elexis european lexicographic infrastructure

D6.1

EARLY ELEXIS INTEROPERABILITY REPORT

Author(s): Andraž Repar, Iztok Kosem, Simon Krek

Date: July 31, 2019

H2020-INFRAIA-2016-2017

Grant Agreement No. 731015

ELEXIS - European Lexicographic Infrastructure

D6.1 EARLY ELEXIS INTEROPERABILITY REPORT

Deliverable Number:	D6.1
Dissemination Level:	Public
Delivery Date:	July 31, 2019
Version:	3
Author(s):	Andraž Repar, Iztok Kosem, Simon Krek

Project Acronym:	ELEXIS
Project Full Title:	European Lexicographic Infrastructure
Grant Agreement No.:	731015

Deliverable/Document Information

Project Acronym:	ELEXIS
Project Full Title:	European Lexicographic Infrastructure
Grant Agreement No.:	731015

Document History

Version Date	Changes/Approval	Author(s)/Approved by
June 30, 2019	Initial draft	Andraž Repar
July 15, 2019	Internal revision	Simon Krek
July 31, 2019	External evaluation	Michal Kren

Table of Contents

1 About the report 3

2 ELEXIS Tools and services developed by M15 4

2.1 Sketch Engine 4

2.1.1 Usage statistics 5

2.1.2 User feedback 8

Prof. Marie-Aude Lefer 8

Elisa Corino 10

Dr Ana Frankenberg-Garcia 11

2.2 Lexonomy 14

2.2.1 Lexonomy community and hackathons 14

2.2.2 User feedback 15

Ivana Filipović Petrović 15

Primož Ponikvar 16

2.3 Elexifinder 18

2.3.1 User feedback 19

Robert Lew 19

2.4 ELEXIS News Feed 21

2.4.1 Usage statistics 21

Facebook 21

Twitter 23

2.4.2 User feedback 24

Edward Finnegan 24

3 ELEXIS infrastructure as an ecosystem of interconnected elements 26

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

4 Conclusion 27

List of Figures

Figure 1. Number of institutions per country who joined the ELEXIS project and took advantage of the free-of-charge availability of Sketch Engine 5

Figure 2. The number of ELEXIS accounts since the start of the project. 6

Figure 3. Number of users and users 1+ per month. The small difference between the two bars indicates that most users use Sketch Engine for extensive periods of time. 7

Figure 4. The number of userhours (and userhours 1+) per month since the start of the project. The peaks are associated with increased activity during the end of summer and winter study terms. 8

Figure 5. Tag cloud of top keywords extracted from the articles related to the search query “dictionary”. 18

Figure 6. A list of authors who wrote the most articles in the results for the search query “dictionary”. 19

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

1 About the report

This report provides an overall assessment of ELEXIS tools and services developed by M15. Its goal is to evaluate and report on the impact of the whole ELEXIS infrastructure as an ecosystem of interconnected elements. It contains quantitative information in the form of usage statistics and qualitative assessment in the form of user feedback.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

2 ELEXIS Tools and services developed by M15

The following tools and services have been made available through the ELEXIS project by M15:

- Sketch Engine
- Lexonomy
- Elexifinder
- ELEXIS News Feed

2.1 Sketch Engine

The Sketch Engine corpus query, corpus building and corpus management system allows users to build and work with 300+ text corpora in over 90 languages and 20 scripts. Sketch Engine contains a number of unique tools to analyse large corpora of up to 30 billion words. Each user can benefit from fully automated dictionary-building functionality.

The access to Sketch Engine is funded by the EU through the ELEXIS project between 2018 and 2022. The access is provided at no cost to academic institutions and ELEXIS observers and applies to non-commercial use only.

Sketch Engine provides the following functionalities:

- Corpus building tool
- Word sketch
- Word sketch difference
- Thesaurus
- Concordance
- Parallel concordance
- Wordlist
- N-grams
- Keywords & terms
- Trends
- OneClick dictionary

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

2.1.1 Usage statistics

We provide usage statistics from the start of the project to the end of June. At the end of June, 296 institutions from 16 countries have signed up for ELEXIS access to Sketch Engine. For details, see Figure 1.

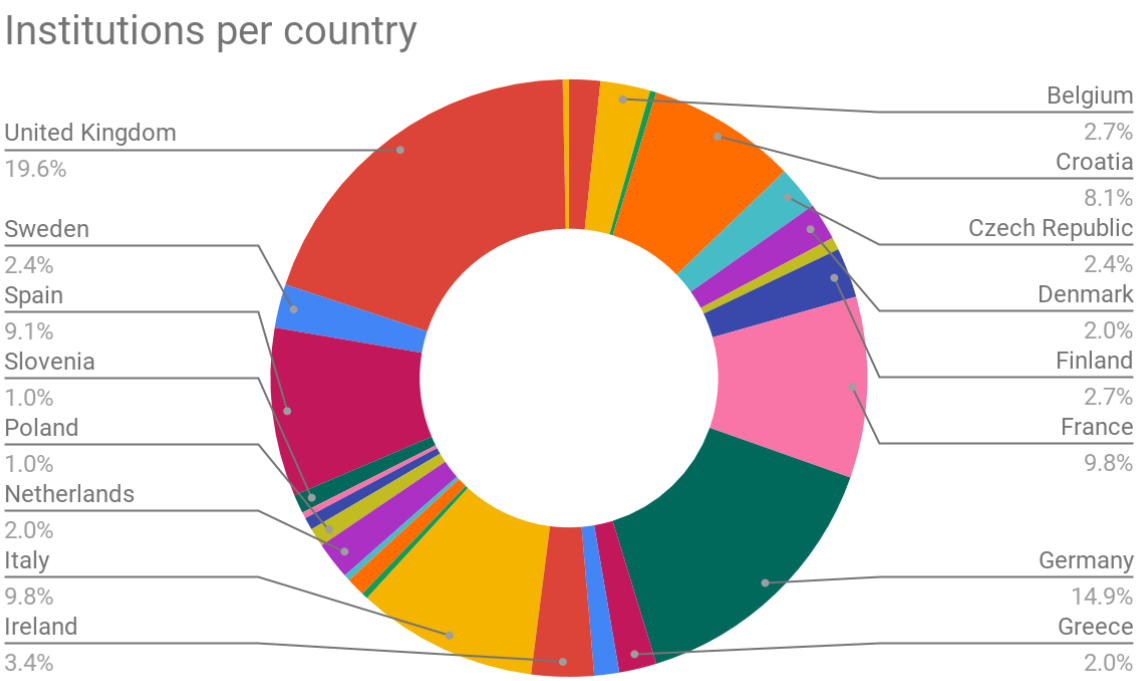

Figure 1. Number of institutions per country who joined the ELEXIS project and took advantage of the free-of-charge availability of Sketch Engine

