on L2 data management.

Building on previous and ongoing COST Actions focused on language learning (e.g., enetCollect, SAREP, European Literacy Network), this Action will continue developing a shared knowledge base for promoting new paradigms in research and practice on language learning, resulting – potentially – in larger funded initiatives (cf. D7.6).

Another main objective of the Action that rests on the quality of its network is the harmonization of research and ethical procedures in conducting research, with an aim to achieve excellence and integrity in research. Thus, for example, the new General Data Protection Regulation (GDPR) entered into force in May 2018 to provide consistent levels of data protection (i.e. stricter rules and more control over personal data) throughout the European Union. Despite the fact that GDPR is legally binding in all EU countries, the current situation in research ethics varies from country to country as the interpretation of GDPR may differ depending on different perspectives and research traditions in different states. By developing and achieving networking excellence, COST Action on L2 data will ensure that in a large number of EU countries, including a maximum number of Inclusiveness Target countries and near neighbour countries, ethical standards are observed and maintained at all stages of L2 research, from data collection to reporting and archiving.

2.2 ADDED VALUE OF NETWORKING IN IMPACT

2.2.1 SECURING THE CRITICAL MASS AND EXPERTISE

This Action brings together existing communities of researchers - already well represented in the proposed network and most likely to grow quickly - from traditionally different fields (Second Language Acquisition, Learner Corpus Research, Foreign Language Teaching, Language Testing, Natural Language Processing). These communities pursue similar fundamental questions regarding second language learning, teaching and development and work with similar data (i.e., L2 data) but they have developed from different research traditions and therefore work in fairly separate “worlds” (i.e., different conferences, different academic journals). With a few exceptions, these fields have largely failed to share insights, which means that a lot of similar work has and still is being done in parallel. The different disciplines all have their strengths and weaknesses in terms of L2 acquisition research, and can complement each other in this respect. In particular, Second Language Acquisition (SLA) researchers have a strong tradition in theory building based on experimental work investigating second/foreign language performance and development; learner corpus researchers have expertise in metadata collection, coding and describing large data sets, while NLP researchers are experts in encoding, annotating, enriching, cross-linking and mining big data. One of the most important potentials of this COST Action for research lies in building bridges between these separate traditions to allow for cross-fertilisation and complementary (in contrast to parallel) efforts drawing on each of the fields’ expertise.

A further aim of the proposed Action is to create fertile connections between practitioners in language pedagogy and assessment on the one hand and researchers on the other hand, in order to give language scientists access to the many forms of rich and authentic L2 data that is being produced in language classrooms, elicited by CALL tools (e.g., vocabulary apps), and assessed by language testing bodies. One target of the COST Action is to facilitate data sharing from these different stakeholders that are in daily contact with second language learners and test takers, with L2 data researchers. These sets of big L2 data form treasure troves that will potentially spark many new research initiatives. Receiving access to authentic L2 data and learning analytics allows the innovative compilation of new corpora that

can be used for further research. Specifically, these kinds of data will allow replication and testing of hypotheses (from traditionally small-scale empirical studies that have been performed in SLA) on larger scales both in terms of research population (i.e., data from many different participants and highly diverse participant groups learning a variety of target languages) as well as regarding time span (i.e., allowing longitudinal studies that investigate development over several months/years, ideally in trans-national comparative studies). Finally, working on big data will provide NLP researchers with the necessary scale of information to train and fine-tune the tools and measures they develop. In line with the European policies, a particular aim of the COST Action will be to work with large L2 data of a variety of source and target languages (e.g., Polish L2 data of French; Spanish L2 data of German) in order to support multilingualism and counter the bias towards English. For participating practitioners, this means that they can build close relationships with cutting-edge, theory-driven and innovative research, where they can bring forward their wishes from teaching/testing practice and guide scholars towards tools and procedures that will support their growth, sustainability and effectiveness.

The Network already involves 70 researchers and practitioners from c. 60 institutions (universities, (non-)government institutions and business companies) and 28 countries, who are at various stages of career development, from Early Career Investigators (38%) to Senior researchers with over 20 years of experience, and will continue to grow. Current proposers are all experts in their respective fields relating to L2 data. For example, expertise in ethical and legal issues is secured by the engagement of lawyers who are experts in data protection and GDPR issues, as well as researchers who work on research ethics in their respective institutions.

Special emphasis will be placed on (1) Early Career Investigators (38% of the current Network, a percentage that we will make sure to multiply by involving many more young scholars who are active in the various disciplines) as they will be the principal recipient of the new knowledge generated by the Action; and (2) stakeholders from Inclusiveness Target Countries (53% of the current Network, a percentage that is also likely to multiply quickly) who proved particularly eager to join the Network as they usually do not receive institutional support at the local or national level for answering the challenges addressed in the Action.

2.2.2 INVOLVEMENT OF STAKEHOLDERS

Stakeholders relevant to this Action are SLA, LCR and NLP researchers, research-active language educators, higher education and research institutions, international academic bodies engaged in the promotion of L2 study quality and L2 data use (e.g., the Learner Corpus Association, the European Second Language Association, the European Association of Language Testing and Assessment), language assessment practitioners, teachers and students, language policy makers, textbook and material developers, systems engineers, CALL developers, and academic publishers.

Non-academic stakeholders will be encouraged to get involved in the Action in a variety of specific ways: for example, the best way to engage teachers and have them collaborate is to show them that their students can learn better and become more independent learners with the resources based on L2 data. Some of the examples of successful L2 data use / L2 data elicitation are commercial applications such as Duolingo, as well as studies demonstrating multiple and long-term benefits of data-driven learning (e.g., Vyatkina, 2016). Language policy makers and language assessment bodies are best motivated by the examples and successes reached in some previous experiments, where real-life L2 data has been used to create tests or to define teaching/testing guidelines. Researchers of various kinds are easier to motivate since access to data opens the possibilities to fast and effective research. Networking and training events for teachers, policy makers, and test/material developers will be a way to show potential synergies of working together, as well as develop closer relationships and foster collaboration among these stakeholders.

The core Network at the moment of application comprises representatives of various disciplines concerned with the use and analysis of L2 data or products developed on top of the produced L2 data. Current communication with potential members of the Network proves that there is both interest and need to collaborate and share insights gained in the separate (and currently disjunct) disciplines. The backgrounds of enlisted core Network proposers is strong evidence of that.

The list of the core Network members has a number of practicing teachers (most of whom are teaching second languages to university students), SLA, NLP, LCR researchers as well as language policy makers and representatives of language testing bodies and business companies. Through these currently enrolled co-proposers, we intend to reach out to more stakeholders in their networks, as well as to promote our activities through the Action and WG websites, actively suggesting relevant Short-Term Scientific Missions and collaboration options.

Benefits to broader society will be ensured by the involvement of the most relevant stakeholders as data providers and data users: researchers, teachers and other education providers, test developers, publishers and policy makers. They will be involved as participants in conferences, workshops and webinars, and will also act as a link to data subjects (i.e. learners producing L2 data). Their involvement will ensure that social responsibility is maintained by showing respect for the privacy of individuals and communities who might be involved in research as data subjects. In this way, the main objectives of GDPR (control over own data and trust, lawfulness, fairness and transparency) will have a great chance to be achieved and maintained, and the practices harmonised among all European and some near neighbour countries.

2.2.3 MUTUAL BENEFITS OF THE INVOLVEMENT OF SECONDARY PROPOSERS FROM NEAR NEIGHBOUR OR INTERNATIONAL PARTNER COUNTRIES OR INTERNATIONAL ORGANISATIONS

The current Network does not include near-neighbour countries (mainly for lack of time) but it is anticipated that stakeholders from these countries will want to join the Action.

