COST Action
Final Assessment Review

(07/03/2017 to 06/09/2021)

CA16105: European Network for Combining Language Learning with Crowdsourcing Techniques

This report is submitted by the Action Rapporteur in fulfilment of the requirements of the rules for COST Action Management, Monitoring and Final Assessment and is confidential to the COST Association and the Management Committee of the Action.

Summaries

Main aim/ objective

The enetCollect Action aims at setting into motion a Research and Innovation trend for combining Language Learning and Crowdsourcing in order to enhance the production of language learning material and language-related datasets.

The Action addressed this as described below

EnetCollect has successfully brought together an interdisciplinary community of stakeholders to explore the high-potential subject of the combination of language learning and crowdsourcing. EnetCollect members came from related and relevant domains, such as Language Learning, Natural Language Processing, Computer Science, E-Lexicography, Learner Corpora and Ethics. They successfully tackled exploratory tasks and implemented working prototypes, allowing to better understand the extent of the high-potential behind this unusual combination mixing a linguistically-oriented field with a technically-oriented one.

Among other things, the efforts of the enetCollect members allowed to better understand how familiar language teachers and learners were with the concept of crowdsourcing, how and what type of language learning material could be crowdsourced from them, what criteria would foster their participation, and what technical, ethical and legal aspects should be taken into account by any solution for crowdsourcing learning content. The efforts also allowed to better understand how interesting the subject could also be to create language-related datasets.

Overall, the efforts resulted in hundredths of participations and presentations to meetings, several dozens of scientific stays and intensive collaborations, several dozens of scientific publications by authors that had never collaborated before and in several new projects funded. All of this despite the major hindrance induced by the Covid-19 health crisis.

In completing such achievements, it has been gradually confirmed that the combination of Language Learning and Crowdsourcing is a subject for which it is worth establishing a dedicated community. In that perspective, enetCollect created a solid groundwork for this brand-new community to thrive beyond the completion of the COST Action itself.

The Rapporteur summarised the Action's major outcomes, impacts and successes as follows

It is obvious that EnetCollect was a successful Cost Action as its members and collaborators managed to develop a Research and Innovation trend for combining Language Learning and Crowdsourcing and to fulfill their aim of enhancement of the production of language learning material and language-related datasets.

The Cost Action facilitated the creation of an interdisciplinary community of researchers and stakeholders, from related and relevant domains, such as Language Learning, Natural Language Processing, Computer Science, E-Lexicography, Learner Corpora and Ethics, who managed to explore the high-potential subject of the combination of language learning and crowdsourcing, even during the current pandemic, which has considerable slowed down the progress of the projects and its research and dissemination activities. The exploratory tasks and working prototypes were successfully tackled based on the link between linguistically and technically oriented fields.

Furthermore, the enetCollect members worked together on such issues as language learning and teaching, teachers' and students' awareness about crowdsourcing, material and curricula development based on crowdsourcing approaches taking technical, ethical and legal aspects into consideration as well as creation of language-related datasets.

In general, the enetCollect triggered the development of cooperation and collaboration on the topic of Language Learning and Crowdsourcing at the EU level, which was fruitful in terms of publications, new joint research proposals, knowledge transfer, presentations and meetings. EnetCollect has become a new community of researchers, educators, practitioners, experts working together on theoretical research and its practical implementation and dissemination, which is going to have a long-term impact on the society even after the Cost Action completion.

Achievement of MoU objectives, deliverables and additional outputs/ achievements

MoU objectives

The Action reported the achievement of the following objectives and their dependence on the Action networking.

MoU objective	Level of achievement reported by Action	Dependence reported by Action	Dependence assessed by Rapporteur
Creating a theoretical framework for achieving a shared understanding of the challenge, ranging from very concrete aspects such as defining a common terminology, to higher-level aspects such as evaluating the potential and limits of related approaches.	51 - 75%	High	High
Gathering evaluation data for complementing the theoretical framework with factual information obtained from prototypical experiments and wide-ranging surveys in order to extend and root the theoretical framework and evaluate which aspects should be approached first.	51 - 75%	High	High
Disseminating the knowledge created and the achievements to any relevant audience.	76 - 100%	High	High
Creating a balanced interdisciplinary core community of active stakeholders.	76 - 100%	High	High
Establishing communication channels allowing to easily share information and reach relevant members or stakeholders.	76 - 100%	High	High
Fostering new funded initiatives complementing the Action.	76 - 100%	Low	Low
Creating an association continuing the work started in the Action after its conclusion.	26 - 50%	High	Medium

Dependence = dependence of the achievement (of each MoU objective) on the Action networking.

Rapporteur assessment of the achievement of MoU objectives that the Action reported as achieved (76-100%)

The Action did achieve all the above objectives that it reported were more than 75% achieved.

Action explanation regarding MoU objectives reported as not fully achieved (less than 76%)

The table below shows the Action's explanation and the Rapporteur's analysis thereof for any MoU objectives that the Action reported as not fully achieved.

MoU Objective that was reported as not fully achieved	Action's explanation	Rapporteur's analysis
Creating a theoretical framework for achieving a shared understanding of the challenge, ranging from very concrete aspects such as defining a common terminology, to higher-level aspects such as evaluating the	This MoU objective was achieved by less than 76% since the pandemic highly impeded the last 1.5 years of the Action. Creating a theoretical framework required to tackle various aspects of the topic in relation to the	The Action Chair's explanation is valid. Due to the current pandemic situation the MoU objective was achieved by less than 76%. The work of the Action was particularly affected during the last 1.5 years of the

potential and limits of related approaches.	five working groups and then bringing together the collected knowledge to formulate the overall theoretical framework of the subject. While results on the first part of the task were achieved on various ends, the second part of the work was hindered by the pandemic which made any large scale physical meetings impossible and virtual initiatives could not make up for it.	Action. The members of the network and of working groups were not able to meet in person and discuss the progress of their work, thus, in comparison to the first stages of the project, the last ones were not fully fulfilled as online meetings and virtual initiatives were not effective enough. Overall, EnetCollect has developed a theoretical framework for the combination of language learning with crowdsourcing techniques. The relevant data was gathered by five working groups. The members of the network have reached consensus regarding key concepts and topics, but due to COVID-19 it was difficult to continue with a deeper analysis, integration and full description of the framework. In general, the members of the network managed to complete literature gathering and review with respect to explicit and implicit crowdsourcing and their application to language learning and language-data creation. They also carried out large-scale surveys with teachers and with learners regarding language learning platforms in line with ethical considerations.
Gathering evaluation data for complementing the theoretical framework with factual information obtained from prototypical experiments and wide-ranging surveys in order to extend and root the theoretical framework and evaluate which aspects should be approached first.	This MoU objective was achieved by less than 76% since the pandemic highly impeded the last 1.5 years of the Action. Since the work on prototypical experiments was meant to build on the description of the theoretical framework, which had to be built first, it naturally started only after about the conclusion of the first year of enetCollect. Progress on the practical evaluation and prototyping took up speed just when the pandemic slowed down everything. While different task forces within enetCollect continued to create prototypes and run experiments, the amount of activities and achievements remained greatly below the expected, due to limited availability and engagement of the Action participants overall. Also, the challenging situation of distant teaching made it more difficult to find teachers and learners willing to participate in extracurricular activities and experiments.	The Action Chair's explanation is valid. During the first stage of the project the researchers were working on the description of the theoretical framework. This stage was not affected by the pandemic. The second stage of the project, in particular prototypical experiments which have to be based on the new theoretical framework, was impeded by the COVID-19. There was a great slow down in the progress experiments and activities that had been planned. Not all Action participants were available and engaged in the same was as prior to the health crisis. The shift to a distant teaching has affected the access to the participants (teachers and learners) and the data collection process. Although the amount of experimental work accomplished was lower as it had been expected and planned, still the working group members managed to carry prototypical experiments and hands-on activities, which helped to develop prototypical language learning applications, to reviews current regulations, ethical principles and conditions, to critically evaluate the existing language learning environments and prepare relevant deliverables. Although the Cost

		Action managed to organize only two Hackathon-like meetings, they are effective in terms of practical development work within enetCollect and the formation of sustainable task forces.
Creating an association continuing the work started in the Action after its conclusion.	This MoU objective was achieved by less than 50% since we decided that we will approach the creation of an association only after confirming that a substantial number of members has proven to continue working together on enetCollect-related thematics. Indeed, since the last 1.5 years of enetCollect were highly impeded by the pandemics, its momentum was noticeably impacted and the number of highly engaged members has decreased. Accordingly, the creation of an association did not seem the most adequate nor realistic approach to organize the efforts of this new community at this point. Instead the core group decided to move forward with applying to become a DARIAH working group to continue enetCollect work for the next 1-2 years and to reconsider the creation of an association after this transitory period.	The Action Chair's explanation is valid. The MoU objective was achieved by less than 50% as it was affected by the pandemic. The main reason was a decrease in member's active participation in the network in comparison to the situation before the COVID-19. Consequently, the creation of an association was not feasible. The core group took a decision to organize a DARIAH working group and to continue enetCollect work for the next 1-2 years, which can allow the creation of an association after this transitory period.