As of June 30, 2019, there have been a total of 17,110 ELEXIS user accounts registered in Sketch Engine with considerable growth recorded since the start of the project. Figure 2 shows the growth trend of user accounts since April 2018.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

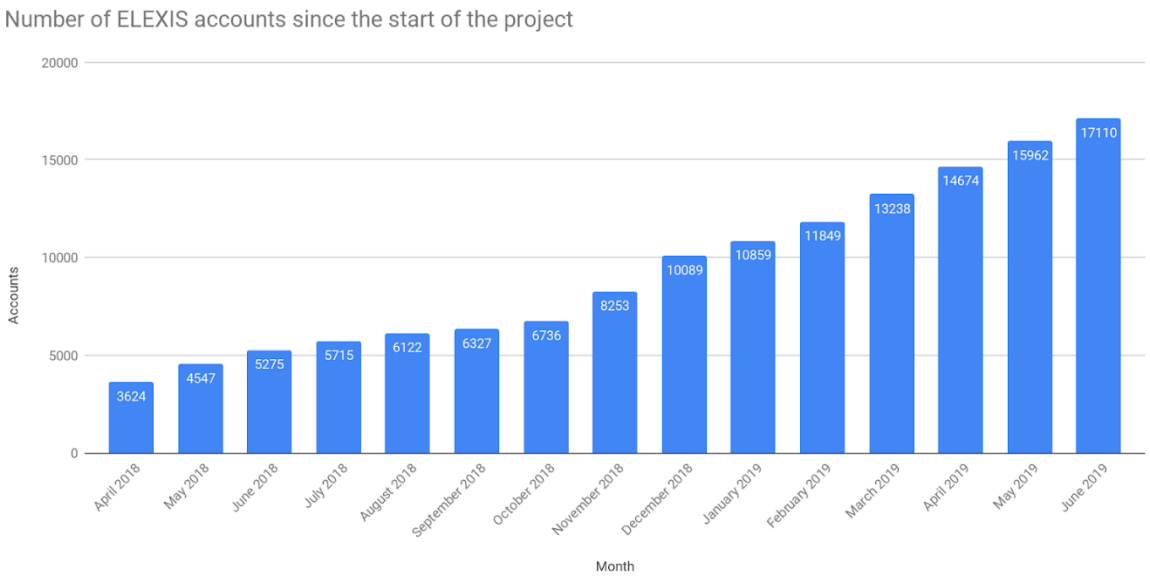

Figure 2. The number of ELEXIS accounts since the start of the project.

We are also monitoring the time spent by these users in Sketch Engine. The following metrics are used:

- **users**: shows the number of unique users who used Sketch Engine in that month
- **userhours**: the number of userhours the users spent working with Sketch Engine

To investigate the ratio of "acidental users", e.g. ones who only logged in once and did not really work with the system, we looked at two additional indicators:

- **users 1+:** shows the number of unique users with 2 or more userhours
- **userhours 1+:** shows the number of userhours produced by users 1+

In Figure 3, we can observe that there is only a small number of “accidental users”. Most users continue using Sketch Engine which indicates that they find the tool useful for their daily work.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

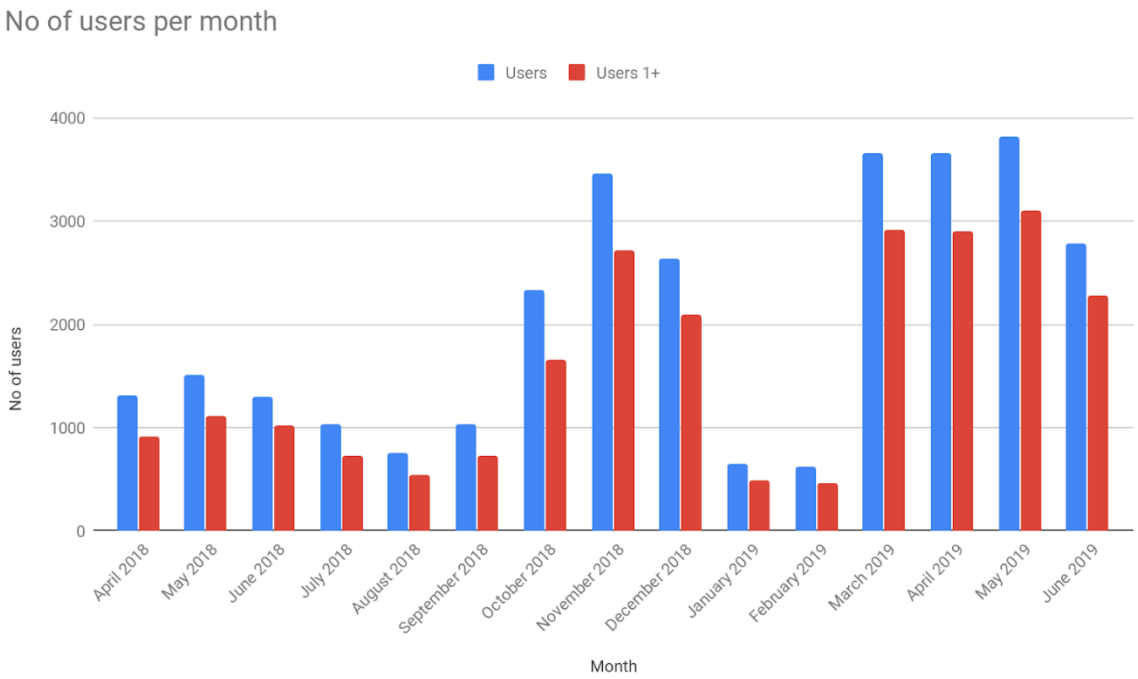

Figure 3. Number of users and users 1+ per month. The small difference between the two bars indicates that most users use Sketch Engine for extensive periods of time.