Collaboration at an international level seems warranted since the Action aims to address challenges that are not limited to Europe: the state of the art described under Section 1.1.1 holds across the world and international colleagues are confronted with the same issues. Given that members from international partner countries are not funded by COST Actions, and despite active collaboration especially with the USA, it was decided not to invite stakeholders from outside Europe at this stage.

3 IMPACT

3.1 IMPACT TO SCIENCE, SOCIETY AND COMPETITIVENESS, AND POTENTIAL FOR INNOVATION/BREAK-THROUGHS

3.1.1 SCIENTIFIC, TECHNOLOGICAL, AND/OR SOCIOECONOMIC IMPACTS (INCLUDING POTENTIAL INNOVATIONS AND/OR BREAKTHROUGHS)

(A) Current disparate practices with regards to study quality, data protection and ethics in research across Europe and near neighbour countries pose a major challenge to both researchers and the wider society. This is mainly due to diverse traditions in research, or the lack thereof in some ITCs and near neighbouring countries. In order to tackle this issue, the proposed COST Action with its tasks and deliverables will have significant impact in both COST members and IT countries. It will raise the levels of competitiveness particularly of ITCs and near neighbour countries, while a large number of countries included in the Action has in itself a high potential for innovation and break-throughs. The stimulation of change towards **more rigorous, open, transparent and reproducible science** in the L2 research community is envisaged through a series of impact-generating activities strategically prioritised at different stages. The key goals for the impact agenda are (a) to bring new evidence about current approaches to research in the European L2-related research community, (b) to develop a set of recommendations and optimally tailor guidance which researchers can use to regulate the workflow of their projects overcoming current weaknesses in research practices, and (c) to raise awareness of the fast developing open source architecture and its robust potential for scientific innovation through transforming behavioural norms (e.g., via online platforms for collaborative research such as IRIS or StudySwap). This agenda will be guided by a set of principles, the main ones being the inclusion of beneficiaries in shaping the pathways to impact as well as targeted dissemination. The most relevant beneficiaries are researchers, language educators and the general public. This Action's ambition is to achieve high impact by breaking new ground in an improved and widely-shared understanding of what it means to practice rigorous, transparent and reproducible SLA research. The planned activities most directly target individuals involved in research, but via the individuals they are expected to eventually also have an impact on the publication system (reduction of publication bias toward the "new" and "significant", and higher overall quality standards in relevant journals), and on the institutional education of future L2 researchers.

(B) Regarding **technology and tools**, the COST Action aims to support standardisation in terms of data collection (e.g., what kind of tasks and learner characteristics are minimally required); in terms of data annotation (e.g., what coding and annotation standards are applied to collected data) to allow for further and wide-spread analyses; and, in terms of research data management and data repositories where L2 data are stored. In addition, new procedures will be established and existing processes will be fine-

tuned that provide scientifically sound and ethically uncontroversial ways of collecting L2 data from a large number of learners (e.g., data from social media; crowd-sourcing; data from online language learning tools and large-scale international language assessment tasks) and learner analytics. Technological impact will also be achieved through the development of workflows that aim to ease the access to shared L2 data that has been collected. In this sense, the COST Action will fertilise cross-disciplinary work on L2 data that goes beyond the boundaries of research as collaboration with stakeholders in L2 education and assessment will ensure tools and technological solutions are user friendly and useful for practitioners too.

(C) This COST Action aims to create **societal** impact in many ways. Most importantly, different stakeholders of second language pedagogy and instructed second language acquisition will benefit from the outcomes. For example, language educators and policy makers will benefit from the insights the Action creates for their core business, i.e., supporting language learners. Similarly, language testing bodies will be able to draw on the outcomes of scientific endeavours of COST Action partners. Importantly, the societal impact will be mutual. That is, on the one hand, stakeholders working with language learners in their day-to-day life (e.g., teachers, assessment bodies) will be encouraged to support the data collection needs of researchers. On the other hand, given that collaborations (for example, between researchers and teachers) are established for data collection, feeding back relevant outcomes from research into education and testing will be a natural component of the joint efforts. As a result, the COST Action will create avenues for research questions to be explored that were formulated by the societal stakeholders based on their practical needs and knowledge of the field as a multilingual society and intercultural business environment. Last but not least, in line with European policies, a particular aim to the COST Action will be to work with large L2 data of a variety of sources and target languages (e.g., Polish L2 data of French; Spanish L2 data of German) in order to support multilingualism and enrich existing work, which is biased towards English (cf. the 2002 Barcelona Objectives, which stated the aim of having all EU citizens speak two additional languages).

3.2 MEASURES TO MAXIMISE IMPACT

3.2.1 KNOWLEDGE CREATION, TRANSFER OF KNOWLEDGE AND CAREER DEVELOPMENT

Measures to maximize impact in knowledge creation will involve:

(A) Regular Working Group meetings for the **scientific** fields involved, where the aim is to (i) update each other on the developments in the different communities allowing for merging cutting-edge interdisciplinary insights, methods and resources from complementary fields; (ii) identify specific joint questions that will be pursued, for example, in joint applications for research funding.

(B) To make sure that questions from practitioners in **society** are heard and acted upon and that insights from research feed into practice in language education and assessment, the COST Action aims to support (i) meetings with practitioners (teachers, policy makers, assessment bodies, language teaching (tool) developers) and researchers where the aim is to identify questions and needs of practitioners to spark new research; (ii) hands-on workshops (and continuous online support) for various practitioner groups where the aim is to integrate research outcomes into language teaching and assessment practice; (iii) the development of intervention studies where the efficacy and effectiveness of research findings could be put to the test in authentic contexts in order to evaluate research outcomes according to educational standards.

(C) **Identification of needs and planning a fitting response (short term)**. One of the steps to maximise impact will be to monitor Open Science demand from scholars and research-active language educators through an OS practices survey (circulated to members of a large network of professionals such as the European Second Language Association and the Learner Corpus Association). This step will provide fresh perspectives on myths about open research (e.g., visualising failure; idea 'scoopability') as well as on how best to bust them.

Measures to maximize impact in transfer of knowledge will involve:

(D) **Capacity building** via workshops, training schools and conferences will promote the idea of fair, transparent and lawful management and sharing of L2 data, having a common benefit in mind. An important measure to maximise impact by knowledge exchange will be the formation of **Knowledge Transfer Partnerships** (KTP) that will aim at connecting the stakeholders in the triangle: Academia – Businesses – Society (i.e., researchers – test providers, publishers – teachers, educators). The benefit for collaborators/partners will be demonstrated by the increased applicability of gained knowledge and expertise in processing, managing and sharing data in an ethical, fair, and lawful manner.

Impact on science and technology will be demonstrated and measured by increased expertise particularly among IT countries, but also among COST members who have not yet developed GDPR fully compliant procedures in conducting research. Knowledge Transfer Partnerships will benefit all involved in these collaborations, which has a potential to reflect on scientific and socio-economic development. Apart from researchers, other stakeholders such as teachers, publishers and test providers will also benefit from KTP. This will be manifested such that their increased knowledge will result in better understanding of research, its benefits for teaching and testing practices and, overall, their involvement will allow for diversifying the pool of potential research participants and collection of larger amounts of research data. Furthermore, it will also allow for achieving more transparency, fairness and lawfulness in dealing with L2 data, enabling the members of wider society to have more control over their data.

