

To: "Ufal-I mail conference, local DL" <ufal-l@ufal.mff.cuni.cz>

Sent: Sunday, 31 December, 2023 12:34:39

Subject: Thanks for 2023 - and looking forward to 2024

Dear all:

After Bára sent the usual overview of ÚFAL's activities and successes in the past year (and it has been rich especially towards the end of the year – my congratulations to all who received grant awards for 2024!), I was not sure if I should continue the tradition and write to the whole of ÚFAL, or just to the LINDAT (sub)group. The main reason was that we are becoming so large that I admit I do not know all the activities of everybody (which I regret very much, but...), and I felt I might forget even people who did contribute to LINDAT itself. In the end, as you can see, I decided to write this letter again to everybody – if not only for the feeling that even if we are split not only between LINDAT and ÚFAL, but also to many other grants and even unfunded or volunteer projects, it is important to at least know about each other as much as possible – both for work-related but also social reasons. So, here is my summary for everybody, of course mostly from the LINDAT point of view.

In 2023, as you know, LINDAT/CLARIAH-CZ has started a new four-year funding cycle. There have been good news (addition of 4 partners from the Holocaust research area) and bad news – budget cuts. Nevertheless, we managed to go on, and get a number of good results, for which I would like to thank everyone. From the infrastructural perspective, two activities that ran throughout the year have been extremely important for everybody, even outside LINDAT: the continued participation of LINDAT team members in the preparation of the Czech EOSC consortium and project (for 2024-2028), where our repository has been included as one of the four repositories for Open Science nationwide, and therefore funds allocated for its future development; and second, the ongoing care of our computing cluster, mainly funded by LINDAT investments, by the UFAL/LINDAT IT team, who have been excellent and helpful as ever – including providing specifications and help for future planning of durable equipment for everybody, from students (AIC) to all researchers. My big thanks to these two small, but very important teams!

LINDAT is about data, tools and services. We have added a number of them throughout 2023 – as usual, two UD releases (and related datasets, such as CorefUD, and new UDPipe models), enhanced treebank annotation datasets (e.g., for discourse), evaluation and related training datasets (for information retrieval, entity linking, reading comprehension, speech, theater scripts, several for MT, and others), multilingual and other lexical resources, and some special software (such as new search tool for the multilingual SynSemClass ontology, model for PDT-like output for UDPipe). It is also good to see venturing into multimodal (video+text) data (for summarization – the MLASK dataset): this will certainly become more widely researched topic in the near future. Non-ÚFAL/LINDATers also deposited data to our repository, most notably by the Czech-U.S. collaboration project, UMR. For everybody involved in these data, tools and services development, my big thanks for continued effort to contribute not only to LINDAT, but to the researchers throughout the world who can then use them.

Many LINDAT projects have continued throughout 2023. On top of the core work on upgrading the repository and preparing it for Charles University central data repository instance and for the EOSC CZ project, the PDT-C analytic annotation and treebank correction project continues, as do the discourse, coreference, MT and lexical projects, many with input from the related research project(s). Many thanks – and please keep the good work all the way to the final data and software releases!

Lot of effort has been also spent on securing future funding. While some promises have been made by the Ministry of Education regarding the additional cuts (I mean, in terms of lowering them slightly for 2024), it is still not clear (even today!) whether it happens. We are also not sure if the accompanying grant for more durable equipment went through (should have started tomorrow). We are optimistic though about the EOSC CZ projects: even though the national repository project has not been submitted yet, we will start working on it right away, with the specific humanities (sub)project has yet to be written and start in 2025. Since EOSC CZ is also focused on data infrastructure, it will allow us to continue all current LINDAT data project even with lower funding, at least for 2024 and possibly beyond.

LINDAT/CLARIAH-CZ will be also helped by international projects, which LINDAT is part of or which it supports, that start or continue in 2024: the infrastructural ATRIUM and EVERSE projects, and the continuing CLS, HPLT, RESQ+ and MEMORISE projects. We also hope that LINDAT will be used (as the source or target for data and services) by other ÚFAL projects that either continue or are now starting, as I have mentioned earlier.

Given the current situation in government budget, we are entering a prolonged period of uneasy funding situation in research. In fact, in the EU, the “classical” funding calls we have used in the past are disappearing, while different ones are emerging, and we will have to look at those (as we are now doing at the national scene with the OP JAK calls, which we were trying to avoid in the past for their administrative difficulties). This brings me to another important group of our team members – even if some of them actually are not formally paid by LINDAT – namely, project managers and administrators. LINDAT, with its 15 partners, is a difficult beast to tame, and the project management and administration deserves big thumbs up for making it possible!

Given our project management and administrative support, as well as LINDAT support data-wise, there is hardly any excuse for anybody not to keep trying to secure additional funding for their research. Please do consider all possible and suitable levels – from GAUKs to the standard portfolio of calls from national and European funding agencies all the way to ERCs. You can do it!

Before I finish, let me also say a few words to the tragic school shooting that occurred just before Christmas at the Faculty of Arts (FF UK). FF UK is not only a significant LINDAT consortium member, but also a fellow faculty within the university, with whom many of us closely cooperate on projects or informally, teach there, are members of their committees (and vice versa). Some of those we knew or cooperated with personally lost their lives and some have been seriously wounded. Some of them have “just” been there, barricaded and under tremendous stress.

My sympathy and thoughts are with all of you who have lost someone you knew, and with everybody affected in any way at ÚFAL or at FF UK.

Many words have been said and written since then about the causes, handling, and the fallout of it; I do not want to add much more, since I am sure you can easily discern the reasonable opinions from the others. Perhaps one thing only: yes, it could have happened at MFF, elsewhere at UK, another university, or any other school or public place. But the worst thing that could happen now is to surrender to fear and act so – that would be an invitation for another such tragedy.

As Bára has already written, and on top of what the Charles University, MFF and FF managements do, we at the ÚFAL and LINDAT management level are ready to help everyone in whichever way we can. Please do support all that the university, MFF and ÚFAL is doing, and we will also be happy to accept your suggestions how to do more.

I wish you all a safe, healthy and less stressful 2024, laced with both personal and work-related successes, and I am looking forward to cooperating with all of you on LINDAT and other projects.

Jan Hajič

PI and coordinator, LINDAT/CLARIAH-CZ & ÚFAL MFF UK