A similar trend can be observed in terms of the number of userhours. As of June 30, 2019, users have logged a total of 322,292 hours in Sketch Engine. The number of userhours 1+ is high indicating a high engagement of Sketch Engine users. See Figure 4 for details.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

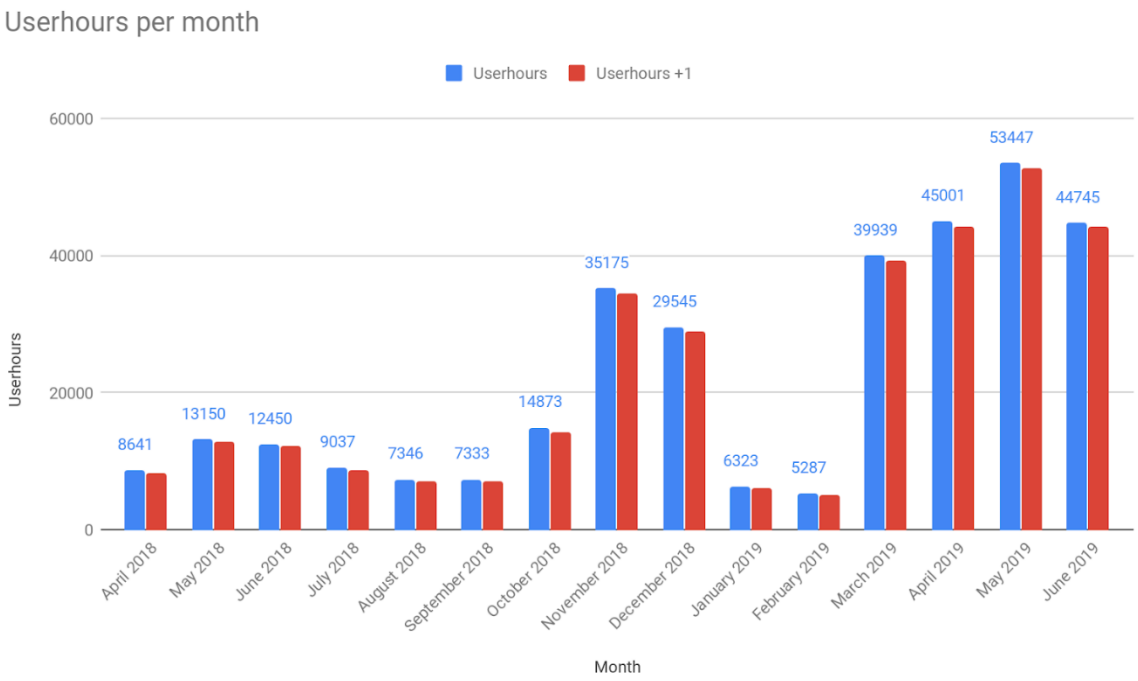

Figure 4. The number of userhours (and userhours 1+) per month since the start of the project. The peaks are associated with increased activity during the end of summer and winter study terms.

2.1.2 User feedback

To evaluate the impact that Sketch Engine availability offered through the ELEXIS project has on researchers in the area of lexicography and related fields, we have asked three users (who have gained access through ELEXIS) to provide a short evaluation summarizing their experience with Sketch Engine. Their contributions can be found below:

Prof. Marie-Aude Lefer
Université catholique de Louvain, Belgium

The lecturers and master's students of the Louvain School of Translation and Interpreting (LSTI) at the UCLouvain (Belgium) rely intensively on Sketch Engine. The tool is mostly used in specialized translation courses (financial translation, technical translation, legal translation, translation of EU-related texts, etc.) – with students translating from a range of source languages (Dutch, English, German, Italian, Russian, Spanish, Turkish), their L2s, into French, their L1. Typically, users upload their own DIY specialized corpora (mostly in the target language, French) to (1) extract terminology in order

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

to build terminological databases that can then be uploaded in computer-aided translation tools and (2) find solutions to specific translation problems encountered while translating a given source text (especially as regards phraseology, which is particularly tricky in specialized translation, even if students translate into their native language). The most frequently used Sketch Engine tools are wordlists, n-grams, keywords and terms, and word sketches (to extract terminology and phraseological patterns) and concordances (to find solutions to specific translation problems). As a matter of fact, these functionalities of Sketch Engine are taught to all master's students in translation within the framework of a compulsory course in year 1, alongside other computer-aided translation tools (e.g. translation memories and terminological databases). All translation students at the LSTI are therefore familiar with Sketch Engine and most of its functionalities and can use the tool for all their translation tasks. Because students are now familiar with the tool and corpus methodologies in general, many build their own DIY specialized corpora and exploit them in Sketch Engine for their master's dissertation (a 60-page translation project they all have to carry out in year 2). It should also be added that in addition to using user-defined corpora in Sketch Engine, lecturers and students also frequently rely on the monolingual reference corpora and parallel corpora available in Sketch Engine.

Users at the LSTI – whether lecturers or students – also use Sketch Engine to do research in the fields of terminology and corpus-based translation and interpreting studies. Several lecturers are currently supervising master's dissertations which make intensive use of Sketch Engine to analyze corpus data. Sketch Engine offers the most powerful extraction techniques to date, and clearly outperforms other available concordancers (whether freely available or not), such as AntConc or WordSmith Tools, which do not offer part-of-speech tagging.

It is also important to add that Sketch Engine is used by many other users at the UCLouvain outside the LSTI, such as language and linguistics students and researchers in linguistics from the Language and Communication research institute (IL&C), many of whom are active in the field of corpus linguistics (including learner corpus research).

The UCLouvain staff and students are happy to be allowed to use Sketch Engine for free as part of the ELEXIS project. The LSTI, in particular, enjoy this free use very much as they need to buy many other computer-aided translation tools, such as SDL Trados Studio, which are particularly expensive. In view of Sketch Engine's free access, Marie-Aude Lefer and Geneviève Maubille, who both teach at the LSTI and are very active users of Sketch Engine, have recently taught Sketch Engine workshops at the

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

translation departments of the Université libre de Bruxelles and the Université de Mons in Belgium, which are both envisaging the use of the tool with their translation students in the near future.

Elisa Corino
Università di Torino, Italy

Sketch Engine (SkE) is a powerful tool for both research and teaching purposes. As manager of the university account, I have observed an increase in its use in the last couple of years, as more and more users ask for a storage space increase.

Personally, I use SkE for research in LSP contexts, using WebBootCat and the Build a corpus function to create customized specialized corpora. Extracting collocations, word lists, n-grams helps me in defining the characteristics of a given langue in a given context (terminology, collocations, constituent order).

I encourage my students of Translation courses to do the very same: they create the specialized corpus and then extract frequencies, significant word sketches and sketch diff to create glossaries and survey the language they are dealing with, in both monolingual and parallel corpora.