General Assessment of MoU objectives

The level of ambition of the MoU objectives was **High**
Overall, **the Action achieved all MoU Objectives**.

Deliverables

Delivery and level of dependence of deliverables reported by Action

Deliverable	Timing deliverable	Dependence reported by Action*
D.1.2 - Summary report on the state of the art of explicit crowdsourcing approaches	**Delivered**	**Medium**
D.1.3 - Summary report on the theoretical framework for producing material through explicit crowdsourcing	**Delivered**	**Medium**
D.1.4 - Design plans, implemented prototypes and related evaluation reports on explicit crowdsourcing for language learning material production	**Delivered**	**Medium**
D.2.2 - Summary report on the state of the art of implicit crowdsourcing approaches	**Delivered**	**Medium**
D.2.3 - Summary report on the theoretical framework for producing material through implicit crowdsourcing	**Delivered**	**Medium**
D.2.4 - Design plans, implemented prototypes and related evaluation reports on implicit crowdsourcing for language learning material production	**Delivered**	**Medium**
D.3.1 - Summary report on existing state of the art solutions for user-oriented design strategies	**Delivered**	**Medium**
D.3.2 - Summary report on the theoretical framework with regards to design strategies for functionalities, interface and usability	**Delivered**	**Medium**
D.3.3 - Design plans and evaluation reports on implemented prototypes	**Delivered**	**Medium**
D.4.1 Technology-oriented specifications to support explicit and implicit crowdsourcing for language learning material production (WG1 and WG2)	**Not foreseen**	**Medium**
D.4.2 Technology-oriented specifications to support user-oriented design strategies for a competitive solution (WG3)	**Not foreseen**	**Medium**
D.5.1 Ethical guidelines to build the grounds for the creation of templates and information sheets	**Delivered**	**Medium**
D.5.2 - Legal guidelines to build the grounds for the creation of templates and information sheets	**Delivered**	**Medium**
D.5.3 - Business guidelines to build the grounds for the creation of templates and relevant business model descriptions	**Not foreseen**	**Low**
D.7.1 - Guidelines for the management and implementation of the Dissemination Plan	**Delivered**	**High**
D.7.2 - Guidelines for the management and implementation of the Exploitation Plan	**Delivered**	**High**
D.7.4 - Action-internal lists of relevant stakeholder groups related to enetCollect's objectives and WGs	**Delivered**	**High**
D.7.5 - Communication means including project website and social media	**Delivered**	**High**
D.7.7 Regular press releases on enetCollect over the entire lifetime of the Action	**Delivered**	**Low**
D.7.8 Newsletters and news announcements about enetCollect over the entire lifetime of the Action	**Delivered**	**Medium**

D.7.10 Final report on the R&I targets, prototypical implementations and achievements of enetCollect in combining language learning with crowdsourcing techniques	**Delivered**	**High**
D.7.12 - First edition of a recurrent scientific event on combining language learning and crowdsourcing	**Not delivered, but foreseen within 2 years**	**Low**
D.7.13 - Preparation of the groundwork for creating an association to continue the work started in enetCollect	**Not delivered, but foreseen within 2 years**	**Low**

* Dependence reported by Action = the extent to which the delivery of the deliverable was dependent on the Action networking

Rapporteur analysis of level of delivery of deliverables

The level of delivery of the deliverables reported above is assessed as follows.

Overall, the level of achievement of the Deliverable reported by the Action is medium. It is important that nearly all of the deliverables have been achieved even though the whole work of the Cost Action has been affected by the current pandemic. The members of the network have prepared the summary reports on the state of the art of explicit/implicit crowdsourcing approaches, on the theoretical framework for producing material through explicit/implicit crowdsourcing, on existing state of the art solutions for user-oriented design strategies, on the theoretical framework with regards to design strategies for functionalities, interface and usability. Design plans and evaluation reports on implemented prototypes and related evaluation reports on explicit/implicit crowdsourcing for language learning material production have been delivered. Technology-oriented specifications to support explicit and implicit crowdsourcing for language learning material production and to support user-oriented design strategies for a competitive solution have not been foreseen. Both ethical and legal guidelines to build the grounds for the creation of templates and information sheets were delivered. Business guidelines to build the grounds for the creation of templates and relevant business model descriptions were not foreseen. Guidelines for the management and implementation of the Dissemination/Exploitation Plan have been delivered.

- **Analysis of deliverables reported by the Action as delivered**

  The deliverables that the Action reported as delivered are confirmed.

- **Analysis of deliverables reported by the Action as not delivered but delivery foreseen within 2 years**

Deliverable	Plans to ensure delivery within two years
D.7.12 - First edition of a recurrent scientific event on combining language learning and crowdsourcing	Regular scientific events on combining language learning and crowdsourcing will be organized in the form of workshops co-located with relevant conferences and separate crowdfest hackathons. The organization will be coordinated by a DARIAH working group, which we are currently preparing (see Dlv. 23) in order to keep the enetCollect community together after the end of the COST Action.
Deliverable	Plans to ensure delivery within two years
D.7.13 - Preparation of the groundwork for creating an association to continue the work started in enetCollect	Since the pandemics partly broke enetCollect's momentum and activities decreased, the creation of an association did not seem the most adequate approach to continue enetCollect's efforts after the COST Action, as it's community is unlikely to make strong commitments at this point. Instead, we opted to create a DARIAH working group

(https://www.dariah.eu/activities/working-groups-list/) to keep the enetCollect community together in a lighter fashion. The application for the DARIAH group has been submitted and should be approved by February 2022. If this is the case, the group should be launched within early 2022.

The plans described by the Action to ensure the delivery within two years are credible.

- **Analysis of deliverables reported by Action as not delivered and delivery not foreseen**

Deliverable	Explanation
D.4.1 Technology-oriented specifications to support explicit and implicit crowdsourcing for language learning material production (WG1 and WG2)	Task 4.1 aimed to provide support on technological challenges that enetCollect members of WG1/WG2 might face when developing approaches and prototypes combining Crowdsourcing and Language Learning. WG1 and WG2 were expected to formulate WG-transversal and mainly content-independent technological needs, which would be tackled by WG4. Things didn't go as expected for two reasons: (1) the rather exploratory WG1/WG2 research didn't yield clear technical requirements. (2) the pandemic slowed down activities in the project phase, where most of the concrete outputs/requirements should have been created. See https://www.enetcollect.net/ilias/goto.php?target=file_1196_download for details.

Deliverable	Explanation
D.4.2 Technology-oriented specifications to support user-oriented design strategies for a competitive solution (WG3)	Task 4.2 aimed to study technical solutions to support the creation of language learning platforms. It identified language learning software and learning management systems as the most relevant software and aimed to evaluate their strengths and weaknesses. We faced two problems: (1) no comprehensive open-source solutions were available for language learning, and commercial software partners were difficult to involve. (2) For the identified open-source solutions for learning management systems, specific support for language learning needs were missing (e.g. ASR). Finally, the pandemic hindered any further attempts. See https://www.enetcollect.net/ilias/goto.php?target=file_1197_download for details.

Deliverable	Explanation
D.5.3 - Business guidelines to build the grounds for the creation of templates and relevant business model descriptions	Despite it's continuous promotion throughout the Action, very few members could be attracted for the business aspect of enetCollect and accordingly very little work could be carried out. Indeed, the work remained limited to the outputs of one STSM on "Drivers of success of crowdsourcing platforms", a review of state-of-the-art literature on existing crowdsourcing platforms from the economic point of view (see report [1]) and two small task forces at the Crowdfests in 2019 and 2020[2]. [1] https://www.enetcollect.net/ilias/goto.php?target=file_416_download [2] https://enetcollect.eurac.edu/cost-tools-events/meetings/

The Action's explanation is analysed as follows The Action does not foresee the delivery of the flowing: Technology-oriented specifications to support explicit and implicit crowdsourcing for

language learning material production (WG1 and WG2); Technology-oriented specifications to support user-oriented design strategies for a competitive solution (WG3) and Business guidelines to build the grounds for the creation of templates and relevant business model descriptions. The Action has provided valid justification for this.

- **Analysis of the level of dependence on the Action networking of the achievement of the deliverables**

The dependence on the Action networking of the achievement of the deliverables reported by the Action is confirmed

General Assessment of deliverables

The level of ambition of the deliverables was **medium**
Overall, **the Action achieved all deliverables**

Additional outputs / achievements

Co-authored Action publications

The Action reported 34 publications on the topic of the Action, co-authored by at least two Action participants from two countries participating in the Action, and for which the Action networking was necessary. The full list of publications appears in Annex I.

Action networking was necessary for ALL of the listed publications

The:

- **Quality** of the Action's co-authored publications is **very good**.
  The publications produced by the members of the Cost Action are of high quality. The research articles have focused on the aims of the Cost Action, in particular the attempt of unlocking a crowdsourcing potential available for all languages, including less widely spoken languages, in order to create language resources and achieve a coverage of material for teaching the languages. They showed that the network has met its research and capacity-building goals by creating an international community of researchers worked on producing a comprehensive theoretical framework and running prototypical experiments to benefit a wide range of users and languages, while considering ethical, legal, and business issues. The authors have informed the readers about the objectives, expected impact and strategic organisation of the Cost Action Network that contribute to reaching its flexible and sustainable success goals. Some of the authors focused on such important topics as edutainment, gamification, computer assistance and a smart User Interface (UI), the use of crowdsourcing websites/games to learn foreign languages taking context specific variables into consideration, providing insights for experts, material developers and teacher trainers striving to create cross-culturally valid crowdsourcing platforms/games. However, there are more proceeding and local journals rather than peer-reviewed international journals.