(E) **Knowledge transfer through online presence (medium and long-term).** A website will be set up with the aim to summarise the activities in alignment with the set deliverables, to raise the project profile and ensure its continuous visibility through communicating research findings, details about related events and summaries of specific content areas separately for each target group of beneficiaries. The site will also include a Resources tab with regular updates on OS training opportunities, methodological courses and links to shareable open software (e.g., for experiment building).

Measures to maximize impact in career development (also based on transfer of knowledge) will involve:

(F) To support **Early Career Investigators** as well as more established members in the different communities, the COST Action aims to (i) organize annual **Training Schools** where those teaching should come from different fields such that those attending receive insights from the different fields and are trained in the various methods used. In this way, the future generation of L2 data researchers will become knowledgeable and skilled in a truly transdisciplinary way; (ii) support **Short Term Scientific Missions** (STSM) where individual researchers can spend some time at a complementary institution in order to learn from the different methodological approaches and perspectives. STSMs also have the potential to lead to shared efforts to identify gaps in existing research and to address joint questions in collaborative research and applications for funding thereof.

(G) To enhance **training** at European universities, Action's members will be encouraged to send MA students to partner universities with Erasmus funding, work towards the development of joint/double MA degrees and co-supervise PhD candidates. The set of guidelines and recommendations produced by the Action will be of direct use for university training, with the potential to revamp the curriculum in degrees such as applied linguistics.

3.2.2 PLAN FOR DISSEMINATION AND/OR EXPLOITATION AND DIALOGUE WITH THE GENERAL PUBLIC OR POLICY

The Dissemination, Exploitation, and Outreach Plan will be elaborated according to COST guidelines and the Action objectives, and will be revised regularly. The Action with its tasks and deliverables will provide the conditions to disseminate the information efficiently and to as large an audience as possible. Each WG will be responsible for its own dissemination strategy, but there will also be a plan for the whole Action.

Tools and activities for dissemination:

- An online platform to present the Action, its activities and results.
- Targeted video "teaser talks" with feedback options directed at the public and at stakeholders informing them about the goals, activities and results of the network.
- Press releases and social media content production for disseminating results and activities to the public and for reaching/involving stakeholders.
- Annual scientific conference with peer-reviewed open-access proceedings.
- Training Schools with sections involving different stakeholders.
- Video recordings of selected talks from the annual conferences and Training Schools will be published online and advertised.
- "How to" videos on L2 data collection and use to be distributed on the Action website and social media
- Hackathon for practical experiments in adapting NLP tools for learner language data.
- Publicly accessible online directory of relevant stakeholders with contact information.
- The scientific results will be presented in conference sections and individual papers at national and international conferences as well as in peer-reviewed academic journals.
- The Action will make all outputs, milestones and deliverables available as Open Data or under Creative Commons licences and deposit them on platforms that allow for long-term archiving.

In line with the Open Accessible Summaries In Language Studies (OASIS) project, the Action will work towards the simplification of its output for wide dissemination.

Exploitation plan:

- With the aim to improve practices in L2 research, WG members will integrate the highest standards of intellectual precision to co-author open access articles that will target high-impact journals. Showcasing best OS practices, the articles will share reusable materials (IRIS), replicable tasks (OSF), and the data will be made available for secondary analyses.
- Training materials, ethics documentation and ethical Code of Conduct will be made publicly available on the Open access platform and widely advertised among researchers and other stakeholders.
- Guidelines for minimal set of necessary knowledge and know-how to be taught at master level (for potential revision of curriculum)
- Integration of recommended representation and quality standards into future language resource initiatives.
- Proposals for follow-up national and international research and development projects.
- Large follow-up proposal (e.g., Horizon2020 Infra) for an integrated web interface for online querying a broad range of corpora for different languages as a resource for the scientific community, for educational stakeholders, for NLP companies and for the public.

Communication with the general public. The project website will provide a link to a form for feedback, and it will highlight the opportunity to ask questions about any aspect of the Action. Outreach activities on the internet will be complemented with press releases and with summaries publicised on social media (Instagram, Twitter, Facebook) and via e-newsletters. Contacts will be made with media, so that the activities undertaken in this Action reach as wide as possible audiences.

In addition, we aim to disseminate the outcomes and insights of the COST Action via the following activities and channels to the wider public and policy makers:

- Members of the COST Action will attend practitioner fairs in the field of language education and technology (e.g., [Didacta](#), [LearnTec](#), [Expolingua](#)) to establish networks beyond the scientific communities.
- The Action's outputs that seem to be specifically useful for the wider public (e.g., in how they relate to the objectives of the European language portfolio) will be made available to all interested stakeholders.
- Newsletters for the public will be sent out on a regular basis informing about progress and possibilities.
- When reaching larger milestones traditional media will be addressed as well.
- Outcomes that would support European policies, most particularly the stated objective that all European citizens should speak two additional languages other than their mother tongue (L1+2), will be shared with policy makers in charge.

4 IMPLEMENTATION

4.1 COHERENCE AND EFFECTIVENESS OF THE WORK PLAN

4.1.1 DESCRIPTION OF WORKING GROUPS, TASKS AND ACTIVITIES

[WG1] Data quality

WG1 is concerned with maximising L2 data impact across fields, in research (e.g., second language acquisition, natural language processing) and practice (e.g., language education and language testing). To this end, WG1 will address issues related to data quality, sharing and re-use, and work towards the development of shared principles and 'best practices' related to data and metadata collection, data transcription, annotation, measurement and documentation. It will also produce conversion guidelines that will be necessary to convert metadata, transcription and annotation in one field to those in the agreed standards.

[WG2] Legal and ethical aspects

WG2 is concerned with the questions of integrity and maintenance of ethical principles in conducting L2 research, with a particular emphasis on upholding the principles of the European General Data Protection Regulation (2018) across Europe and among various stakeholders interested in L2 data.

Specifically, the WG will promote Europe wide knowledge exchange, capacity building and raising awareness about the principles of GDPR that refer to lawfulness, fairness and transparency in collecting, processing, managing and sharing L2 data. This group aims at actively contributing to the issues arising from the need to implement the new Regulation in all areas of research, with a special interest in assisting the development of best and responsible practices in Open Science.

[WG3] L2 data processing and enrichment

The main objectives of WG3 are to review currently available NLP tools and techniques that were often developed and tested on the basis of L1 data, evaluate their performance on L2 data and work towards their improvement in collaboration with WG1 (e.g., enrichment of L2 data with annotation and metadata) and WG2 (e.g., automatic approaches to anonymization). The WG will place emphasis on tools and techniques developed for other L2 languages than English as these tools often form the basis of many kinds of further corpus processing and analyses of learner language data that are still currently lacking for languages such as Bulgarian, Croatian or Slovenian. Special attention will therefore be placed on NLP for the wide range of languages used in Europe.

[WG4] Research reproducibility, replicability and transparency

WG4 aims at improving quality standards for research studies based on L2 data. This will primarily be achieved with the help of COST training instruments. The main objectives of WG4 will be to advance rigour in research methodology (including quantitative and qualitative analyses, as well as mixed methods), enhance the understanding of the value of reflexive and cumulative research and transparent research design, and promote the as-yet-underused affordances of OS practices in the wider domain of L2 research. The approach will be critical, interdisciplinary and cross-regional, encouraging changes in national and international research and publication systems.