And my colleague teaching English language does the very same. Many of my students also use SkE for their MA thesis in linguistics and translation, enriching their work on language with figures and statistics extracted from the corpora available on the platform or creating their personal ones.

My students of the course in Approaches to language teaching have a whole module on corpus linguistics where they learn how to use SkE for didactic purposes.

I also recommend the use of SkE in CLIL contexts. During training courses for teachers, I always present the tools and show them how to use the different features to a) prepare materials for language classes (both FL and Italian classes), and b) devise Data-driven learning activities where students actively use the corpora.

After a period of adjustment, I particularly appreciated the new interface, with word clouds and visual aids that help in sketching figures for scientific papers. The most used functionalities I generally use

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

are: the word sketch, the KWIC, the frequency lists. I really appreciate the CQL query format that allows me a more syntax-oriented research.

When I compile my own corpus I also like to create subcorpora. I would like a function to create parallel corpora easily; the current way of doing it is definitely time consuming, having a text aligner integrated to the system would be of great benefit.

Dr Ana Frankenberg-Garcia
University of Surrey, UK

I have been a regular user of corpora since 1999, and a regular user of Sketch Engine since 2011. I have absolutely no hesitation in stating that Sketch Engine is the most powerful yet user-friendly corpus software that I am aware of.

I began to use Sketch Engine in my work as Chief Editor of the Oxford Portuguese Dictionary, to develop the Portuguese framework of the dictionary and bilingual equivalences, and now as Reader in Translation Studies at the University of Surrey, to support my research and teaching.

Initial use of Sketch Engine at the University of Surrey was limited to 20 user accounts. This was due to funding limitations. Now, with the support from Elexis, the entire university has access to the software and resources distributed by Sketch Engine.

Staff and students in my department – the Centre for Translation Studies – are all introduced to Sketch Engine and are able to use its basic features (simple search concordances, word sketches and thesaurus), which they utilize regularly in everyday translation assignments and other activities. Students taking the Corpora & Translation module at Masters level to learn to use Sketch Engine in further depth, and rely on it to build corpora, extract terminology and address a wide range of research questions.

I currently have 3 Masters students and one PhD student using Sketch Engine in their research: their work involves researching terminology in the work of a contemporary American philosopher; compiling a specialized bilingual Norwegian-English corpus and terminology database in the domain of textile crafts; compiling a specialized bilingual Spanish-English corpus in the domain of Tourism in Spain to investigate the translation of culture-specific items; and compiling a major Chinese-English

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

parallel corpus to investigate translation norms from Chinese into English. One of my former PhD students used Sketch Engine to research ideology in Arabic-English political speech translations. There are also a number of other Masters and PhD students who I am not directly supervising also using Sketch Engine in their research (for example, to analyse metaphor in popular science, and to investigate the language of closing arguments in American trials).

Outside my research centre, other departments at the University of Surrey – like the people in TESOL (Teaching English to Speakers of Other Languages), Education and Tourism – have also been using Sketch Engine on a regular basis.

In my research, Sketch Engine is an essential tool for the ColloCaid project, where we are developing a text editor/writing assistant to help writers with academic English collocations. The project website has links to several papers in which we reference Sketch Engine and explain detail how it is used in ColloCaid.

Sketch Engine is also being used in a research collaboration between my university in the UK and two Brazilian universities, to support the internationalization of Brazilian research. Although Brazilian universities are not eligible to Elexis funding, Sketch Engine has helped us promote four academic writing workshops in Brazil, where Brazilian researchers learned to build corpora in their specialist domains to help them improve their research writing in English.

A similar workshop was hosted in collaboration with the University of Leon, Spain, in the beginning of June 2019. This time, researchers at the University of Leon were able to access Sketch Engine thanks to Elexis.

There are also colleagues in my research centre using Sketch Engine in the MEMAD project (in collaboration with other European universities), to develop novel methods and models for managing and accessing digital audiovisual information in multiple languages.

Our research plans for the immediate future on the intelligent integration of human and machine in different modalities of translation and interpreting will also mean continuing to make intensive use of Sketch Engine.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

Outside university, I have used Sketch Engine in a series of Professional Development workshops to train translators on the use of corpora for practical translation. These highly-successful one-day workshops hosted by the UK Institute of Translation and Interpreting show that it is possible to learn to use Sketch Engine very quickly. Extended trial licenses for the workshop sponsored by Lexical Computing allowed participants to continue to use Sketch Engine for three months.

In summary, Sketch Engine is an integral part of teaching and research at the University of Surrey Centre for Translation Studies, and we are grateful that Elexis has allowed us to use it more widely with our own students and to disseminate its use to other departments in our university. The use of Sketch Engine in professional training workshops for the UK Institute of Translation and Interpreting and in the workshops to support researchers in Brazil only serve to show that the impact of Elexis goes well beyond institutions directly benefitting from Elexis.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

2.2 Lexonomy

Lexonomy is a cloud-based dictionary writing and also online dictionary publishing system which is highly scalable to adapt to large dictionary projects as well as small lexicographic works such as editing and online publishing of domain-specific glossaries or terminology resources.

Lexonomy already interacts with Sketch Engine and the aim of the project is to develop and expand this interaction further. Sketch Engine can push lexicographic data into Lexonomy to create automatically-generated dictionary drafts and Lexonomy can pull data from Sketch Engine's corpora during the entry editing process.

2.2.1 Lexonomy community and hackathons

In order to facilitate Lexonomy development and collaboration between Lexonomy users (lexicographers and developers), we have created a project on GitHub (https://github.com/elexis-eu/lexonomy) and a Google Group (elexis-lexonomy@googlegroups.com). On GitHub, one can follow the latest development activities, including reported issues, bugs and feature requirements. Google group is used more for general discussions, arranging meetings etc.

In addition to regular online meetings, in 2019 the decision was made to introduce regular Lexonomy hackathons where all members of the Google Group meet and intensively work (for a few days) on Lexonomy improvements. While the developers focus on technical aspects, the lexicographers work on identifying missing features, prioritizing features for future implementation, and preparing/updating Lexonomy documentation (instructions for use, help.

In the first half of-2019, there were two Lexonomy hackathons. The first one took place from 23rd to 25th April 2019 in Brno and was attended by nine people. The first day focused on a general development plan for Lexonomy, while the other two focussed on addressing specific issues, especially more problematic ones such as opening up large files, crashing when opening invalid DTD etc. The second hackathon took place from 12th-14th June 2019 and was attended by ten people. More focus was on solving specific persistent issues and bugs, whereas general discussions covered topics such as cross-references, format support, XSL schema support and documentation.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

2.2.2 User feedback

Ivana Filipović Petrović
Institute for Linguistic Studies, Croatian Academy of Sciences, Croatia

Lexonomy is used for writing the Online Dictionary of Croatian Idioms (ODCI) which is currently under development in the Institute for Linguistic Studies of the Croatian Academy of Sciences. The project of creating this born-digital corpus-driven dictionary started in 2018 and we chose Lexonomy for our DWS for several reasons.