- **Significance** of the Action's co-authored publications is **very good**.
  Overall, the publications of the Cost Action are of great significance. The authors of the articles discuss the importance of their research. Crowdsourcing not only opens new perspectives within the general concept of Computer Assisted Language Learning (CALL), but also raises questions about ethics, motivation, and fair contribution. Technology offers platforms such as Duolingo, Bussu, and Babbel for learning languages with active contribution of the learners. Such applications reach millions of users. Thus, there is a need for initiatives to explore the potential of crowdsourcing for language learning. According to the researchers, one of them is enetCollect CA16105 Combining Language Learning with Crowdsourcing Techniques, which is a European project within COST action. The aim of the Cost Action publications is to disseminate the project's ideas as well as present some results of the research done by the authors, members of the Cost Action. Although most of the papers are either proceedings or local journal papers rather than peer-reviewed international journal, the authors investigate very important issues that can contribute to the development of the field, for example, general frameworks for synchronous educational language games that simultaneously allow researchers to crowdsource learner answers in a controlled environment.

- **Relevance** to the Action of the Action's co-authored publications is **very good**.
  The publication produced by the members of the network are relevant to the aims, objective and scope of the Cost Action. The topics covered in publications are related to automated paraphrasing of informal into formal language, corpus linguistics, automatic solutions for the normalization of informal syntactic inadequacies found in the researched structures into standard structures typical of formal or professional writing through the application of very generic transformational grammars. EnetCollect: A New European Network for combining Language Learning with Crowdsourcing Techniques has produced a lot of publications are in Spanish. A lot of publications explain the activities and deliverable of the Cost Action itself, promotion of a research trend combining the well-established domain of Language Learning with recent and successful crowdsourcing approaches; fostering the language skills of all citizens regardless of their backgrounds by enhancing the production of language learning material using Crowdsourcing techniques. The authors of journal

articles described the research conducted within the network with respect to combination of crowdsourcing with IT technologies used in areas such as language learning and Natural Language Processing (NLP), bringing together interdisciplinary researchers to foster language learning of all European citizens from diverse sociodemographic, cultural, educational, and linguistic backgrounds.

- **Quantity** of the Action's co-authored publications is **very good**.
  Overall, 34 articles have been submitted for review via the submission system. It should be noted that most of the articles are published in local journals and proceedings rather than in peer-reviewed international journals. In general, the report shows that the Cost Action members were prolific in terms of publications that come from their collaboration in the Cost Action. This could be the indication of the Cost Action success and the impact on the society. It should be noted that publications are presented not only in English language but also in Spanish and other local languages. The papers are focused both on theoretical and practical issues, for example a cross-European survey on teachers and crowdsourcing, which examines how familiar language teachers are with the concept of crowdsourcing and addresses their attitude towards including crowdsourcing into language teaching activities. The researchers were able to collect data on (a) teachers' experience with organizing crowdsourcing activities for students/pupils, (b) the development of crowdsourced resources and materials as well as (c) teachers' motivation for participating in or employing crowdsourcing activities, in over 30 European countries.

Projects and proposals resulting from Action activities

The Action reported the following projects resulting from Action activities involving at least one Action participant, and for which the Action networking was necessary.

Title	Main proposer name	Funder
WIKOLLECT (resubmission)	Alberto BARRÓN CEDEÑO	H2020
CALLECTOR	Emmanuel Rayner	National
Multimodal Child Language Acquisition	Costanza Navarretta	National
Large-Scale Information Extraction and Gamification for Crowdsourced Language Learning	Pavel Smrz	National
From parallel corpora to multilingual exercises - Making use of large text collections and crowdsourcing techniques for innovative autonomous language learning applications	Johannes Graën	Other
LegiCrowd Onto	Alain Couillault	Other EU - NGI The Next Generation Internet initiative by the Digital Single Market of the European Commission.
Empirical foundations for digitally-supported development of writing skills	Špela Arhar Holdt	National

In addition the Action reported 5 proposals resulting from Action activities involving at least one Action participant, and for which the Action networking was necessary.

Relevance of the Action's proposals and/ or projects is **very good**

Quantity of the Action's proposals and/ or projects is **very good**

Action networking was necessary for ALL of the listed proposals / projects

Other Outputs / Achievements

The table below shows the other outputs / achievements and level of dependence on the Action networking reported by the Action and the Rapporteur's assessment thereof.

Other Output / Achievement reported by Action	Dependence reported by Action	Dependence assessed by Rapporteur
A prototypical architecture for implicitly crowdsourcing language-related data while generating vocabulary exercises was developed during the first Hackathon-like event we organized in January 2019. The work on the architecture was continued as a virtual collaboration of the taskforce throughout the year. Another session of the initiative was also organized at the following Hackathon-like event in early 2020. The work has resulted in multiple publications. The code of the project is open source and available on Gitlab. https://gitlab.com/crowdfest_task3 Related publications: Rodosthenous, C. T. (2019, May 21). Designing a Prototype Architecture for Crowdsourcing Language Resources [Poster presentation]. LDK 2019, 2nd Conference on Language, Data and Knowledge, Leipzig, Germany. Lyding, V., Rodosthenous, C. T., Sangati, F., ul Hassan, U., Nicolas, L., König, A., Horbacauskiene, J., & Katinskaia, A. (2019, April 8). v-trel: Vocabulary Trainer for Tracing Word Relations - An Implicit Crowdsourcing Approach. Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2019. International Conference Recent Advances in Natural Language Processing, RANLP 2019, Varna, Bulgaria. Rodosthenous, C., Lyding, V., Sangati, F., König, A., ul Hassan, U., Nicolas, L., Horbacauskiene, J., Katinskaia, A., & Aparaschivei, L. (2020). Using Crowdsourced Exercises for Vocabulary Training to Expand ConceptNet. Proceedings of LREC 2020, 12th Language	**High**	**High**

Resources and Evaluation Conference, 307--316. Nicolas, L., Aparaschivei, L., Lyding, V., Rodosthenous, C., Sangati, F., König, A., & Forascu, C. (2021). An Experiment on Implicitly Crowdsourcing Expert Knowledge about Romanian Synonyms from L1 Language Learners. Proceedings of 10th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2021), 1–14.		
A prototypical role-playing program which allows to implicitly crowdsource lexical data for non-standardized languages was created during the first Hackathon-like event we organized in January 2019. The role-playing game fosters the collaboration of the player with a person with good command of the non-standardized language (e.g. a grandparent) to answer the questions provided in the game. The work was continued within the PhD project of an enetCollect member. A second collaborative effort on the game was carried out in the second Hackathon-like event in early 2020. The work including a demo has been published. Millour, A., Araneta, M. G., Lazić Konjik, I., Raffone, A., Pilatte, Y.-A., & Fort, K. (2019, May 17). Katana and Grand Guru: a Game of the Lost Words (DEMO). Proceedings of the Ninth Language & Technology Conference. 9th Language \& Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC'19), Poznan, Poland.	**High**	**Medium**
A Zotero group has been created to foster the collaborative collection of references provided by members that are relevant to enetCollect. It was intensively used in relation to literature reviews during STSMs. It is currently populated with 203 references, and its use by all members will be fostered further. https://www.zotero.org/groups/2174884/_enetcollect_bibliography_ This Zotero group was also completed by crowdsourcing bibliographical information from	**High**	**High**

participants of the WG1 meeting celebrated in Gothenburg at the end of 2018 as a condition for participants to join the meeting. https://docs.google.com/spreadsheets/d/1oIbeMIyym97iC5nK4EbdfMoOERDInHx8rtubIW64w9k/edit#gid=0		
A dedicated channel was created on videolectures.net and video recordings of the presentations held at enetCollect meetings are being made available on the channel, including video and aligned slides. http://videolectures.net/enetCollect/ More than 50 videos of several enetCollect meetings are available online, and have more than 100 and often even between 300-400 views.	**High**	**High**
We created a number of documents related to the COST Action workflow itself (e.g. procedures), and not to the subject of the COST Action, which we believe could be of interest for other COST-related stakeholders. We created a rather strict and comprehensive STSM procedure issuing STSM calls every 2 months. http://enetcollect.eurac.edu/cost-tools-events/stsms/stsm-procedure/ http://enetcollect.eurac.edu/cost-tools-events/stsms/stsm-approval/ We introduced STSM webinars, informal sessions in which former STSM grantees shared their experiences with members interested in approaching an STSM. https://groups.google.com/g/enetcollect-all/c/3RN2hWoRFEA/m/eDrKmcyXBgAJ?hl=de We created templates for Social Media contributions related to STSMs which all STSM grantees needed to complete to inform members about their STSM activity: https://docs.google.com/forms/d/e/1FAIpQLSfCc-4HEE-jljEtCZy7Fe9-eZZUy1uYiyxtPY_7OpvrR20Ydw/viewform We created a set of logos and templates to harmonize communications and give enetCollect an identifiable identity that members can relate to. http://enetcollect.eurac.edu/logos_and_templates/ We created a dedicated page for funding opportunities that members	**High**	**High**

