Evergreen Conference

Localization is so Much More Than Just Translation

May 26, 2021

>> ROGAN HAMBY: All right it is 11 AM Eastern. Good morning, good afternoon, that evening, wherever you are -- we should have a link for closed captions let me try to find that. I will follow up with that just a second.

The Evergreen International Conference and Mobius are our sponsor. And we will have Eva Cerninakova and Linda Jansova talking to us about localization. I will now go quiet but I'm really looking forward to this.

>> LINDA JANSOVA: Good morning everyone. It is an honor to be here my name is Linda Jansova and together with my colleague Eva Cerninakova, we will be speaking about our experience with using Evergreen in a non-English speaking country.

First a little technical note. In case you wish to download the slides or read them in their original Google slides format so you can instantly click on the URLs mentioned. Please click on the URL shown. Eva will also copy this URL into the chat. Before we proceed further, let's have Outlook at two terms that describe today's topic. Internationalization and localization. Expressions i18n and L10n are often used with i18n representing the numbers and letters between the first and last term. In both cases, we speak about means of adapting computer software to different languages, regional picture regularities and technical requirements of a target locale.

While internationalization is a process of designing a software application so that it can be adapted to various languages and regions, localization is a process of adapting internationalized software for a specific region or language.

That is the theoretical background which we will use for today's talk. Internationalization and localization are rather complex topics, and we will be sharing your insights from our own experience, this means that this presentation can be considered a case study focusing on Evergreen and the Czech Republic.

It is not the first time we've been present at the Evergreen conference however, back in 2015, we contributed with our readings introducing all the members of our Czech micro community. Last year is a conference was held we took part as participants, this year marks the first conference which we have prepared a full presentation, not everyone may be familiar enough with our geographic location and we've included a map showing our country, the Czech Republic is situated in the central European region.

Our country is also known as Czechia, it is small having less than 11 million inhabitants in an area of less than 80,000 km. We have a beautiful capital called Prague, now you can enjoy the view of the Prague Castle overlooking a number of other historic monuments. Actually, one of them located across the river is the seat of the library that I work for during the day.

The last picture from Prague which has been carefully selected to confuse you little bit, you can see something which looks like the Eiffel Tower but in fact it is a lookout tower in Prague that's been inspired by the Parisian landmark. Enough of geography and history for now. If you take a closer look at the landscape you might find out that we probably have the densest library network in the world. We have more than 5000 libraries, which are in the ministry of culture's list. Huge number of libraries stems from the fact that in 1919, our first library law was passed in this law made it compulsory for all municipalities to have a library.

Fast forward to the 21st century know, quite a number of these libraries, although they have different founding bodies, collaborate on various nationwide projects. Mostly seeking efficiencies or ways of improving services to their patrons. It is good to always consider joining such projects because libraries are stronger together.

There is a number of such national projects in our country come out of these we will briefly touch on four of them. Let's get started with the first project, Obalkyknih.cx and edit content provider focused on materials available in the trend 21 libraries started back in 2008 is a book chapter provider and over the years more types of edit content such as tables of contents, summaries or user ratings have been included. New can concurrently it contains about 2.3 million book jackets and almost 600,000 tables of contents. It is a service which is very popular with the Czech libraries.

When it comes to Evergreen, we actually implemented the service twice, for the first time in 2012 and the second time in 2016. Obviously both Evergreen and the edit content provider change over time and so it is rather likely that no solution can be considered permanent, after having created the current module, we monitored to master’s in early 2017.

This screen shows evidence that we have indeed done so. It shows that our developer has committed the module. When the module is switched on, the book jackets can be seen in the usual Evergreen Way showing both in the surface display and on the webpages which individual records. The image shows the sample record. If other edit content data are available and switched on, patrons can access them the lower part of the page. The second project I mentioned is a project dealing with Authority Records. In that Czech public about 70 libraries collaborate in the creation of these records and this started about 20 years ago and the key technology used for sharing the data is still the same protocol. Z39.50. The result saying records are provided by the national library.

What is significant is the fact that the Czech libraries are used to clean bibliographic and authority data themselves without external companies involved. The usual procedure of the library catalog consists of downloading the record in question from the national library via the 3950 protocol, however Evergreen currently saves all authority records downloaded via this protocol as bibliographic records.

One more note, although we are a small country, the number of authority records is rather high. Currently about 1 million records are in the file, in comparison the Library of Congress provides about 11 million authority records.

Therefore, using the whole authority file can be rather difficult due to its enormous size. It goes without saying that in practice, most libraries use just a fraction of the total number.

Back in 2017, we carried out a project to help Evergreen libraries use the Czech authority records, given the authority records sides and the need to have the records directed to the database we came up with the following work around grade we selected only the authority records which have their identifiers presented in the bibliographic records, and the algorithms cover cases when no authority identifiers were included in the record.

We prepared documentation available via our documentation portal Evergreen docu wiki.cz.

the union catalog of the Czech Republic is another important project, since 1995 it was made available in the electronic format. More than 500 libraries cooperate meaning that they contribute with their data. The catalog contains about 8 million bibliographic records. The preferred way of contributing is via the open archives initiative protocol for metadata harvesting, OAI-PMH, sending them to the server via the FTP, file transfer protocol is also an option.

In 2018, we carried out a project aiming at checking that Evergreen libraries contribute to the union catalog via the OAI-PMH protocol.