Firstly, we needed a DWS with a lexicographer-friendly dictionary editing interface, i.e. for users who have little or no knowledge in programming world. When we started, my colleague and I had some ideas about how entry should look and elements it should contain. So, in a less than two weeks, two linguists and lexicographers without any experience with XML data format, edited all elements and attributes of an entry in our dictionary in a wanted way. I find the interface in Lexonomy very user-friendly and cozy to work with. Once when you define the parts of an entry in your dictionary and the appearance of elements, you can very easily and quite fast build new entries, filling the boxes with text or choosing from list that you created earlier. Also, if you pick laic mode, you won't even know that there is an XML data format in the whole story.

We started to create our dictionary from the scratch manually, given that we don't have the usual list of headwords. When it comes to idiom dictionaries, we believe that the network of alphabetically ordered single-word headwords is a feature of the paper dictionaries, created for easing the search, and is not needed in e-dictionary, given that the user can search an idiom by any of its components. Accordingly, our headword is an entire citation form of an idiom, shaped according to the data about frequency from Croatian web corpus hrWaC, available in Sketch Engine.

Moreover, the conceptual (thematic) organization is a very important feature of ODCI and we accomplished satisfying results in Lexonomy in this area as well. The most helpful function in this sense so far is clickable cross-reference, by which we interconnect idioms with equal, similar or opposite meaning. It would be useful if thesaurus function is more worked out in a way that we can include thematic fields in the entry and that user can search the dictionary by thematic fields and that results include all idioms that are interconnected in this particular field. Given that thematic organization has become (almost) a norm in e-dictionaries, thanks to the fact that space constrains are no longer an

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

issue as they were in print lexicography, it is very important to us to provide a dictionary for those users who seek for words in order to express their ideas, and not (just) the other way around. Also, important part of ODCI are usage notes. These notes usually contain further explanations of how the idiom is used or varied, interesting facts about where it comes from and how it can be used to play with language. While configuring the entry in ODCI in Lexonomy, we chose for this type of information to be in a box with a thin border, and this is possible due to fact that dictionary visualization in Lexonomy is quite rich and very adjustable. When it comes to other functionalities, sometimes we use automatic extraction of examples from Sketch Engine. Actually, we are continuing exploring functionalities of Lexonomy and adjusting our entries according to new insights. Also, when dictionary become publicly accessible, a drop-down menu would make us happy.

Finally, the fact that Lexonomy is an open-source tool greatly facilitated and accelerated our work. In other words, we were able to start with this project without significant institutional or financial support, all we needed was an internet connection and a computer. In addition, we can count on Lexonomy also for online publishing of ODCI. To a lexicographer with (too) many years of experience in writing dictionary in Word, Lexonomy is a very helpful, user-friendly, visually very pleasantly designed tool that meets our lexicographical desires quite successfully.

Primož Ponikvar
Center for Language Resources and Technologies, University of Ljubljana, Slovenia

We have been using Lexonomy for over a year now, mainly for the compilation of Slovenian-Hungarian dictionary but also a few other dictionaries. Our lexicographers are very satisfied with what Lexonomy has to offer, and also with the way it has developed over the past year in order to meet their needs. On the lexicographer side, we have begun in a rather simple manner, by using the nerd or laic mode of editing XML. On the user side, we collaborated with the developers to prepare an XSLT script for the publication of sample entries; using the script, it was possible to include user-friendly features such as More/Less buttons for collocations and examples, audio files etc.

More benefits of further development of Lexonomy within ELEXIS started to show after the Observer event in Vienna where the first glimpses of using a customised Entry Editor were shown. This was of

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

particular importance to us as our lexicographers felt that a lot of time and effort could be saved if the editor was made simpler, i.e. the editing would not be done on XML directly. Together with our developer, we thus prepared our version of Entry Editor which had two important features: a) editing was done in format as presented to the users (not in XML), and b) we were able to limit the editing only to the elements relevant for the lexicographers (translation, labels and a few others). In addition, we have adopted the Flags feature in Lexonomy for our workflow; in fact, while initially we thought that Lexonomy is lacking a feature to monitor lexicographic workflow, we found Flags to be quite sufficient for our needs.

At this point, I should mention that some of these aforementioned features are a direct consequence of the fact that our developer has joined the Lexonomy community of GitHub and attended both Lexonomy hackathons that took place in the first half of 2019. In this way, we were able to not only learn from other projects but also collaborate on identifying and solving most pressing needs. At the time of writing, we have already started testing the connection with the Sketch Engine, especially pulling of (good) corpus examples from the corpora.

In sum, our experience with Lexonomy has been only positive. Our lexicographers have nothing but praise for it, they like it that they can use it without installing any special software, that it is very intuitive and simple to use, and that it can be customised and further developed to meet their needs – in other words, that it "listens" to them. I hope that the Lexonomy community will grow further and more collaborations between Lexonomy users will develop; I see this as one of the main advantages that Lexonomy has over other dictonary-writing systems I have encountered and used so far.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

2.3 Elexifinder

The search tool ELEXIFINDER is dedicated to helping lexicographers and other researchers find scientific output in lexicography and related fields. The enables users to search through papers and videos, using concepts, i.e. words or set of words with a Wikipedia page, and various other conditions, e.g. source (conference etc.), author, language etc. Each paper/video is linked to its page where the users can download or view it. The database currently contains 1755 papers from EURALEX and eLex conferences and 86 video presentations, in 11 different languages. ELEXIFINDER uses the latest visualization and summarization user interface techniques helping users find exactly what they want faster and with less effort (for details, see Figures 5 and 6).