can pursue autonomously. http://enetcollect.eurac.edu/autonomous-funding-opportunities/ We also created guidelines about how to perform literature reviews: https://docs.google.com/document/d/128gDF-RCNIaNfF5WrJMScN_cJwD3UYOvSTWurN44X5s/edit		
Two explicit crowdsourcing experiments on ranking Multi-Word Expressions (MWE) by difficulty for the purpose of creating L2 teaching materials were carried out in preparation for a WG1 hand-on workshop. Accordingly, a list of English (verbal or adverbial) multi-word expressions was associated with levels of complexity, where complexity is treated in terms of CEFR (Common European Framework of Reference) levels of proficiency (COE, 2001). The list was created through the two explicit crowdsourcing experiments where participants of a WG1 meeting were treated as a (fake) prototypical crowd of language teachers due to their high proficiency in English and their overall linguistic skills derived from their language-related R&I profession. The aim of each crowdsourcing experiment was to collect a number of subjective ratings from crowdsourcers regarding the relative difficulty of these expressions from the perspective of a language learner. In other words, the goal was to rate expressions based on how difficult they are for a language learner to produce. Slides on the experiment and results: https://spraakbanken.gu.se/sites/spraakbanken.gu.se/files/eNetCollect_WG1_GOT_JC_Experiment.pdf https://spraakbanken.gu.se/sites/spraakbanken.gu.se/files/Clustering%20and%20visualizing2.pdf Experiment: Link to the adverbial MWE experiment: https://mnozicenje.cjvt.si/project/enetcollect_mwe_ranking_adverbs_ns/ Link to the verbal MWE experiment: h	**Medium**	**Medium**

ttps://mnozicenje.cjvt.si/project/enetcollect_mwe_ranking_verbs_ns/ Link to the guidelines: https://docs.google.com/document/d/1XicAizG0nAe3gPiT47X3GWqF12SMOSyohLX7RAzV3d4/edit		
The survey (enetCollect survey for Crowdsourcing practices, https://www.1ka.si/a/161288), had the MC members of the enetCollect action as a target group . The idea was to use the survey designed in Alice Millour's STSM to gather information about the crowdsourcing practices from the countries which are members of the Action. Overall information about 31 crowdsourcing resources was collected while the survey was active. The summary statistics of the survey are available in two different reports. Summary: https://drive.google.com/file/d/1ia-95syaBBJzbMwkzZT5JhD9ZDh-vret/view?usp=sharing Chart Data Analysis: https://drive.google.com/file/d/1n2NjoZ2lR2_zu2ShVmbXXUcaxjv4kumJ/view?usp=sharing	**High**	**High**
A survey was conducted among language teachers (N= 1129) including participants from over 30 European countries (over 35 members of the network were involved in the survey preparation and dissemination). In the survey, the following questions were addressed: (a) Are teachers familiar with the concept of crowdsourcing? (b) Are they including crowdsourcing in their teaching? (c) Are they themselves active in crowdsourcing activities? (d) If yes, what are these activities? (e) What is their attitude towards crowdsourcing in general? (f) What would motivate them to include crowdsourcing into their teaching? While the participating teachers were generally positively inclined towards the idea of crowdsourcing, their familiarity with actual possibilities to use crowdsourcing activities in the classroom was very limited. The participants reported the need for learning about crowdsourcing as a concept as well as the need for concrete ideas about which specific crowdsourcing activities are suitable for language teaching. Slides discussing the results => https://spraakbanken.gu.se/sites/spraakbanken.gu.se/files/enetCollect-Presentation-ARHAR-Teacher_Survey.pdf	**High**	**High**

The results of the survey have been published: Arhar Holdt, Š., Zviel-Girshin, R., Gajek, E., Durán-Muñoz, I., Bago, P., Fort, K., Hatipoglu, C., Kasperavičienė, R., Koeva, S., Lazić Konjik, I., Miloshevska, L., Ordulj, A., Rodosthenous, C., Volodina, E., Weber, T., & Zanasi, L. (2020). Language Teachers and Crowdsourcing: Insights from a Cross-European Survey. Journal of the Institute of Croatian Language and Linguistics // Časopis Instituta Za Hrvatski Jezik i Jezikoslovlje, 46(1), 1–28.		
A Zotero group and related library has been created to collect and annotate references related to the crowdsourcing of NLP datasets through any types of explicit and implicit crowdsourcing. It has been curated by a PhD student working on WG2 objectives and a taskforce for preparing a survey paper on crowdsourcing of NLP datasets. https://www.zotero.org/groups/2352007/_enetcollect_bibliography_-_wg2_review_group_ The library currently contains 334 items, which are processed and annotated in the collaborative effort of the abovementioned taskforce. The survey paper is in preparation and is expected to be submitted to a computational linguistics journal in early 2022.	**High**	**High**
A Synchronous Educational Language Game for Simultaneous Teaching and Crowdsourcing (Substituto) was developed during the second Hackathon-like event in February 2020. It proposes a general framework for synchronous educational language games that simultaneously allows researchers to crowdsource learner answers in a controlled environment by having teachers and students interacting in real-time. In the initial setup for demonstration and testing Substituto was used for	**High**	**High**

training VPC replacement in sentences, also known as phrasal verbs. The work has been published and the code of the prototype is available on gitlab: https://gitlab.com/substituto/nlp4call2020 Related publication: Grace Araneta, M., Eryigit, G., König, A., Lee, J.-U., Luís, A., Lyding, V., Nicolas, L., Rodosthenous, C., & Sangati, F. (2020). Substituto - A Synchronous Educational Language Game for Simultaneous Teaching and Crowdsourcing. Proceedings of the 9th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2020), 1–9.		
A gamified crowdsourcing approach for collecting language learning materials for idiomatic expressions has been explored in the second Hackathon-like event of enetCollect at early 2020. It resulted in Dodiom an asynchronous multiplayer game created on top of a messaging bot. The game is designed for native speakers who compete with each other while providing idiomatic and nonidiomatic usage examples and rating other players' entries. As opposed to classical crowdprocessing annotation efforts in the field, for the first time in the literature, a crowdcreating & crowdrating approach is implemented and tested for idiom corpora construction. A publication describing the efforts is published on arXiv.org and a journal publication is in press. Gülşen Eryiğit, Ali Şentaş, Johanna Monti (2021) Gamified Crowdsourcing for Idiom Corpora Construction https://arxiv.org/abs/2102.00881	**High**	**Medium**
Proceedings of the joint WG3&WG5 workshop held in Leiden, NL in October 2018 have been published with CEURS. http://ceur-ws.org/Vol-2390/ EnetCollectWG3&WG5 2018 - enetCollect WG3 & WG5 Meeting Supplementary Proceedings of the enetCollect WG3 & WG5 Meeting 2018	**High**	**High**

Leiden, The Netherlands, October 24-25, 2018. Edited by Katerina Zdravkova *, Karën Fort **, Branislav Bédi *** * Univeristy Ss Cyril and Methodius, Faculty of Computer Science and Engineering, Skopje, N. Macedonia ** Sorbonne University, School of Arts and Humanities, Paris, France *** University of Iceland, Faculty of Humanities, Reykjavik, Iceland		
During a task at 2019 Hackhaton-like Crowdfest, a shared effort was started to map a large number of existing language learning platforms that offer language learning exercises. A large number of references was obtained by running during the summer 2019 a survey that received 628 answers from participants of 68 different countries speaking 43 different mother tongues. The annotation and study of the references is still on going but a preliminary list of platforms was shared with the members. https://groups.google.com/g/enetcollect-all/c/GiXgrXZMXEw	**High**	**High**

The quality, quantity and dependence (on the Action networking) of the other outputs/ achievements was assessed as follows.
Overall, the Other Outputs and Achievements of the Cost Action are of high quality and quantity and have from medium to high dependency on the Action networking. The researchers managed to highlight the most significant outputs, which include relevant publications based on the joint, interdisciplinary research as part of the activities within the Network. The Cost Action members provided examples of the Cost Action events such as the first Hackathon-like event they organized in January 2019, which helped to develop a prototypical architecture for implicitly crowdsourcing language-related data while generating vocabulary exercises. The work on the architecture was continued as a virtual collaboration of the taskforce throughout the year. They also organized the following Hackathon-like event in early 2020, which resulted in multiple publications. A very important output is also A Zotero group that has been created to foster the collaborative collection of references provided by members that are relevant to enetCollect. It was intensively used in relation to literature reviews during STSMs. In addition, a dedicated channel was created on videolectures.net and video recordings of the presentations held at enetCollect meetings have been made available on the channel, including video and aligned slides.

Assessment of additional outputs and achievements (including co-authored publications and proposals/ projects)

The level of ambition of additional outputs and achievements was **medium**.
Overall, **the Action achieved > 4 valid Additional Outputs / Achievements**.