[WG5] Maximising impact for research and society

WG5 aims at creating fertile connections between researchers and practitioners in language teaching and assessment by integrating and advancing perspectives on L2 data collection, and facilitating the development of studies on language learning in ecologically valid environments (e.g. classrooms and language assessment contexts). This will generate new, richly embedded types of L2 data and will enhance the impact of L2 data on research and practices. WG5 will compile a set of scenarios that can be integrated into language learning classrooms in the form of digital activities. It will foster a “new philosophy” in the teaching and testing practices for the involved stakeholders encouraging uptake of new methods and workflows.

4.1.2 DESCRIPTION OF DELIVERABLES AND TIMEFRAME

WG1: Data quality		
Overall objectives:	Development of shared principles and ‘best practices’ related to data and metadata collection, data transcription, annotation, measurement and documentation	
Task 1.1: Examine and review the current situation: What are current practices and needs of data collection, handling and sharing in each field		Month 1–12
<u>Activities:</u>	A systematic review on the existing standards within each of the disciplines (SLA, LCR, NLP), complemented by discussion groups and surveys with stakeholders.	
<u>Milestones</u>	Inventory of current practice and needs that highlights where convergence might be easily made (e.g., agree on a standardized name for the same phenomenon) while also bring to the foreground aspects that need bridging.	
<u>Deliverable 1.1</u>	Summary report (and/or academic paper) on review identifying features where there is a lot of divergence and where the Action will need to work towards procedures and tools that bridge gaps and allow convergence.	
Task 1.2: Agreed set of (minimal) standards		Month 13–24
<u>Activities:</u>	(1) Work towards agreement on a minimal set of standards on those features that are shared and those where the fields could learn from each other. (2) Meetings as a result, work towards a set of features where convergence is not needed or very difficult to establish. For the latter cases, these meetings will pursue the goal to come up with creating conversion tools and documentation that allows for comparison across fields.	
<u>Milestones</u>	Minimal set of standards identified for (a) data collection (e.g., metadata)	

	(b) data handling (e.g., transcription, annotation) in collaboration with WG2 (c) data sharing (conform OS) in collaboration with WG2 and WG4	
<u>Deliverable 1.2</u>	Document (digital) that specifies minimal standards to guide stakeholders in data collection, handling and sharing.	
Task 1.3: Training of (future generations of) researchers		Month 12–48
<u>Activities:</u>	(1) Organise training schools to educate (future generations of) researchers (2) Organise joint PhD student supervision by experts from all involved fields.	
<u>Milestones</u>	(1) Minimal set of standards (Deliverable 1.2) (2) Inventory of (complementary) expertise across universities involved in the network	
<u>Deliverable 1.3a-c</u>	(a) Training school on shared principles and 'best practices' related to data and metadata collection and documentation (c. month 24) (b) Training school on shared principles and 'best practices' related to data transcription, annotation and measurement (c. month 36) (c) Joint PhD programme agreements	
WG2: Legal and ethical aspects		
<u>Overall objectives:</u>	Development of shared principles and 'best practices' related to data and metadata collection, data transcription, annotation, measurement and documentation	
Task 2.1: Examine and review the current situation: how are ethical issues dealt with among European researchers?		Month 1–12
<u>Activities:</u>	A survey to be sent out to a large number of universities in Europe (in each country); templates for ethical consent to be collected from different institutions/universities	
<u>Milestones</u>	Analysis of the survey responses. Presentation of the results at a conference.	
<u>Deliverable 2.1</u>	An academic paper on the current state of L2 data ethical treatment among European researchers, and associated countries.	
Task 2.2: Knowledge exchange, capacity building in dealing with ethics for different types of data and different types of research (e.g., qualitative, quantitative, Open Science, corpus data, research using data from social media, etc)		Month 13–24
<u>Activities:</u>	Organisation of workshops in ITCs, and a conference. Liaising with experts to provide expert advice and training.	
<u>Milestones</u>	Reaching a number of researchers (100) from ITCs and neighbouring countries, as well as from other COST countries involved as participants in workshops and the conference	
<u>Deliverable 2.2</u>	Bank of protocols and templates for consent to be used when conducting research (to be available via an open access platform)	
Task 2.3: Harmonisation and formalisation of GDPR requirements in dealing with L2 data across Europe		Month 25–48
<u>Activities:</u>	Organisation of a series of meetings and workshops to develop community-wide understanding of legal definitions and requirements under GDPR (such as Fairness, legitimate interest, appropriate safeguards), and formalise these in a Code of Conduct.	
<u>Milestones</u>	First draft of Code of Conduct produced, and submitted to associations for comments and feedback.	
<u>Deliverable 2.3</u>	Code of Conduct submitted to European Data Protection Board for approval, and made available via an open access platform	
WG3: L2 data processing and enrichment		
<u>Overall objectives:</u>	review currently available NLP tools and techniques that were often developed and tested on the basis of L1 data, evaluate their performance on L2 data and work towards their improvement	
Task 3.1: Overview of existing models and tools for automatic processing, linguistic and error annotation and analyses of language learner corpora		Month 1–18
<u>Activities:</u>	Review of published research, available corpus and data management tools	
<u>Milestones</u>	Collaboration with WG1 on a review of annotation taxonomies available in current tools and projects Collaboration with WG2 on a review of automatic anonymization procedures	
<u>Deliverable 3.1</u>	Summary report on existing formats and tools for automatic processing of L2 data	
Task 3.2: Recommendations of standards for NLP processing of learner corpus data		Month 13–48

<u>Activities:</u>	Establish requirements regarding the special properties of learner language resources (e.g., rich metadata, annotation layers, multimodal, legal aspects) Organizing a hackathon and summarizing its results
<u>Deliverable 3.2.1</u>	Organisation of a hackathon about collecting and processing learner data
<u>Deliverable 3.2.2</u>	Recommendations for adapting resources and integrating them into corpus research and management tools
<u>Deliverable 3.2.3</u>	Guidelines for the creation and collection of datasets for NLP tools
Task 3.3: Develop NLP workflow for quality check of L2 corpora: legal and ethical standards	
<u>Activities:</u>	Designing automatic workflow for ethical use of L2 data according to recommendations of WG2
<u>Milestones</u>	Presentation of the designed workflow to COST members for feedback. Presentations at conferences for the different disciplines represented in the Action.
<u>Deliverable 3.3</u>	White paper documenting best practices implemented in corpus projects associated with the Action
WG4: Addressing challenges: reproducibility, replicability and transparency	
Overall objectives:	Improve quality standards for research studies based on L2 data
Task 4.1: Explore current awareness and use of Open Science practices and related needs in the wider L2 community	
<u>Activities:</u>	Conduct an online survey about current awareness and needs related to OS practices
<u>Milestones</u>	Completion of survey data collection and analysis
<u>Deliverable 4.1.</u>	Publication of survey results in a journal paper
Task 4.2: Meta-analysis-based best practice recommendations for study quality	
<u>Activities:</u>	Systematic review of measures used in existing work to ensure study quality and deduction of best practice recommendations
<u>Milestones</u>	Best practice recommendations are developed
<u>Deliverable 4.2</u>	Publication of position paper
Task 4.3: Support researchers to adopt or develop best practices in research methodology and Open Science	
<u>Activities:</u>	Training researchers and providing sustainable information on best practice
<u>Milestones</u>	Organize and conduct methodology training school and determine the structure and content of a dedicated website
<u>Deliverable 4.3.1</u>	Research methodology and OS training school
<u>Deliverable 4.3.2</u>	A website with information about best practices in research methodology and OS
Task 4.4: Critical, interdisciplinary discussion on advances and limitations brought along by the methodological turn	
<u>Activities:</u>	Interdisciplinary research conference on advances and limitations connected to the methodological turn
<u>Milestones</u>	Organize and conduct conference
<u>Deliverable 4.4.</u>	An interdisciplinary research conference
WG5: Maximising impact for research and society	
Overall objectives:	Create fertile connections between researchers and practitioners in language education and assessment
Task 5.1: Establishing a common ground: Approaches to the use of technology for L2 data collection and analysis	
<u>Activities:</u>	Review of current status of usage of (I)CALL and other language applications in real-life L2 data contexts (assessment, classroom, self-studies); surveys, discussions, experience sharing through workshops, analysis and structuring of the collected