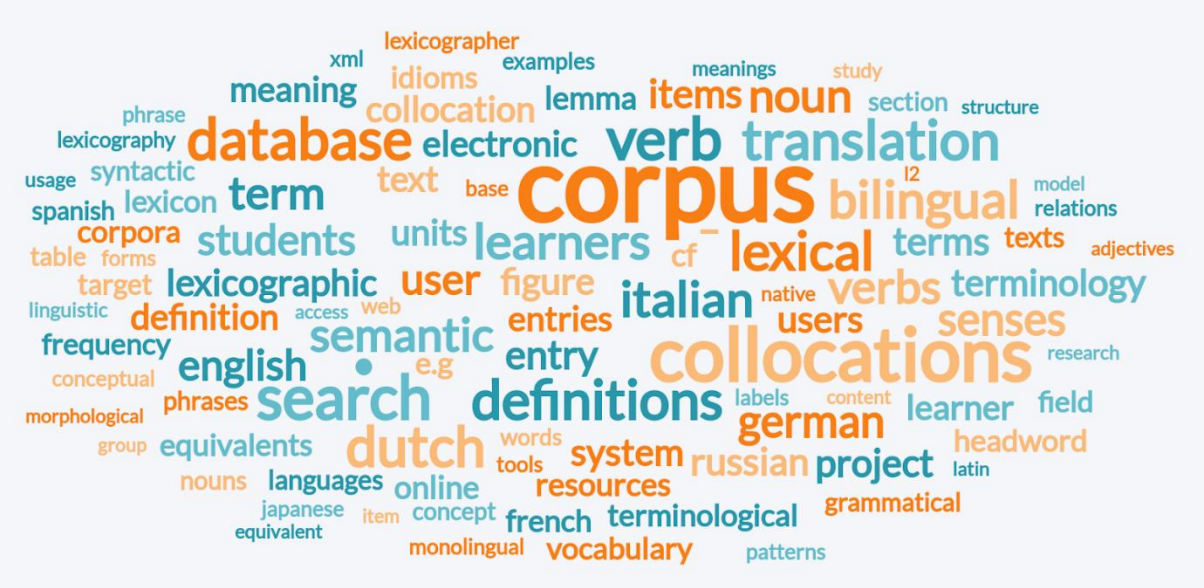

Figure 5. Tag cloud of top keywords extracted from the articles related to the search query "dictionary".

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

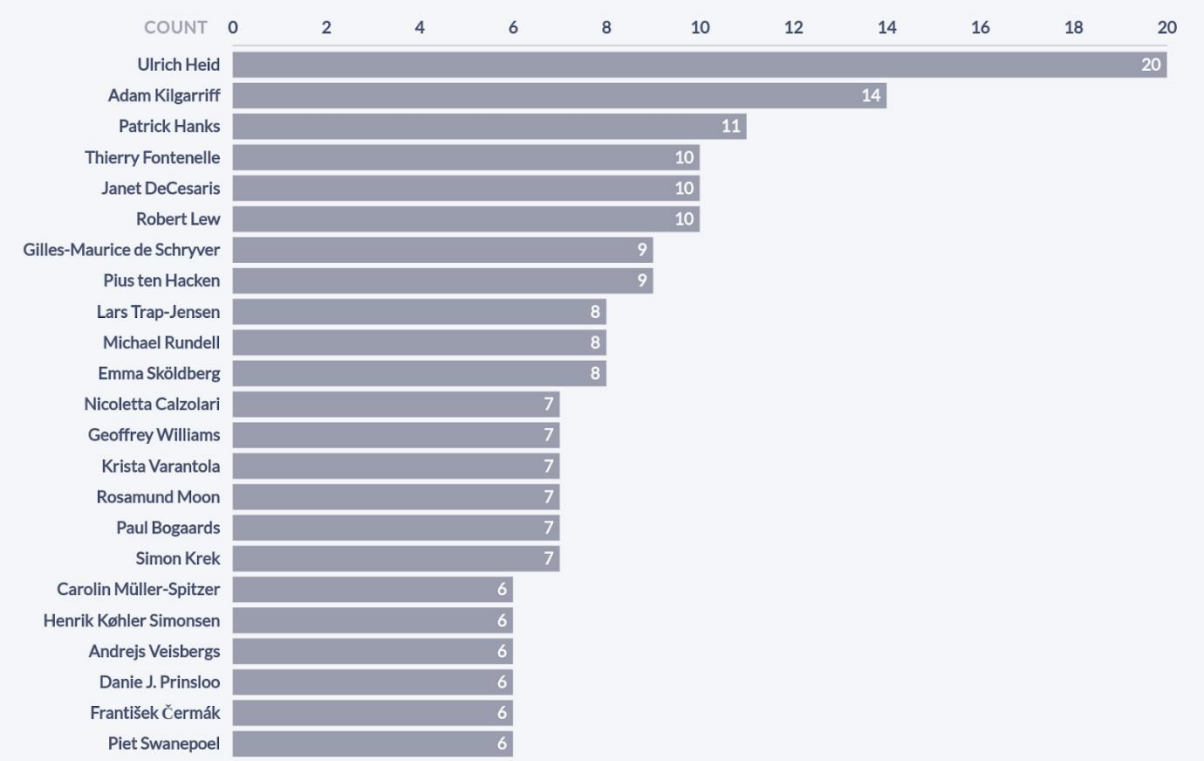

Figure 6. A list of authors who wrote the most articles in the results for the search query "dictionary".

2.3.1 User feedback

Robert Lew
Adam Mickiewicz University, Poland

ELEXIFINDER is an online search tool currently under development (http://er.elex.is) that seeks to offer a modern inteface for searching publications, broadly conceived, and including media such as presentation video recordings, within the area of lexicography.

ELEXIFINDER provides a clean modern interface with a salient text search box across the top, and six selection boxes immediately below, representing six different ways of filtering the results. The main searchbox suggests keywords and placenames (apparently places associated with publications, such as affiliations). The keywords come from a closed set. However, it's also possible to enter words that are not keywords as well as phrases combining several words. The filters include

1. geographical location, which can be entered at varied levels of granularity (e.g. country or city);

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

2. sources, which includes conference sets, though it is referred to by the generic and in this case somewhat misleading term “news sources”; I suggest that this narrative be edited down to just “sources”;
3. categories, which includes a very broad taxonomy of topics which in its majority doesn’t seem applicable to lexicographic publications (e.g. Games, Recreation, Shopping, with finer subdivisions); I suggest that this generic taxonomy be replaced with one more appropriate for lexicography; alternatively, this filter could just be removed;
4. time, which is self-explanatory;
5. language, which probably refers to the language in which a given article has been written; however, for our purposes at least as useful (or more) would be the language with which the publication deals; for example, I haven’t found anything under Polish, although there are articles dealing with Polish lexicography that are written in English; it would be useful to be able to locate those using Polish as a language tag;
6. miscellaneous, which includes the media type: articles and videos; presentation and audio files might also be a possibility in the future

The tool includes a map-based presentation, which does not load correctly as of this writing (though I recall having used it in the past), so I can’t currently comment on this functionality.