General Assessment

The Action's outputs and achievements are **very Good**.

Impacts

The Action reported the following impacts (the short- to long-term scientific, technological, and / or socioeconomic changes produced by a COST Action, directly or indirectly, intended or unintended) that have resulted, or might result, from the Action.

Description of the impact	Type of impact	Timing of impact
The world is confronted with an intensified need for effective multilingual and cross-lingual communication, both on a personal and professional level, induced by educational, professional, economic and geopolitical circumstances such as the Erasmus programs, market globalization, political conflicts or global crises. Already in 2012, a report on "Europeans and their Languages" stated that 21% of Europeans aged over 14 (84.4% of the European population) were learning a language. If we include learners under 14, we can estimate that the 2012 population of European language learners was composed of at least one hundred million learners (and is most likely much larger today). The multilingual reality is thus especially true for Europe, with its dense population with varied linguistic backgrounds. As such, efforts aiming at enhancing the overall language learning offer are highly desirable. At the same time, CALL contributes to distant learning, a subject of particular importance nowadays because of the COVID-19 pandemic. EnetCollect contributed to enhancing the language learning offer by creating a new R&I trend that has as its core aim to benefit language learning, as well as a heterogeneous R&I community exploring this trend from various perspectives (e.g. data perspective, language learning perspective, ethical perspective etc.). The novel nature of this community can already be attested by publications co-authored by authors who had in most cases not published together before. The recent, yet mature nature, of this community can be observed through the increasing number of members registering to the mailing lists and on the intranet (more than 200 email addresses registered on the general mailing list and 206 profiles registered on the intranet) as well by the amount of enetCollect-related publications (51 publications recorded so far) and the amount of project proposals submitted (12 proposals submitted), 7 of which were funded for an overall amount of around one million euros. While in practice, the results achieved by this community are more related to research than innovation and, as such, haven't yet led to the creation of much language learning material, we expect that its future achievements will in time reach the general public and impact the language learning offer.	• Scientific / Technological • Economic • Societal	Achieved

Validity, relevance and significance (in particular importance and timeliness) of the impact reported by the Action: It is obvious that EnetCollect has a high impact on society since it has a high scientific, technological and economic value. The aim of the Cost Action was to enhance the overall language learning in Europe and worldwide, especially in relation to distant learning during the current pandemic. There is no doubt that multilingual and multicultural Europe needs effective multilingual and cross-lingual communication, on a personal and professional level, induced by educational, professional, economic and geopolitical circumstances such as the Erasmus programs, market globalization, political conflicts or global crises. A new R&I trend has been created by the EnetCollect in order to enhance language learning. The Cost Action has triggered the creation of heterogeneous R&I community that can

explore this trend from various perspectives (e.g. data perspective, language learning perspective, ethical perspective etc.). The EnetCollect proved to be a collaborative network of researchers working together on research, data collection, data analysis and dissemination, which is reflected in joint publications (51 publications) and project proposals (12 proposals submitted with 7 funded). There is a growing number of members from different countries of Europe.

Language-related domains, such as the NLP (Natural Language Processing) community, are research domains of worldwide importance which have kept on growing in alignment with the fast unfolding of the digital age. They are also a cornerstone of recent modern AI efforts, as its technologies allow approaches to generate information from large amounts of text and to process the underlying knowledge. Even though the vast majority of related approaches rely on language-related datasets, perfect datasets are still nowadays a chimera, and these research fields have always been hindered by this major and unresponded demand. By having researched high-potential methods to enhance language-related datasets through crowdsourcing, enetCollect opened new paths for language-related dataset creation on a large scale given the size of the crowds potentially targetable. As such, it created the groundwork to address in a viable and sustainable fashion the aforementioned major and unresponded demand. Should the achievements continue to unfold after the completion of enetCollect, the impacts on language-related domains such as the NLP would induce a catalyst effect resulting from the removal of a major hindrance to their developments. As such, the long-term impacts could become vastly significant.	• Scientific / Technological	Achieved
Validity, relevance and significance (in particular importance and timeliness) of the impact reported by the Action: The work of the Cost Action is important in terms of validity, relevance and significance and its impact on society regarding scientific, technological, economic and societal development. Among the topics that have been investigated by the members of the Cost Action NLP (Natural Language Processing), corpus linguistics, language learning and teaching. The Cost Action had its aim to research high-potential methods and to enhance language-related datasets through crowdsourcing and its members managed to achieve this aim through collaboration and joint research. It has the potential to continue its long-term impacts even after its completion		
Within the R&I field of computer-assisted language learning (CALL), the automatic generation of exercise content from language-related datasets is almost non-existent, despite the fact that some of these datasets encode the knowledge that learners are often tested on (e.g., lexical knowledge). This absence is probably due to differences in expectations with respect to linguistic accuracy: learning materials are usually close to perfect, whereas NLP resources rarely are. Generating content from imperfect datasets poses a challenge in terms of its suitability for learning. By exploring approaches that combine the generation of language learning exercises from language-related datasets while crowdsourcing linguistics knowledge from the answers to the exercises and use it in order to improve the datasets themselves, enetCollect members have provided a clear solution to unlock the stalemate on this research question of noticeable potential for CALL.	• Scientific / Technological	Achieved
Validity, relevance and significance (in particular importance and timeliness) of the impact reported by the Action: Based on the report, it is evident that the Cost Action has its impact on the society and scientific/technological advancement. The researchers, members of the Cost Action Network managed to fill in the existing gap with respect to the R&I field of computer-assisted language learning (CALL), the automatic generation of exercise content from language-related datasets. This is a significant advancement in the field of corpus linguistics, CALL, language teaching and languages learning. The enetCollect members explored approaches that combine the generation of language learning exercises from language-related datasets while crowdsourcing linguistics knowledge from the answers to the exercises and use it in order to improve the datasets themselves.		

The extent to which the Action has advanced the careers, skills and networks of researchers including ECIs (as described by the Action) is very good.

General assessment of impacts

The Action's impacts are best described as follows.
Multiple highly significant impacts are reasonably foreseen for the future OR one highly or moderately significant impact is already observed [Very Good]

Dissemination and exploitation of Action results (other than co-authored Action publications listed previously)

Dissemination meetings funded by the Action

The following Dissemination meeting(s) funded by the Action added value for the Action:

- Dissemination meeting 2, 22-08-2018 - 25-08-2018, Finland
- Dissemination meeting 1, 07-05-2018 - 12-05-2018, Japan
- Dissemination meeting 1, 30-05-2019 - 31-05-2019, Poland
- Dissemination meeting 2, 28-08-2019 - 31-08-2019, Belgium
- Dissemination meeting 2, 20-08-2020 - 21-08-2020, Denmark

Action website

http://enetcollect.eurac.edu/

The:

- openness and user-friendliness of the Action website are very good
- content of the Action website (programmes and minutes of all events present, all outputs/ deliverables accessible from website) is very good

The Action website was an effective means of disseminating the Action.

Other dissemination activities

The following other dissemination activities reported by the Action were effective and added value

Item/activity	Invited talk at the 7th DaFWEBCON - 'Webkonferenz für Deutschlehrende' (Online Conference for Teachers of German) Title: 'Sprachenlernen und Crowdsourcing – ein innovatives Projekt.' (Language Learning and Crowdsourcing, an innovative project) Presented by: Verena Lyding & Lionel Nicolas Date: March 2018
Target Audience	Teachers of German as second or foreign language participating on the 7th DaFWEBCON on 'Deutsch – global vernetzt – lokal aktiv!' (German - globally connected - locally active)
Outcome of the activity	Communication of enetCollect's ambition to teachers of German as second or foreign language. Collection of feedback and expressions of interest for being contacted for future experiments with teachers from conference participants (German teachers).
Hyperlink	https://dafwebkon.com/Veranstaltung/lyding-sprachenlernen-und-crowdsourcing-ein-innovatives-projekt/

Item/activity	Invited talk during STSM at University of Iceland, Faculty of Languages. Title: 'Crowdsourcing in language learning and teaching: "small" languages' achievements' Presented by: Katerina Zourou Date: March 2018
Target Audience	Interested members of the Faculty of Languages at the University of Iceland
Outcome of the activity	Communication of research topics of the STSM candidate, raising awareness for the enetCollect COST Action.
Hyperlink	https://www.slideshare.net/Web2Learn_eu/crowdsourcing-in-language-learning-and-teaching-small-languages-achievements

Item/activity	Invited talk during STSM presented at the project "Industria 4.0", Corso di Letteratura italiana;

	Corso de Filosofia del diritto, Piacenza. Cite Title: Crowdsource Natural Language Processing Presented by: Alain Couillault Date: May 2018
Target Audience	Participants of the 'Corso di Letteratura italiana' and the 'Corso de Filosofia del diritto' at the University of Piacenza
Outcome of the activity	Communication of project activities, raising awareness on the enetCollect COST Action
Hyperlink	https://www.enetcollect.net/ilias/goto.php?target=file_530_download