	information.
<u>Milestones</u>	Identify joint aims, connections between how different stakeholders use different types of resources and methods & approaches for compiling, analyzing, visualizing L2 data
<u>Deliverable 5.1</u>	Report summarizing the state-of-the-art, open demands, and where combining and integrating perspectives is particularly fruitful / required to make progress and maximize impact on research and society
Task 5.2: Fostering interaction between research, practice, and language policy stakeholders	
	Month 12-36
<u>Activities</u>	Collect scenarios focusing on fostering novel ways of data collection and analysis. Experience sharing, hypothesis building, surveys, hackathons.
<u>Milestones</u>	Collection and evaluation of research scenarios and their integration into the classroom/testing/digital platform environments, incl formative and summative assessment. Analysis of potential pitfalls and ways to address those. Outline of the characteristics of (I)CALL systems supporting real-life assessment, teaching and learning & data collection as basis for (I)SLA research
<u>Deliverable 5.2</u>	Summary report with strategies, scenarios and suggestions on integration of research questions into language learning and assessment environments
Task 5.3: Facilitating the creation and integration of tools and interfaces for stakeholders in research and practice	
	Month 24-48
<u>Activities</u>	Conceptually: identify needs of researchers and practitioners of different profiles and suggest “guidelines” for how to set up user-friendly interfaces (and infrastructures) and search facilities to the L2 data collection and exploration. Form: hackathons and workshops.
<u>Milestones</u>	Theoretical framework, guidelines for setting up interfaces / infrastructures from a user/researcher point of view, design plans for prototypical setup
<u>Deliverable 5.3</u>	Design plans & guidelines for prototypical interfaces / infrastructures for collection, uploading & exploration of L2 data

WG transversal Tasks and Deliverables

Task 6: Management		Month
D6.1	General meetings of all network members	bi-annually
D6.2	Activity reports (including description of Network)	bi-annually
Task 7: Dissemination, exploitation and outreach		Month
D7.1	Communication means including project website and presence on social media	12
D7.2	Targeted press releases on the Action activities	12, 30, 45
D7.3	Newsletters	annually
D7.4	Training schools on the automated collection of L2 data, together with hackathons (WG3 with the collaboration of WG1 and WG2)	37, 48
D7.5	Follow-up / complementary national project proposals	24, 36
D7.6	Follow-up European project proposal (follow up programme to Horizon 2020)	48

4.1.3 RISK ANALYSIS AND CONTINGENCY PLANS

The risks related to the objectives of the Action will be monitored and controlled. There are several risks that could be identified within the tasks of the WGs and which are discussed below. Since the strategy defined for reaching the Action's objectives systematically includes the reaction to new input (produced by different forms of end-user involvement and evaluation) as part of the process, the concept is expected to be flexible enough to react to these risks and treat them in a way which will neither affect the work plan nor the short- and long-term success of the Action.

Coordination-related risks: As the experience of other COST actions prompts, the effectiveness of a network relies to a great degree on the enthusiasm and involvement of the leading figures in the action. To ensure that the coordination of the L2data4All action is driven towards the aims of the action, the following measures have been taken: leading figures in the five fields, equivalent to the number of working groups (WGs), have been identified and invited to a starting workshop prior to proposal writing. Care has been taken (1) to invite people active in the areas that have been proposed to them for leadership, so that awareness of the challenges as well as main players in the field could be brought into the action; and (2) to set up WG leadership consisting of at least two persons, preferably with different research profiles to safeguard continuity and versatility of views on the same issues. In case one of the WG leaders is unavailable or needs to drop-out, the continuity of the group work is safeguarded by the vice-leader stepping forward, and substituting the vacant place with an identified new player. This policy on maintaining the double leadership and background versatility of the leading figures will be promoted throughout the action, so that the planned group work is kept on track.

Task-related risks: Since the work of the COST actions is built around community-building, spreading new ideas and finding new synergies (rather than doing research or implementations), the main proposed mechanisms for achieving the aims are surveys, common guidelines and technical specifications. The primary risk with surveys is biased sampling due to participant non-response to the initial survey. The WGs will mitigate this risk through purposive sampling, working with a series of partner organisations (e.g., CLARIN) with established networks, and promotion of the benefits of harmonised standards, procedures and added value of versatility of views. With respect to common guidelines and technical specifications one estimated risk is that the Action members experience difficulty in reaching a consensus on best practice across different national and institutional frameworks. This risk is estimated as low because previous cooperation in COST actions has shown that there is overlap regarding ideas about what may be reasonable and practical ways to go about these issues. Another risk is that the nature of the tasks and challenges might be too heterogeneous with respect to the different languages represented in the network, or the conventions and traditions of the established language or country-specific research communities (e.g., need/tradition to use a certain tagger and/or tag set). This risk will be treated by, first of all, setting a "minimal" set of as that can be shared between the involved languages standards as the primary goal, and further focusing on methods that are language-independent or on language-specific solutions.

Community-building/Expertise risks: A further risk is connected to successful knowledge exchange and capacity building, which is twofold: insufficient numbers of researchers from ITCs, and lack of required expertise. The first risk will be mitigated by again working with partner institutions with established networks in ITCs, and the second through dedicated enrolment of the needed experts into the action through management structures that include Dissemination, Exploitation and Outreach strategies. The risks will be monitored through the action lifetime.

4.1.4 GANTT Diagram

Months	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48
Deliverables																
WG1				D1.1				D1.2								D1.3a-c
WG2				D2.1				D2.2								D2.3
WG3						D3.1				D3.2.1			D3.2.2		D3.2.3	D3.3
WG4				D4.1				D4.2		D4.3.1		D4.3.2		D4.4		
WG5								D5.1				D5.2				D5.3
WG1-WG5 transversal		D6.1-2		D6.1-2		D6.1-2		D6.1-2		D6.1-2		D6.1-2		D6.1-2		D6.1-2
			D7.1-2-3					D7.3-5		D7.2		D7.3-5	D7.4		D7.2	D7.3-4-6