Once a set of articles has been returned, there are options to generate a tag cloud on the fly, as well as statistics of authors (quite interesting and useful for selecting reviewers), countries, and conferences (which themes were big at which conference?). A concept graph is also available along with a few other visualizations.

As of this writing, the database underlying the tool includes 1755 papers from the past EURALEX and eLex conferences, as well 86 video presentations. Fairly large as it seems, this is still fairly limited. The search tool will be of much more use if it succeeds in including a greater variety and volume of materials. As far as I am aware, this is exactly the idea, and work towards that end is in progress.

Overall, this looks like a versatile and well-designed tool with much potential, as long as a a more extensive database of articles are included.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

2.4 ELEXIS News Feed

Lexicographic news feed is an ELEXIS service that uses the Event Registry API to extract latest news articles identified to be related to lexicography. News articles are extracted from 30,000 news sources, and over 35 languages are currently supported.

2.4.1 Usage statistics

Starting in March 2019, news articles from the Elexis News Feed service have been published on the Elexis social media channels (Facebook and Twitter). Comparison to other project-related social media posts shows that the News Feed posts have respectable performance on social media.

Facebook

The Facebook post performance is measured using two standard measures:

- **Reach**: Reach is the number of people that have seen the content within a certain period.
- **Engagement**: The total number of actions (likes, comments, shares, clicks) that people take involving your ads.

In Table 1, we compared the performance of Facebook posts linking to news articles found via the Elexis News Feed to the overall top performing Facebook posts. We can observe that while the top performing posts have better reach and engagement, the News Feed posts nevertheless achieve respectable results occassionally even coming close to the top 3 overall spots (i.e. the two posts in March). The total reach of the News Feed posts from March to June was 2319 (total reach of all posts: 32.411) and the total engagement of the News Feed Posts was 318 (total engagement of all posts: 4880).

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

March

Type	Title	Reach	Eng
General	Asil TG Portrait	910	256
General	Summer School Dagstuhl	872	31
General	eLex CfP	733	69
News Feed	**"Les secrets des mots" by Jean Pruvost:**	**612**	**67**
News Feed	**Professor & the mad man**	**413**	**69**

April

Type	Title	Reach	Eng
General	Lisbon Summer School 2019	1892	66
General	Lexonomy Hackathon	923	47
General	DH workshop	773	30
News Feed	**Leopoldo Lugones, la lengua castellana de América**	**177**	**8**
News Feed	**Announcement Newsfeed**	**154**	**34**
News Feed	**Cinq mots a ressusciter de toute urgence**	**100**	**4**

May

Type	Title	Reach	Eng
General	Lisbon Summer School	1370	61
General	Villa Vigoni Thesen II	1305	71
General	totH2019	280	23
News Feed	**the secret life of dictionaries**	**172**	**46**
News Feed	**Iox' word that hasn't changed in sound or meaning**	**151**	**33**
News Feed	**Make the world a better place by using theses rarely heard words**	**113**	**10**
News Feed	**Los secretos detras de la cuenta qza la rae us en twitter para reponder consultas**	**88**	**12**

June

Type	Title	Reach	Eng
General	TEI lex 0 ASIALEX	692	79
General	lexicography in the nordic region	520	18
General	summer datathon ssdllod sina	280	46
News Feed	**What do they call french toast in france?**	**146**	**20**
News Feed	**RAE Dissionario de la lengua de signos**	**111**	**12**
News Feed	**Blacks law dictionary, T. Reuters**	**82**	**3**

Table 1. Perfomance of News Feed Facebook posts compared to the top 3 performing posts on the ELEXIS Facebook page.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

Twitter

The Twitter performance is measured using two standard measures:

- **Impressions**: Times a user is served a Tweet in timeline or search results
- **Engagement**: Total number of times a user interacted with a Tweet. Clicks anywhere on the Tweet, including Retweets, replies, follows, likes, links, cards, hashtags, embedded media, username, profile photo, or Tweet expansion

Table 2 below contains statistics about the top performing tweets on the ELEXIS Twitter profile and the tweets coming from the ELEXIS News Feed. We can observe a similar pattern as with Facebook, News Feed tweets have respectable performance but are usually not among the top three tweets in a single month—with the exception of the the tweet "Professor & the mad man" which has the second best engagement in March.

The total number of impressions of the News Feed posts from March to June was 9177 (total impressions of all posts: 187,400) and the total engagement of the News Feed Posts was 124 (total engagement of all posts: 1782).

March

Type	Title	Imp	Eng
General	Survey of Lexicographers' Needs	2958	31
General	Women in Culture and Tech	2425	19
General	Elexis workshop	2048	63
News Feed	**Professor & the mad man**	**1563**	**44**
News Feed	**"Les secrets des mots" by Jean Pruvost:**	**511**	**6**

April

Type	Title	Imp	Eng
General	eLexicography between Digital Humanities & Artificial Intelligence	2609	42
General	Linked Open Data in Linguistics	2046	38
General	Update from #travelgrant winner @asil_cetinElexis workshop	2044	84
News Feed	**Announcement Newsfeed**	**609**	**12**
News Feed	**Cinq mots a ressusciter de toute urgence**	**403**	**12**
News Feed	**Leopoldo Lugones, la lengua castellana de América**	**285**	**3**

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

May

Type	Title	Imp	Eng
General	Extended deadline until 6 May 2019 for the #DH2019 pre-conference	3087	24
General	6th Edition of Lisbon Summer School	2218	34
General	3rd summer Datathon in linguistic linked open data	1893	34
News Feed	**Iox' word that hasn't changed in sound or meaning**	**455**	**2**
News Feed	**Make the world a better place by using theses rarely heard words**	**364**	**5**

June

Type	Title	Imp	Eng
General	Blacks law dictionary, T. Reuters	2982	26
General	live stream to @IztokKosem's keynote 'Responsive Dictionaries'	2757	30
General	Job Klaxon: Apply for Data Analyst	2001	17
News Feed	**What do they call french toast in france?**	**764**	**35**
News Feed	**Make the world a better place by using theses rarely heard words**	**457**	**3**
News Feed	**RAE Dissionario de la lengua de signos**	**216**	**2**

Table 2. Perfomance of News Feed Twitter posts compared to the top 3 performing posts on the ELEXIS Twitter profile.