Item/activity	SPECIAL: CROWDSOURCING UND EINE REVOLUTION IM SPRACHENLERNEN at the News Section on Website of Eurac Research Authors: Verena Lyding & Lionel Nicolas Date: March 2017
Target Audience	Interested public, in particular stakeholders of Eurac Research in South Tyrol, Italy
Outcome of the activity	Increased visibility of enetCollect's objectives, raising awareness of its potential for language learning
Hyperlink	http://www.eurac.edu/en/news/news/pages/newsdetails.aspx?entryid=122107

Item/activity	Two short radio transmissions at 'Morgenmagazin „Thema Wissen"' (Morning show on 'Knowledge/Science'), RAI Südtirol Topics: 'Crowdsourcing and Language Learning' and 'Computer-based Language Learning' Invited interview partner: Verena Lyding Date: June 2017
Target Audience	General public in South Tyrol, Italy, interested in transmissions on knowledge and science
Outcome of the activity	Communication of the project objectives, raised awareness on the crowdsourcing potential for language learning.
Hyperlink	http://www.raibz.rai.it/index.php

Item/activity	Presentation at mini-workshop collocations and sentence selection on the creation of language learning materials and automatic sentence selection for learner reference resources at Språkbanken, University of Gothenburg during an STSM. Title: "Semi-automated lexicography in Slovenia: collocations dictionary and beyond" Presented by: Iztok Kosem Date: May 2017
Target Audience	Researchers from the research group at Språkbanken, University of Gothenburg, as well as external research fellows.
Outcome of the activity	Communication and discussion of automatic data extraction methods used in Slovene lexicography and language learning; and presentation of plans for implementing crowdsourcing techniques into the workflow
Hyperlink	https://sweclarin.se/swe/mini-workshop-collocations-and-sentence-selection

Item/activity	Invited presentation at the research group for Second Language Studies at the University of Iceland during an STSM. Title: "The scope of language intuition: some insights from crowdsourcing" Presented by: Karën Fort Date: November 2017
Target Audience	Member of the research group for Second Language Studies at the University of Iceland, http://rim.hi.is
Outcome of the activity	Communication of relevant crowdsourcing work of the STSM applicant to researchers at the host institution.
Hyperlink	https://enetcollect.net/ilias/goto.php?target=file_379_download

Item/activity	Presentation for researchers from the tagged Icelandic corpus team (http://www.malfong.is/index.php?lang=en&pg=mim) Topic: Demonstration of ZombiLingo (dependency syntax), Bisame (POS tagging) andRigor Mortis (multi-word expressions) Presented by: Karën Fort Date: November 2017

Target Audience	Researchers from the tagged Icelandic corpus team (http://www.malfong.is/index.php?lang=en&pg=mim) interested in crowdsourcing examples.
Outcome of the activity	Communication about three existing crowdsourcing tools (GWAPs) for language resources. Participants to the presentation showed some interest and could be willing to collaborate on a Game with a purpose for Icelandic multi-word expressions or to adapt ZombiLingo.
Hyperlink	https://enetcollect.net/ilias/goto.php?target=file_379_download

Item/activity	Presentation at the Economics Department and Business School of the University of Iceland. Title: "Contractual governance and performance Presented by: Nina Gorovaia Date: January 2018
Target Audience	Researchers and members of the Economics Department and Business School of the University of Iceland.
Outcome of the activity	Communication of research of the STSM grantee for networking. Very useful contacts have been made that might result in future collaboration and consulting opportunities for the public sector.
Hyperlink	https://www.enetcollect.net/ilias/goto.php?target=file_416_download

Item/activity	Presentation to the members of Språkbanken, the Swedish Language Bank, University of Gothenburg (SB research meeting series) Topic: Previous work on exercise generation Presented by: Itziar Aldabe Date: March 2018
Target Audience	Researchers and members of Språkbanken, University of Gothenburg, interested in exercise generation.
Outcome of the activity	Communication of research work by the STSM grantee relevant for WG1.
Hyperlink	https://www.enetcollect.net/ilias/goto.php?target=file_518_download

Item/activity	Guest talk at the University of Ljubljana and Trojina Institute Title: "Towards a research infrastructure for Second Language Acquisition and teaching: case of L2 Swedish" Presented by: Elena Volodina Date: June 2018
Target Audience	A good portion of audience represented the highly needed group of L2 researchers, teachers and assessors.
Outcome of the activity	Communication of research activities of the STSM grantee to a relevant target group; networking and creation of opportunities for involvement into WG1.
Hyperlink	https://www.enetcollect.net/ilias/goto.php?target=file_544_download

Item/activity	Presentation at meeting with DALI team (Disagreements and Language Interpretation) at Queen Mary University at London (QMUL academic leader -Massimo Poesio). Topic: Presentation of enetCollect, WG1-5, Smart User Model and Linguistics Apps Development for Second Language Learning Presented by: Rina Zviel-Girshin Date: September 2018
Target Audience	Members of the DALI team (Disagreements and Language Interpretation) at Queen Mary University at London.
Outcome of the activity	Communication of the enetCollect Action and ongoing research of the STSM grantee.
Hyperlink	https://www.enetcollect.net/ilias/goto.php?target=file_658_download

Item/activity	Post-meeting proceedings for the WG3/WG5 meetings celebrated in Leiden at the end of 2019.
Target Audience	The target audience is composed of all stakeholder groups involved in enetCollect, especially the ones related to WG3 and WG5.

Outcome of the activity	The proceedings should be published during the summer 2019 and are thus not available yet. Nonetheless, a call for paper was launched, 13 submissions were made and reviewed by two peers. 12 papers were selected for publication.
Hyperlink	https://groups.google.com/d/msg/enetcollect-all/tJH4p74OC6w/hlrE95aTBQAJ

Item/activity	Presentation at IXA seminar series, University of the Basque Country. Topic: Towards a lightweight solution for less-resource languages Presented by: Alice Millour Date: February 2018
Target Audience	Researchers and members of the IXA NLP research group, University of the Basque Country.
Outcome of the activity	Communication of research work by the STSM grantee relevant for WG2.
Hyperlink	http://www.enetcollect.net/ilias/goto.php?target=file_802_download&client_id=enetcollect

Item/activity	Professional video recording of presentations held at enetCollect meetings and creation of a channel on videolectures.net to publish and archive the videos together with slides.
Target Audience	Action members as well as any interested stakeholders of the Action.
Outcome of the activity	Dissemination of the presentations held at enetCollect meetings to a wider audience. Formation of interested stakeholders on enetCollect related topics. Currently, 30 videos are available online and ca. 40 more videos are in processing.
Hyperlink	http://videolectures.net/enetCollect/

Item/activity	Creation of a Zotero group (reference management software) for the collaborative collection of references relevant to enetCollect's domains of interest.
Target Audience	EnetCollect members as well as any stakeholders of the Action; researchers and students of the related research domains.
Outcome of the activity	Centralized and collaboratively curated collection of relevant related work on the R&I trend on Language Learning and Crowdsourcing.
Hyperlink	https://www.zotero.org/groups/2174884/_enetcollect_bibliography_

Item/activity	Several announcements about the enetCollect Action as well as multiple transfer-oriented presentations and articles were published by members on their local/institutional webpages etc.
Target Audience	Local research communities of members, interested public, any visitors of members' webpages.
Outcome of the activity	Information about enetCollect, its objectives, community and preliminary outputs is communicated to a wider (also non-scientific) audience.
Hyperlink	http://enetcollect.eurac.edu/results/general_dissemination/

Item/activity	Presentation at the 4th Conference for Blended Learning, Zayed University, Dubai. Title: European Network for Combining Language Learning and Crowdsourcing Techniques – a New Perspective in Blended Learning. Presented by: Elżbieta Gajek Date: April 25th-27th, 2019
Target Audience	Stakeholders of the International Association for Blended Learning in Partnership with the Center for Educational Innovation, Zayed University, Dubai.
Outcome of the activity	Promoting the enetCollect Action to professionals and practitioners related related to blended learning.
Hyperlink	https://iabl.wildapricot.org/resources/Documents/WCBL_Proceedings.pdf

Exploitation activities

The following activities to ensure exploitation (use, in particular in a commercial context) of the Action's achievements reported by the Action were effective and added value

Item/activity	WG 2 together with members of the Core Group organized an Hackathon-like initiative in January 2019 to kick-start the implementation or theoretical elaboration of approaches/experiments/applications of relevance for the different WGs. The Core Group intends to repeat this successful event in 2020.
Target Audience	Any members of enetCollect interested in making a concrete and practical contribution to a subject of shared interest among various members, possibly from different WGs.
Outcome of the activity	During two days, more than 35 members worked in subgroups targeting 6 tasks related to different WGs. A number of tangible outputs resulted from such efforts, including various prototypes and publications.

Item/activity	During the Hackathon-like event organized in January 2019, one of the 6 tasks on which the members worked on was related to "Business models in language learning platforms" and aimed at exploring the business opportunities that enetCollect could create over time by first studying as a reference point the business models of existing platform.
Target Audience	Any members of enetCollect with a business-oriented mindset and interested in creating commercial solutions from the developments made within enetCollect.
Outcome of the activity	The preliminary results achieved were presented during the 3rd Annual Meeting in Lisbon and an STSM proposal will be submitted to extend this work further.