References

- Al-Hoorie, A. H., & Vitta, J. P. (2018). The seven sins of L2 research: A review of 30 journals' statistical quality and their CiteScore, SJR, SNIP, JCR impact factors. *Language Teaching Research*. <https://org/10.1177/1362168818767191>
- Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing Natural Language Processing techniques. *Language Learning*, 67(S1), 180-208.
- Alexopoulou, T., Geertzen, J., Korhonen, A. & Meurers, D. (2015). Exploring big educational learner corpora for SLA research: Perspectives on relative clauses. *International Journal of Learner Corpus Research*, 1(1), 96-129. <http://dx.doi.org/10.1075/ijlcr.1.1.04ale>
- European Commission. (2019) Cost of not having FAIR research data. Luxembourg: Office for Official Publications of the European Communities. Retrieved from <https://publications.europa.eu/s/IKIM>
- Glaznieks, A., Abel, A., Lyding, V., Nicolas, L. & Stemle, E. (2014). Establishing a standardised procedure for building learner corpora. *Apples - Journal of Applied Language Studies* 8(3), 5-20.
- Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E. J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, 1(1), 70-85.
- Khany, R., & Tazik, K. (2019). Levels of statistical use in applied linguistics research articles: From 1986 to 2015. *Journal of Quantitative Linguistics*, 26(1), 48–65.
- MacWhinney, B. (2017). A shared platform for studying Second Language Acquisition. *Language Learning*, 67(S1), 254–275. <https://doi.org/10.1111/lang.12220>
- Marsden, E., & Plonsky, L. (2018). Data, open science, and methodological reform in Second Language Acquisition research. In A. Gudmestad & A. Edmonds (Eds.), *Critical Reflections on Data in Second Language Acquisition* (pp. 219–228). Amsterdam: Benjamins.
- Megyesi, B., Granstedt, L., Johansson, S., Prentice, J., Rosén, D., Schenström, C. J., Sundberg, G., Wirén, M. & Volodina, E. (2018). Learner Corpus Anonymization in the Age of GDPR: Insights from the Creation of a Learner Corpus of Swedish. In *Proceedings of the 7th workshop on NLP for Computer Assisted Language Learning* (pp. 47-56). <https://aclweb.org/anthology/W18-7106>
- Meurers, D., & Dickinson, M. (2017). Evidence and interpretation in language learning research: Opportunities for collaboration with computational linguistics. *Language Learning*, 67(S1), 66–95. <https://doi.org/10.1111/lang.12233>
- Open Science Collaboration, & Kappes, H. B. (2014). The Reproducibility Project: A model of large-scale collaboration for empirical research on reproducibility. In V. Stodden, F. Leisch & R. D. Peng (Eds.), *Implementing Reproducible Research* (pp. 299–324). New York: CRC Press.
- Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research*, 3(1), 61–94. <https://doi.org/10.1075/ijlcr.3.1.03pag>
- Rebuschat, P., Meurers, D. & McEnery, T. (2017). Language learning research at the intersection of experimental, computational, and corpus-based approaches. *Language Learning*, 67(S1), 6-13.
- Ross, M. W., Iguchi, M. Y., & Panicker, S. (2018). Ethical aspects of data sharing and research participant protections. *American Psychologist*, 73(2), 138–145. <https://doi.org/10.1037/amp0000240>
- Thomas, M., & Pettitt, N. (2017). Informed consent in research on second language acquisition. *Second Language Research*, 33(2), 271-288.
- Tracy-Ventura, N. & Myles, F. (2015). The importance of task variability in the design of learner corpora for SLA research. *International Journal of Learner Corpus Research*, 1(1), 58-95.
- Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS one*, 8(7), e67332.
- van Rooy, B. (2015). Annotating learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 79–106).

COST Mission and Policies

Promoting and spreading excellence

L2data4all will drive innovation, multiply returns on public investment, and promote excellence in second language (L2) research by maximising the impact of L2 data for a wide range of stakeholders, including scientific researchers, language educators, materials developers and language policy specialists, and commercial developers of speech and language technology. It will do so by using the networking tools provided by COST, namely networking in larger groups (action meetings), smaller groups (work group meetings/workshops, organization of tutorials and summer schools), practical individual research (Short-Term Scientific Missions) and support for dissemination of outputs through social media, dedicated website/platform, research conferences, etc. The Action will promote geographical, age and gender balance.

L2data4all has secured a network of 70 proposers representing 28 countries with 53% of Inclusiveness Target Countries (ITCs) and 27 young researchers, from a broad variety of stakeholders covering research (26), education (34), business (6), as well as private (4), government (2) and non-profit (2) sectors. The range of expertise in the network, including the technological expertise of big data scientists, will foster new and more effective solutions for the exploitation of large, multidimensional datasets, and guide a new generation of young experts.

L2data4all will influence regulatory bodies, policy makers and national decision makers, and impact the private sector by involving stakeholders from different fields, institutions and countries. Its commitment to FAIR data, Open Access, Open Data and Open Science also allow *L2data4all* to reach, involve and impact more stakeholders, and provide an implementation mechanism for HorizonEurope goals.

L2data4all will promote and spread excellence most particularly in Inclusiveness Target Countries by organising networking events (training schools, group meetings, etc.) in these countries.

Fostering interdisciplinary research for breakthrough science

L2data4all will generate powerful synergies by coordinating research across disciplines and national borders. To date, divergent disciplinary perspectives, purposes, and methods have hindered dialogue within or between fields that use L2 data. The action will enable L2 researchers to overcome the structural barriers to sustainability and excellence posed by disparate and often mutually incompatible data formats, databases, metadata, licences, and search tools.

L2data4all will enable breakthrough scientific developments by agreeing on standards and creating tools for researchers to make their data usable, promote their data, and clarify how users can reuse their data. The synergies generated by bringing different research communities together will enable the network not only to accelerate the scientific research cycle, but also to address (1) key societal challenges linked to increasing linguistic diversity and need for linguistic inclusion, education and integration at the European level, and (2) the need to ensure that advances in data collection and exploitation (e.g. webscraping) are accompanied by the highest standards of ethical and legal behaviour.

Empowering and retaining young researchers and innovators

L2data4all will empower a new generation of researchers and innovators that work with L2 data. This Action will put strong focus on transfer of knowledge and organize Training Schools (with a particular focus on Early Career Investigators and Inclusiveness Target Countries) as well as promote joint PhD supervision by experts from all involved fields, thus educating the next generation of researchers to bridge the gap between fields. At master level, the Action will also facilitate international student mobility via Erasmus+ student exchanges and internship placements, as well as the creation of interdisciplinary double degrees. The Action is inclusive and includes people of different age groups, gender and location, to maximise benefits for all.

Involvement of ITC countries and researchers

L2data4all complies with the open and inclusive policy of COST, and has up to now made efforts to reach out to researchers active in ITC countries (15 ITC countries on board at the moment of proposal) and an effort will be made to expand ITC involvement further both in terms of number of ITC countries and number

of individual researchers. Researchers from ITCs will be encouraged to take up leading roles in the Action to foster their career development and enhance visibility of ITCs' S&T development across European borders.

During the Action lifetime, *L2data4all* has the potential to increase the involvement of COST Near Neighbour Countries.

Addressing European challenges

L2data4all will enhance conditions for using learner language data in novel and innovative ways to address major European challenges related to foreign language use by all its citizens, most particularly:

- Europe's stated aim of having all EU citizens speak two additional languages (cf. Barcelona Objectives 2002 and "Conclusions on Multilingualism and the Development of Language Competences, Council of Europe, 2014)
- Europe's recognition that there is a need for uniform, valid and reliable language testing across Europe, as foreign language tests can have major societal and individual consequences (e.g. access to university, access to immigration) (Study on comparability of language testing in Europe, European Commission, Final Report, September 2015).

To this end, the action will enhance conditions for using language learner data for empirical research in the Humanities and Social Sciences and for developing innovative Language Technology tools for a wide range of languages other than English. In collaboration with stakeholders, the action will create and advance ethical and technical standards and quality criteria for building and providing collections of language data (corpora) alongside best practices recommended by acknowledged language resource infrastructure initiatives. The action results will thus pave the ground for establishing a sustainable international infrastructure for L2 data.