2.4.2 User feedback

Edward Finnegan
Professor of Linguistics and Law, Emeritus, USC Gould, USA

Let me characterize a small part of my engagement with the Elexis News Feed and illustrate ways in which I find ENF unique and extremely valuable.

I am a professor of linguistics and law (emeritus and fully engaged) and the chief editor of the journal of the Dictionary Society of North America, the oldest major lexicographical journal in the world. I am also an active linguistics textbook writer and a consultant to law firms litigating linguistic matters, including intellectual property such as trademarks. Given that breadth of engagement, no matter how many journals, newspapers, and magazines I subscribed to, I could not adequately cover the news in the arenas I must know about. While I am more familiar with English-language materials, I can glean enough from articles in French, Spanish, and German to know whether they are worth greater attention. There is no better way than ENF for me to stay in touch with a broad array of topics and venues across many languages on matters of concern to me. ENF is something of a lifeline to what is

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

going on in arenas that, as a practical matter, I could not otherwise know about, and as my familiarity with ENF arises from my engagement with the ELEXIS project, that is one way for me in which ELEXIS is invaluable.

One example. Because regarding language matters and linguistics from the perspective of business or finance is not of direct relevance to my work, I do not subscribe to business publications. Fortunately, at a glance, ENF keeps me informed. A recent ENF link to Knowledge @Wharton (the University of Pennsylvania's business school website) brought to my attention a US Supreme Court decision about attempts by a clothing manufacturer to trademark the term FUCT. Both as a textbook writer and as a lexicographical consultant, I need to know about that decision at the earliest possible time. I would of course have learned of that decision eventually. ENF gave me a timely alert.

Another example. The British Internet media company and website UNILAD has 60 million followers. It carried an article titled "The Correct Way to Pronounce 'Scone' Is 'Scone'." The pronunciation of "scone" is of little significance to me, but how writers on popular websites and in newspapers discuss whether words have correct pronunciations and, when they do, how those pronunciations are determined is significant. As editor and textbook writer, every day I face the challenge of making technical knowledge accessible to wide audiences. The "scone" article includes prototype maps of the Scottish Atlas of Variation, which showed graphically how the pronunciation of "scone" varies systematically from the north of Scotland to the south of England and from Northern Ireland through Ireland. That graphic is engaging and compelling—not merely about the pronunciation of "scone" but about how language varies and what is deemed common or correct in one place differs from what is deemed common or correct elsewhere. Further, the map shows that such variation is by no means random. The relevance to lexicography lies in a statement about the beta version of the Atlas (whose focus, incidentally, is not language but health care): "we are continually working to make it easier for you to access, understand and use our statistics." As editor of a lexicography journal, as a university instructor, and as someone committed to making the best scholarship available to the widest audience, I often need to do exactly that. Among many other virtues, the Elexis New Feed exposes me to ways of doing that.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

3 ELEXIS infrastructure as an ecosystem of interconnected elements

ELEXIS tools and services will form an ecosystem of interoperable elements allowing users to seamlessly switch between apps. To achieve this, existing interoperability solutions offered through CLARIN/DARIAH infrastructures, such as federated login and depositing services at CLARIN centres, will be incorporated or, where needed, new functionalities will be developed.

While initial efforts since the start of the project have been focused in the development of individual applications (with the plan being that they would be integrated with each other in the later stages of the project), the first steps towards interoperability have already been taken:

- Federated SSO login has been integrated into Sketch Engine allowing users to login using their institutional credentials and users can also use their Sketch Engine account to login into Lexonomy.
- Lexonomy and Sketch Engine are integrated with one another via the push/pull model, where the *push* model means that users can »push« corpus data from Sketch Engine into Lexonomy and get, for example, a list of headword candidates which can serve as a starting point for their dictionary, and the *pull* model means that users can, for example, check the usage of a particular word directly in the corpus data without having to manually switch from Lexonomy into Sketch Engine.

Additional interoperability features will be added by the end of the project.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

4 Conclusion

This report contains an overall assessment of the tools and services offered by the ELEXIS project. Sketch Engine, as the most complex and mature product, is widely used among ELEXIS members and observers. As indicated above in Figure 2, the number of users is constantly rising since the start of the project. Lexonomy is currently offered to several large institutions and is still under development, while the other two services, Elexifinder and News Feed, are useful for lexicographers who want to stay abreast of the latest news and research involving dictionaries or similar topics.

The user feedback collected in this report is overwhelmingly positive. Users appreciate the free-of-charge availability of the tools offered through ELEXIS contrasting it with the high costs associated with other commercially available tools (e.g., SDL Trados Studio). However, some of them also pointed out a few areas where improvement is needed.

As the most mature tool, Sketch Engine is used for a variety of tasks. Students at Université catholique de Louvain use it in specialized translation courses to manage their corpora and Sketch Engine is frequently utilized for master's dissertations. A similar usage pattern can also be observed at Università di Torino and the University of Surrey. Apart from students, Sketch Engine is also used by researchers in terminology, lexicography and translation studies. In terms of areas for improvement, Elisa Corino of Università di Torino said that the current method for parallel corpora creation is time consuming and that having a text aligner integrated into the system would be of great benefit.

Lexonomy is still under active development with regular hackatons taking place every few months bringing together developers and lexicographers for a few days of intensive planning and development sessions. In addition, it is currently deployed at several large academic institutions. At the Croatian Institute for Linguistic Studies, users found the interface very user-friendly and pointed out that Lexonomy allows users to work with XML data formats without without being familiar with XML. They are also taking advantage of the push and pull integration with Sketch Engine. On the other hand, they said that the thesaurus function could be improved by adding the ability to include thematic fields in the entry and making these fields searchable. Another user from the Center for Language Resources and Technologies at the University of Ljubljana reported that they particularly like the fact that they can actively contribute to the development and described how they are using specific features in the compilation of a Slovenian-Hungarian dictionary.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

User feedback about Elexifinder and ELEXIS News Feed is also positive. Robert Lew of the Adam Mickiewicz University in Poland believes Elexifinder has much potential if more papers are added into the database, while Edward Finnegan of the USC Gould School of Law said that the News Feed helps him stay in touch with a broad array of topics and venues across many languages.

Significant progress has also been made in terms of the interconnectedness of the ELEXIS ecosystem of tools and services. There is already tight integration between Sketch Engine and Lexonomy via the push (creating a Lexonomy dictionary from a corpus in Sketch Engine) and pull (accessing word information from a Sketch Engine corpus in Lexonomy) models. Additional integration functionalities will be developed in the coming months.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.