Item/activity	WG 2 together with members of the Core Group organized a second Hackathon-like initiative in February 2020 to foster the implementation or theoretical elaboration of practical approaches/experiments/applications of relevance for the different WGs.
Target Audience	Any members of enetCollect interested in making a concrete and practical contribution to a subject of shared interest among various members, possibly from different WGs.
Outcome of the activity	During two days, 38 members worked in subgroups targeting 6 tasks related to different WGs. A number of tangible outputs resulted from the efforts, including various prototypes and publications.

Item/activity	During the second Hackathon-like event organized in February 2020, one of the 6 tasks on which the members worked on was related to "Configurational perspective for the study of business models in crowdsourcing platforms" and aimed at exploring some of the business opportunities that enetCollect could create over time by first studying as a reference point the business models of existing crowdsourcing platform. It followed up on an effort already started in the previous 2019 event which focussed on language learning platforms.
Target Audience	Any members of enetCollect with a business-oriented mindset and interested in creating commercial solutions from the developments made within enetCollect.
Outcome of the activity	None to report (so far). The shared effort stoppped because of the covid crisis and hasn't restarted yet.

Item/activity	Several major language learning platforms were contacted and invited to join the network. Online discussions were held with employees of three major platforms: Babbel, Buolingo and Busuu.
Target Audience	Involving such major players could have impacted both their users and enetCollect members with whom they would start a collaboration.

Outcome of the activity	The head of education in Busuu, together with two junior colleagues, took part in the 2019 Annual meeting and showed interest in collaborating with enetCollect members. A researcher at Duolingo was planned as keynote in the 2020 Annual meeting that was cancelled because of the covid crisis.

Assessment of Action dissemination and exploitation

The effectiveness of the Action's dissemination and exploitation approach (other than co-authored publications) is assessed as follows:

It is obvious that the Cost Action has secured efficient methods of dissemination and exploitation of the results of their collaboration and research activities. There were dedicated members of the Cost Action who were responsible for monitoring different dissemination activities, they had specific task in order to foster various initiatives related to exploitation and dissemination of valuable data, information, knowledge transfer. It is evident from the report that members of the Cost Action were encouraged to work jointly in order to write new research proposals, many of which were successful and were funded. Members of the enetCollect created several channels in order to promote dissemination, for example ResearchGate and Twitter accounts.

Assessment of Action dissemination and exploitation activities:

All Action activities focusing on dissemination of Action results were effective [Very Good]
All Action activities focusing on exploitation of Action results were effective [Very Good]

Action Success(es)

The following table shows the success(es) reported by the Action and the Action Rapporteur's comment.

Success reported by Action	Action Rapporteur comment
EnetCollect has established and practiced Hackathon-like meetings called "Crowdfests". Unlike Hackhatons that are generally focused on completing coding- or computer-related practical tasks, the Crowdfests aimed at kickstarting collaborative efforts on any relevant tasks, ranging from theoretical to practical ones. These practical working sessions over 2-3 days have proven to be very powerful networking instruments for defining interdisciplinary research objectives and achieving first results both on theoretical and practical levels (e.g. preparing surveys, implementing prototypes). The Crowdfests incentivized several collaboration initiatives among Action members from different domains, which in most cases evolved into enduring collaborations and led to increasingly maturing outputs and related scientific publications.	The Success described above is valid as it builds capacity in in bridging separate fields of science and technology and in valorising and implementing advances and applications in science and technology. The report provides evidence of organization various events such as Hackathon-like meetings called "Crowdfests", which aimed at kickstarting collaborative efforts on any relevant tasks, ranging from theoretical to practical ones. It is obvious that such events and related activities were useful in promoting networking and collaboration, interdisciplinary research of practical and theoretical significance, which led to scientific publications and outputs and further research cooperation.
EnetCollect has been particularly successful in bringing together an interdisciplinary community of stakeholders from many related and relevant domains, such as Natural Language Processing, Computer Science, Language Learning, lexicography, learner copora and Ethics. Besides having involved a large number of more than 200 researchers and practitioners, enetCollect has managed to create first links between crowdsourcing and language learning professionals, and to raise specific awareness about the potential of crowdsourcing for language learning. By accomplishing such results, enetCollect created a solid groundwork for this new research and innovation to develop and thrive further.	The Success described above is valid and it is related to the capacity in bridging separate fields of science and technology and to the capacity in a new or emerging field of science and technology. It can be judged from the report that the success of the EnetCollect was related to the ability of the Cost Action members to collaborate with an interdisciplinary community of stakeholders from many related and relevant domains, such as Natural Language Processing, Computer Science, Language Learning, lexicography, learner corpora and Ethics. More than 200 researchers, professionals and practitioners have been involved in the network and they together managed to specific awareness about the potential of crowdsourcing for language learning. Without no doubt, enetCollect has created the groundwork and the basis for further collaboration, research and innovation even after its completion.

Other matters

Added value of extension

The validity of the Action's description of the added value of the extension is as follows: The added value of the extension as described by the Chair is valid. It important that despite the current pandemic, which prevented the smooth progress of the Cost Action and its activities, the members of the Cost Action still managed to continue with their practical and experimental research, relevant activities, online meetings, webinars, STSMs and Virtual Mobility grants.

Difficulties in implementing the Action

The Action Rapporteur made the following observations regarding difficulties in implementing the Action:

No difficulties in implementation of the Action have not been observed apart from the negative impact due to COVID-19.

Suggestions for improvements to COST framework / procedures

The Action Rapporteur made the following suggestions for changes to the COST framework:

N/A

Emerging topics / developments in the field of the Action

The Action reported the following emerging topics / developments in the field of the Action.

- We have no emerging topics or future developments to report on besides the ones originally targeted by enetCollect itself.

The Action Rapporteur made the following comment on the emerging topics / developments in the field reported by the Action.

There are no emerging topics or future developments to report on besides the ones originally targeted bu enetCollect itself.

Action Rapporteur

This Final Assessment Report was submitted on 2022-01-22 by:
Dr Sviatlana Karpava
University of Cyprus
Cyprus

Annex 1: List of publications

The Action reported 34 publications on the topic of the Action, co-authored by at least two Action participants from two countries participating in the Action, and for which the Action networking was necessary.

Co-authored Action publications - peer-reviewed

Title	Parafraseamento Automático de Registo Informal em Registo Formal na Língua Portuguesa
Authors	Anabela Marques Barreiro; Ida Rebelo-Arnold; Jorge Baptista; Cristina Mota; Isabel Garcez
DOI	doi:10.21814/lm.10.2.282
Type	Journal article
Published in	Linguamática
Published by	University of Minho
ISSN	1647-0818
Link	http://linguamatica.com/index.php/linguamatica/article/download/282/447

Title	enetCollect: A New European Network for combining Language Learning with Crowdsourcing Techniques
Authors	Rodrigo Agerri; Montse Maritxalar; Verena Lyding; Lionel Nicolas
DOI	doi:10.26342/2018-61-25
Type	Journal article
Published in	Procesamiento del Lenguaje Natural
Published by	Sociedad Española para el Procesamiento del Lenguaje Natural
ISSN	1989-7553

Title	Introducing the European NETwork for COmbining Language LEarning and Crowdsourcing Techniques (enetCollect)
Authors	Verena Lyding; Lionel Nicolas; Branislav Bédi; Karën Fort
DOI	doi:10.14705/rpnet.2018.26.833
Type	Chapter
Published in	Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018
Published by	Research-publishing.net
Link	https://research-publishing.net/publication/chapters/978-2-490057-22-1/833.pdf

Nicolas, Lionel, Verena Lyding, Luisa Bentivogli, Federico Sangati, Johanna Monti, Irene Russo, Roberto Gretter, and Daniele Falavigna. "EnetCollect in Italy." In *Proceedings of the Fifth Italian Conference on Computational Linguistics*. Torino, 2018.

Zviel-Girshin, Rina, Branislav Bedi, and Nathan Rosenberg. "EnetCollect User Model for Intelligent Gamification Crowdsourcing Language Learning IS." In *Proceedings of the 12th Israel Association for Information Systems (ILAIS) Conference*. Ashdod, 2018.

Zviel-Girshin, Rina, Nathan Rosenberg, and Christos Rodosthenous. "SUSTAIN-ALL Language Learning." In *Proceedings of the 12th Israel Association for Information Systems (ILAIS) Conference*. Ashdod, 2018.

Sangati, Federico, Ekaterina Abramova, and Johanna Monti. "DialettiBot: A Telegram Botfor Crowdsourcing Recordings of Italian Dialects." In Proceedings of the Fifth Italian Conference on Computational Linguistics. Torino, 2018.

Alice Millour, Marianne Grace Araneta, Ivana Lazić Konjik, Annalisa Raffone, Yann-Alan Pilatte and Karën Fort, "*Katana and Grand Guru: a Game of the Lost Words (DEMO)*" in Proceedings of the ninth Language & Technology Conference, Poznan, Poland, 2019.

Designing a Prototype Architecture for Crowdsourcing Language Resources, Christos Rodosthenous and Verena Lyding and Alexander König and Jolita Horbacauskiene and Anisia Katinskaia and Umair ul Hassan and Nicos Isaak and Federico Sangati and Lionel Nicolas, in Proceedings of the 2nd Language, Data and Knowledge (LDK), 2019.