Network of Proposers - Features

COST Inclusiveness target countries

53.57 %

Number of Proposers

70

Geographic Distribution of Proposers

Country	ITC/ non ITC/ other	Number of institutions from that country	Number of researchers from that country	Percentage of the proposing network
Belgium	non ITC	7	8	11.43 %
Bosnia and Herzegovina	ITC	1	1	1.43 %
Bulgaria	ITC	1	1	1.43 %
Croatia	ITC	3	3	4.29 %
Cyprus	ITC	1	1	1.43 %
Czech Republic	ITC	6	6	8.57 %
Estonia	ITC	2	2	2.86 %
Finland	non ITC	1	1	1.43 %
France	non ITC	4	4	5.71 %
Germany	non ITC	6	6	8.57 %
Greece	non ITC	1	1	1.43 %
Ireland	non ITC	1	1	1.43 %
Israel	non ITC	1	1	1.43 %
Italy	non ITC	3	3	4.29 %
Latvia	ITC	1	1	1.43 %
Lithuania	ITC	2	3	4.29 %
Malta	ITC	1	1	1.43 %
Netherlands	non ITC	2	2	2.86 %
Norway	non ITC	2	2	2.86 %
Poland	ITC	1	1	1.43 %
Portugal	ITC	3	3	4.29 %
Romania	ITC	1	1	1.43 %
Serbia	ITC	1	1	1.43 %
Slovenia	ITC	3	3	4.29 %
Spain	non ITC	2	2	2.86 %
Sweden	non ITC	3	3	4.29 %
Turkey	ITC	1	1	1.43 %
United Kingdom	non ITC	6	7	10 %

Gender Distribution of Proposers

40.0% Males
60.0% Females

Average Number of years elapsed since PhD graduation of Proposers with a doctoral degree
10.5

Number of Early Career Investigators
27

Core Expertise of Proposers: Distribution by Sub-Field of Science

72.9% Languages and literature
10.0% Computer and Information Sciences
5.7% Educational sciences
2.9% Other social sciences
1.4% Electrical engineering, electronic engineering, Information engineering
2.8% Other
4.3% Unspecified

Institutional distribution of Network of Proposers

85.7% Higher Education & Associated Organisations
8.6% Business enterprise
2.9% Government/Intergovernmental Organisations except Higher Education
2.9% Private Non-Profit without market revenues, NGO

Higher Education & Associated Organisations:60

- Number by Field of Science of Department/Faculty of Affiliation
Languages and literature:36
Interdisciplinary:6
Educational sciences:5
Computer and Information Sciences:9
Other social sciences:2
Law:1
- Number by Type
Education Oriented:34
Research Oriented:26
- Number by Ownership
Fully or mostly public:55
50-50 Public and Private:1
Fully or mostly private:4

Business enterprise:6

- Number by Market sector of unit of affiliation
Information And Communication:2
Professional, Scientific And Technical Activities:2
Education:1
Other Service Activities:1
- Number by Type
Private enterprises:6
- Number by Ownership and International Status
Independent Enterprise:5
Enterprise owned by a foreign multinational group:1
- Number by Size
SME (EU Definition provided underneath after selection):6

Government/Intergovernmental Organisations except Higher Education:2

- Number by Level
Local government:1
Central and Federal Government:1
- Number by Type
Government department or government-run general public services:1
Other Public Non-Profit Institution:1

Private Non-Profit without market revenues, NGO:2

- Number by Type
Charity:1
Other:1
- Number by Level
National:1
International or European:1

COST Country Institutions(28) : Belgium , Bosnia and Herzegovina , Bulgaria , Croatia , Cyprus , Czech Republic , Estonia , Finland , France , Germany , Greece , Ireland , Israel , Italy , Latvia , Lithuania , Malta , Netherlands , Norway , Poland , Portugal , Romania , Serbia , Slovenia , Spain , Sweden , Turkey , United Kingdom

Near-Neighbour Country Institutions(0)

COST International Partners(0)

European Commission and EU Agencies(0)

European RTD Organisations(0)

International Organisations(0)

Network of Proposers - Details

Main Proposer's Details

Title:	Dr		
First Name:	Magali	Gender:	F
Last Name:	Paquot	Years from PhD:	12
Institution:	Université catholique de Louvain	Type of Institution:	Higher Education & Associated Organisations
Sub-field of Science of Department:	Languages and literature	Core Area of Expertise:	Languages and literature (Second language teaching and learning)

Secondary Proposers' Details

Belgium

Dr Jennifer Thewissen (Universiteit Antwerpen)

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 7

Prof Fanny Meunier (Université catholique de Louvain)

Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: F
Years from PhD: 19

Prof Thomas François (Université catholique de Louvain [FIAL])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 8

Ms Sophie Roekhaut (Altissia International SA)

Participating as Secondary Proposer
Core Expertise:
Gender: F
Years from PhD: No PhD

Dr Amandine Dumont (Université catholique de Louvain [Institut des Langues Vivantes])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 1

Dr Orphée De Clercq (Ghent University [LT3, Language and Translation Technology Team])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 4

Dr Frederik Cornillie (KU Leuven [ITEC & Research Unit Linguistics])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: M
Years from PhD: 5

Bosnia and Herzegovina

Dr Nihada Delibegovic Dzanic (University of Tuzla)

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 9

 **Bulgaria**

Dr Kiril Simov (Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences)

Participating as Secondary Proposer

Core Expertise: Computer and Information Sciences: Natural Language Processing

Gender: M

Years from PhD: 14

 **Croatia**

Dr Tihana Kras (University of Rijeka [Faculty of Humanities and Social Sciences, Department of English Language and Literature])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: F

Years from PhD: 11

Prof Nives Mikelic Preradovic (University of Zagreb, Faculty of Humanities and Social Sciences)

Participating as Secondary Proposer

Core Expertise: Other social sciences: Computational Lexicography

Gender: F

Years from PhD: 11

Dr Petra Bago (Faculty of Humanities and Social Sciences)

Participating as Secondary Proposer

Core Expertise: Other social sciences: Information Sciences

Gender: F

Years from PhD: 5

 **Cyprus**

Mr Christos Rodosthenous (Open University of Cyprus [Computational Cognition Lab])

Participating as Secondary Proposer

Core Expertise: Computer and Information Sciences: eLearning and Digital Content Management

Gender: M

Years from PhD: No PhD

 **Czech Republic**

Dr Alexandr Rosen (Charles University [Faculty of Arts, Institute of Theoretical and Computational Linguistics])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: M

Years from PhD: No PhD

Dr Maarten Janssen (Charles University - ÚFAL)

Participating as Secondary Proposer

Core Expertise: Computer and Information Sciences: Computational linguistics

Gender: M

Years from PhD: 17

Mr Jiri Hana (Genea Analytics)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: M

Years from PhD: 12

Mr Miloš Jakubiček (LEXICAL COMPUTING CZ S.R.O.)

Participating as Secondary Proposer

Core Expertise: Computer and Information Sciences: Artificial intelligence, intelligent systems, multi agent systems

Gender: M

Years from PhD: 2

Prof Pavel Smrz (Brno University of Technology)

Participating as Secondary Proposer

Core Expertise: Computer and Information Sciences: Machine learning algorithms

Gender: M

Years from PhD: 21

Dr Barbora Hladka (Charles University in Prague [Institute of Formal and Applied Linguistics])

Participating as Secondary Proposer

Core Expertise: Computer and Information Sciences: natural language processing

Gender: F

Years from PhD: 19



Estonia

Ms Jane Klavan (University of Tartu [Faculty of Arts and Humanities])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: F

Years from PhD: 7

Dr Jelena Kallas (Institute of the Estonian Language)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: F

Years from PhD: 6



Finland

Dr Therese Lindström Tiedemann (Helsinki University)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: F

Years from PhD: 15



France

Prof ROBIN Agnès (University of Montpellier [UMR 5815 "Dynamiques du droit"])

Participating as Secondary Proposer

Core Expertise: Law: Private law

Gender: F

Years from PhD: 18

Dr Céline Poudat (UNIVERSITY NICE COTE D'AZUR - Université Nice Côte d'Azur [Bases, Corpus, Langage])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology

Gender: M

Years from PhD: 13

Dr Fanny Rinck (Université Grenoble Alpes [Laboratoire Lidilem])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology

Gender: F

Years from PhD: 13

Dr Lydia-Mai Ho-Dac (CLLE-ERSS - University of Toulouse jean Jaurès)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: F

Years from PhD: 12

 **Germany**

Dr Valentin Werner (University of Bamberg [English Linguistics])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology

Gender: M

Years from PhD: No PhD

Dr Katrin Wisniewski (Leipzig University - Herder-Institut für Deutsch als Fremd- und Zweitsprache)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Second language teaching and learning

Gender: F

Years from PhD: 6

Prof Detmar Meurers (Universität Tübingen)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: M

Years from PhD: 20

Prof Robert Fuchs (University of Hamburg [Institute of English and American Studies])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology

Gender: M

Years from PhD: 5

Dr Steffen Brandt (Campus Business Box e.V. [opencampus.sh])

Participating as Secondary Proposer

Core Expertise: Educational sciences: Item Response Models and Machine Learning

Gender: M

Years from PhD: 3

Mr Christian Ebert (cabuu GmbH)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: M

Years from PhD: No PhD

 **Greece**

Mr Alexandros Nousias (National Center for Scientific Research - Demokritos [Institute of Informatics & Telecommunications])

Participating as Secondary Proposer

Core Expertise:

Gender: M

Years from PhD: No PhD

 **Ireland**

Dr Liam Murray (University of Limerick)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Computer-Assisted Language Learning

Gender: M

Years from PhD: 26

 **Israel**

Mr Ilan Kernerman (K Dictionaries)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology

Gender: M

Years from PhD: 6

 **Italy**

Mr egon stemle (Eurac Research [Institute for Applied Linguistics])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: M

Years from PhD: No PhD

Ms Luciana Forti (University for Foreigners of Perugia)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Second language teaching and learning

Gender: F

Years from PhD: 0

Prof Stefania Spina (Università per Stranieri di Perugia)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Second language teaching and learning

Gender: F

Years from PhD: No PhD

 **Latvia**

Dr Ilze Auzina (Institute of Mathematics and Computer Science, University of Latvia [Artificial Intelligence Laboratory])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: F

Years from PhD: 13

 **Lithuania**

Dr Jolita Horbačauskienė (Kaunas University of Technology)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology

Gender: F
Years from PhD: 8

Dr Ramunė Kasperavičienė (Kaunas University of Technology)

Participating as Secondary Proposer
Core Expertise: Languages and literature: Translation and interpretation
Gender: F
Years from PhD: 10

Prof Jurate Ruzaitė (Vytautas Magnus University [Centre of Multilingualism and Intercultural Communication])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 13

 **Malta**

Dr Claudia Borg (University of Malta)

Participating as Secondary Proposer
Core Expertise: Electrical engineering, electronic engineering, Information engineering: Human computer interaction and interface, visualization and natural language processing
Gender: F
Years from PhD: 3

 **Netherlands**

Dr Marije Michel (Groningen University - Faculty of Arts [Applied Linguistics])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: F
Years from PhD: 8

Dr Darja Fiser (CLARIN ERIC [Board of Directors])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 10

 **Norway**

Dr Ann-Kristin Helland Gujord (University of Bergen [Department of Linguistic, Literary and Aesthetic Studies])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: F
Years from PhD: 6

Ms Eli Moe (Kompetanse Norge [Avdeling for prøver])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Language testing, second language acquisition, language assessment, language of schooling
Gender: F
Years from PhD: No PhD

 **Poland**

Dr Agnieszka Lenko-Szymanska (University of Warsaw [Institute of Applied Linguistics])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Second language teaching and learning

Gender: F

Years from PhD: 22

Portugal

Dr Iria del Río (Center of Linguistics (CLUL) - University of Lisbon - University of Lisbon)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: F

Years from PhD: 5

Dr Xavier Anguera Miro (Elsa Corp. - Sucursal em Portugal - Elsa Corp. Sucursal em Portugal [Portugal division])

Participating as Secondary Proposer

Core Expertise: Computer and Information Sciences: Artificial intelligence, intelligent systems, multi agent systems

Gender: M

Years from PhD: 13

Dr Amalia Mendes (University of Lisbon [Faculty of Arts])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: F

Years from PhD: 18

Romania

Dr Madalina Chitez (West University of Timisoara (WUT) [Department of English])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: F

Years from PhD: 8

Serbia

Dr Maja Miličević Petrović (University of Belgrade)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Second language teaching and learning

Gender: F

Years from PhD: 11

Slovenia

Dr Monika Kavalir (University of Ljubljana [Department of English])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: F

Years from PhD: 5

Dr Gasper Cankar (Državni izpitni center (National Examinations Centre) [Research and development])

Participating as Secondary Proposer

Core Expertise: Psychology: Databases, data mining, data curation, computational modelling

Gender: M

Years from PhD: 13

Dr Nikola Ljubescic (Jožef Stefan Institute [Department of Knowledge Technologies])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Databases, data mining, data curation, computational modelling

Gender: M

Years from PhD: 10

 **Spain**

Dr Mikel Iruskieta (University of the Basque Country (UPV/EHU))

Participating as Secondary Proposer

Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics

Gender: M

Years from PhD: 5

Dr Maria Belen Diez-Bedmar (University of Jaen UJA - University of Jaen)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Second language teaching and learning

Gender: F

Years from PhD: 9

 **Sweden**

Dr Elena Volodina (University of Gothenburg - Department of Swedish)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Language Technology for Language Learning (ICALL)

Gender: F

Years from PhD: 21

Dr Tove Larsson (Södertörn University [School of Culture and Education])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Second language teaching and learning

Gender: F

Years from PhD: 3

Mr Niss Jonas Carlsson (Språkkraft)

Participating as Secondary Proposer

Core Expertise: Educational sciences: Education: training, pedagogy, didactics

Gender: M

Years from PhD: No PhD

 **Turkey**

Prof Çiler Hatipoğlu (MIDDLE EAST TECHNICAL UNIVERSITY)

Participating as Secondary Proposer

Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology

Gender: F

Years from PhD: 16

 **United Kingdom**

Dr Nadia Mifka-Profozic (University of York [Department of Education])

Participating as Secondary Proposer

Core Expertise: Languages and literature: Second language teaching and learning

Gender: F

Years from PhD: 7

Dr Cylcia Bolibaugh (University of York [Department of Education, Centre for Research in Language Learning and Use])

Participating as Secondary Proposer
Core Expertise: Educational sciences: Second language acquisition
Gender: F
Years from PhD: 5

Dr Norbert Vanek (University of York [Department of Education, Centre for Research in Language Learning and Use])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 6

Dr Akira Murakami (University of Birmingham [Department of English Language and Linguistics])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: M
Years from PhD: No PhD

Prof Florence Myles (University of Essex [Department of Language and Linguistics])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 28

Dr Mark Brenchley (Cambridge Assessment English - Cambridge Assessment)

Participating as Secondary Proposer
Core Expertise: Educational sciences: Corpus Linguistics; Language Assessment
Gender: M
Years from PhD: 4

Dr Andrew Caines (University of Cambridge)

Participating as Secondary Proposer
Core Expertise:
Gender: M
Years from PhD: No PhD