Title	Crowdsourcing for widening participation and learning opportunities: a view from language learners' window
Authors	Çiler Hatipoğlu; Elżbieta Gajek; Lina Miloshevska; Nihada Delibegović Džanić
DOI	doi:10.14705/rpnet.2020.48.1169
Type	Book chapter
Published in	CALL for widening participation: short papers from EUROCALL 2020
Published by	Research-publishing.net
Link	https://research-publishing.net/publication/chapters/978-2-490057-81-8/1169.pdf

Title	Crowdsourcing in language learning as a continuation of CALL in varied technological, social, and ethical contexts
Author	Elżbieta Gajek
DOI	doi:10.14705/rpnet.2020.48.1168
Type	Book chapter
Published in	CALL for widening participation: short papers from EUROCALL 2020
Published by	Research-publishing.net
Link	https://research-publishing.net/publication/chapters/978-2-490057-81-8/1168.pdf

Title	Ethical issues of crowdsourcing in education
Author	Katerina Zdravkova
DOI	doi:10.1016/j.jrt.2020.100004
Type	Journal article
Published in	Journal of Responsible Technology
Published by	Elsevier BV
ISSN	2666-6596
Links	https://api.elsevier.com/content/article/PII:S2666659620300044?httpAccept=text/xml; https://api.elsevier.com/content/article/PII:S2666659620300044?httpAccept=text/plain

Title	Substituto - A Synchronous Educational Language Game for Simultaneous Teaching and Crowdsourcing
Authors	Marianne Grace Araneta; Gülsen Eryigit; Alexander König; Ji-Ung Lee; Ana Luís; Verena Lyding; Lionel Nicolas; Christos Rodosthenous; Federico Sangati
DOI	doi:10.3384/ecp201759
Type	Proceedings article
Published in	Proceedings of the 9th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2020)
Published by	Linköping University Electronic Press

Title	Language Teachers and Crowdsourcing
Authors	Špela Arhar Holdt; Lorenzo Zanasi; Tassja Weber; Elena Volodina; Christos Rodosthenous; Antonia Ordulj; Lina Miloshevska; Ivana Lazić Konjik; Svetla Koeva; Ramunė Kasperavičienė; Ciler Hatipoglu; Karën Fort; Petra Bago; Isabel Durán-Muñoz; Elżbieta Gajek; Rina Zviel-Girshin
DOI	doi:10.31724/rihjj.46.1.1
Type	Journal article
Published in	Rasprave Instituta za hrvatski jezik i jezikoslovlje
Published by	Institute of Croatian Language and Linguistics
ISSNs	1849-0379; 1331-6745
Subjects	Linguistics and Language; Language and Linguistics
Link	https://hrcak.srce.hr/file/353213

Title	LARA PORTAL: A TOOL FOR TEACHERS TO DEVELOP INTERACTIVE TEXT CONTENT, AN ENVIRONMENT FOR STUDENTS TO IMPROVE READING SKILL
Author	Hanieh Habibi
DOI	doi:10.21125/iceri.2019.1954
Type	Proceedings article
Published in	ICERI2019 Proceedings
Published by	IATED

Title	VEGETARIAN VAMPIRES: WHY THE CALL TECHNOLOGY PROVIDER DOESN'T HAVE TO SUCK THE TEACHER'S BLOOD
Authors	Cathy Chua; Manny Rayner
DOI	doi:10.21125/iceri.2019.1863
Type	Proceedings article
Published in	ICERI2019 Proceedings
Published by	IATED

Title	EASY CONSTRUCTION OF MULTIMEDIA ONLINE LANGUAGE TEXTBOOKS AND LINGUISTICS PAPERS WITH LARA
Authors	Matthias Butterweck; Cathy Chua; Hanieh Habibi; Manny Rayner; Ghil'ad Zuckermann
DOI	doi:10.21125/iceri.2019.1737
Type	Proceedings article

Miloshevska, L., Delibegović Džanić N., Hatipoğlu Ç. & Gajek E. (2021) Crowdsourcing for language learning in Turkey, Bosnia and Herzegovina, Republic of North Macedonia and Poland, Journal of Narrative and Language Studies, 9(16), 106-121.

Nicolas, L., Aparaschivei, L., Lyding, V., Rodosthenous, C., Sangati, F., König, A., & Forascu, C. (2021). An Experiment on Implicitly Crowdsourcing Expert Knowledge about Romanian Synonyms from L1 Language Learners. Proceedings of 10th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2021), 1–14.

Nicolas, L., Ostanina-Olszewska, J., Arhar Holdt, Š., Čibej, J., Borg, C., Lyding, V., & Barreiro, A. (2021). Introducing an implicit crowdsourcing opportunity to teachers. In CALL for Background (pp. 115–136). Peter Lang.

Bédi, B., Bernharðsson, H., Chua, C., Björg Guðmarsdóttir, B., Habibi, H., & Rayner, M. (2020, August 20). Constructing an interactive Old Norse text with LARA. Proceedings of EUROCALL 2020, European Association for Computer Assisted Language Learning. EUROCALL 2020, European Association for Computer Assisted Language Learning., Copenhagen.

Rodosthenous, C., Lyding, V., Sangati, F., König, A., ul Hassan, U., Nicolas, L., Horbacauskiene, J., Katinskaia, A., & Aparaschivei, L. (2020). Using Crowdsourced Exercises for Vocabulary Training to Expand ConceptNet. Proceedings of LREC 2020, 12th Language Resources and Evaluation Conference, 307--316.

Nicolas, L., Lyding, V., Borg, C., Forascu, C., Fort, K., Zdravkova, K., Kosem, I., Čibej, J., Arhar Holdt, Š., Millour, A., König, A., Rodosthenous, C., Sangati, F., ul Hassan, U., Katinskaia, A., Barreiro, A., Aparaschivei, L., & HaCohen-Kerner, Y. (2020). Creating Expert Knowledge by Relying on Language Learners: a Generic Approach for Mass-Producing Language Resources by Combining Implicit Crowdsourcing and Language Learning. Proceedings of LREC 2020, 12th Language Resources and Evaluation Conference, 268--278.

Horbach, A., Aldabe, I., Bexte, M., Lopez de Lacalle, O., & Maritxalar, M. (2020). Linguistic Appropriateness and Pedagogic Usefulness of Reading Comprehension Questions. Proceedings of LREC 2020, 12th Language Resources and Evaluation Conference, 1753–1762.

Bédi, B., Chua, C., Habibi, H., Martinez-Lopez, R., & Rayner, M. (2019, August 29). Using LARA for language learning: a pilot study for Icelandic. CALL and Complexity – Short Papers from EUROCALL 2019. EUROCALL 2019 - CALL and Complexity, Louvain-la-Neuve, Belgium.

Rodosthenous, C., Lyding, V., König, A., Horbacauskiene, J., Katinskaia, A., ul Hassan, U., Isaak, N., Sangati, F., & Nicolas, L. (2019, May 21). Designing a Prototype Architecture for Crowdsourcing Language Resources. Poster Session of the 2nd Conference on Language, Data and Knowledge 2019. Conference on Language, Data and Knowledge 2019.

Alice Millour, Marianne Grace Araneta, Ivana Lazić Konjik, Annalisa Raffone, Yann-Alan Pilatte, et al.. Katana and Grand Guru: a Game of the Lost Words (DEMO). 9th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC'19), May 2019, Poznań, Poland.

Lyding, V., Rodosthenous, C. T., Sangati, F., ul Hassan, U., Nicolas, L., König, A., Horbacauskiene, J., & Katinskaia, A. (2019, April 8). v-trel: Vocabulary Trainer for Tracing Word Relations - An Implicit Crowdsourcing Approach. Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2019. International Conference Recent Advances in Natural Language Processing, RANLP 2019, Varna, Bulgaria.

Zdravkova, K., Fort, K., and Bédi, B. (eds) (2018). CEUR Workshop proceedings of the enetCollect WG3 & WG5 Meeting 2018, Leiden, The Netherlands, October 24-25, 2018, http://ceur-ws.org, volume 2390.

Zingano Kuhn, T., Dekker, P., Šandrih, B., Zviel-Girshin, R., Arhar, Holdt, Š. & Schoonheim,T. (2019). Crowdsourcing Corpus Cleaning for Language Learning Resource Development. In EuroCALL 2019: European Association of Computer Assisted Language Learning.

Zingano Kuhn, T., Todorović, B. Š., Holdt, Š. A., Zviel-Girshin, R., Koppel, K., Luís, A. R., & Kosem, I. (2021). Crowdsourcing Pedagogical Corpora for Lexicographical Purposes. In L. Mitits & S. Kiosses (eds.), Lexicography for Inclusion: Proceedings of the 19th EURALEX International Congress, 7-11 September 2021, Alexandroupolis, Vol. 2., pp. 771–779. Democritus University of Thrace.

Alfter, D., Lindström Tiedemann, T. & Volodina, E. (2021). Crowdsourcing Relative Rankings of Multi-Word Expressions: Experts versus Non-Experts. Northern European Journal of Language Technology, Vol.1.