The code is not in the Evergreen master yet. It offers plenty useful features but unfortunately at the option to organize the data into sets. Therefore instead of using sets we opted for adding a value in them MARC 21 Bili graphic record field to indicate that the record should be sent to the union catalog. We are an extensible stylesheet language transformation and we have the data so we send only the bibliographic records quickly marked for inclusion. For those not familiar with how the OAI-PMH delivery works, if included an example. Basically it is similar to Evergreen's super greater feature.

Of course there are some costs involved in localization. In terms of integrating Evergreen libraries to the Czech library landscape now it is a great opportunity to say aloud thank you to the ministry of culture of the Czech Republic, it is kindly provided financial support for the aforementioned project via its grand scheme for libraries. The projects have been prepared, submitted and coordinated by Jabok Library and Jan Langos library.

We would like to start collaboration with Knihovny.cz providing unified access to the check libraries and their holdings. And now let us have a look at some of the other localization issues we have organized them into eight categories which are statistical reports, templates for printing and cataloging, accession numbers, check currency, cataloging interface, UTF-8 encoding, sorting, and deployment.

We will start with the statistic reports. Czech libraries are provided to provide statistical data about their activities to authorities. Czech authorities require first data which can be easily abstracted from Evergreen and second, data which are rather difficult to put together. One example of the latter can be patron account registration date needed to report how many users have registered. using customized reports can be a solution or a partial solution.

In order to avoid complex reports or SQL queries we use more customized reports to put that required set of data together and the screenshots show two Evergreen reports providing information on in-house use on circulations all designed to get data from the statistical form which you see on the left.

Local customization is also necessary in templates for printing. These are localized into Czech.

When we speak about templates, these are not used just for printing but also for cataloging. In this case we can mention for example 3 defined language, Czech monographs because more direct predefined values saves time of the cataloger.

Now let us move on to accession numbers. Assigning accession numbers is mandatory for Czech libraries. Nowadays some libraries incorporate accession numbers into barcodes as seen on the left image. Some do not and for example, the call numbers under the barcodes of course I mean the barcodes in the physical format now.

Also, not every material should be assigned an accession number. For example, individual serial issues should not have an accession number, but once the issues are bound together, the resulting volume should be seen.

Another difference between our country and the US is currency. We do not use dollars. Although the word dollar is said to be of Czech argent but we use Czech Crowns, these are abbreviated as CZK internationally and Kc, in the check environment.

When it comes to value one US dollar can currently be exchanged to about 20 CZK and also the place and expressions. While US dollars typically used before the amount we use the abbreviation after the amount.

In the billing interface in Evergreen, there are some places where the Czech Crowns can be easily hardcoded. The USD is replaced and put in the correct place after the amount. Some places are more difficult to customize. You see in the screenshot in the lower part of the side you can see the US dollars. It goes without saying that these interface inconsistencies can be confusing to library staff.

Now let us move on to cataloging interface and other challenges for library spread while in the US probably more lepers focus on copy cataloging, it Czech lepers often have to do original cataloging when the original bibliographic records not available for download. Evergreen cataloging interface makes heavy use of MARC 21 text such as 100 while check libraries stuffer appreciate strings and in this case it is main entry personal name. They expect the strings to be in Czech.

context-sensitive help also tends to show up untranslated as can be seen in the screenshot from the cataloging interface. Czech language uses these marks. Evergreen uses UTF-8 encoding but there are places where localization tends to be rather difficult. We have learned that library org units short names are much safer in ASCII only characters for example. Also to make sure that our special letters do not get corrupted along the way, we use HTML entities to correctly display characters from our local added content provider. currently we don't use serve terms highlighting feature because letters with diacritical remarks do not get properly highlighted and it would be confusing for patrons who would see some of the search terms highlighted and some not.

When the Evergreen library serves as a Z 3950 target it may be difficult to search with diacritics.

Another area where localization can play is sorting. This looks about the same as the English one but it is not quite the same. Getting back to the example reversely, this is not supposed to be equal to C, there are examples catalog search results and from browsing showing the correct order according to the Czech alphabet is not observed. Another case when alphabetical order is not correct is interfaced when all strings with diacritical marks are at the end of the list which is not where they are supposed to be using the Czech alphabet.

We have mentioned a number of areas where there are tweaks to better localize Evergreen are needed and can be difficult to maintain across multiple instances of Evergreen. In 2015 it became easier though, as we started to move from Multiple plantings to single consortium.

This consortium is called common catalog also abbreviated as SPOK and now it has eight member libraries.

Jan Langos Library this is the homepage of the library catalog which uses different graphics. Personal library is a different type of consortium were the previous examples were libraries, the personal library is a consortium of librarians. It serves as a common installation for individuals who wish to catalog their home libraries.

This is how the personal library catalogs looks.

There are differences in how libraries handle certain tasks depending on local conditions and also how particular software tools work. It is clear that software from abroad can serve as a source of inspiration. It is true that sometimes things can indeed be done differently.

>> ROGAN HAMBY: Jane has a question about the OAIPMH would the union catalog harvest certain sets of records?

>> TERRAN MCANNA: Yes. It is like that because there are some documents or bibliographic records which should not be sent to the union catalog. In libraries there can be records which are not finalized etc. Libraries need to mark somehow which records should be sent to the catalog and which not.

>> ROGAN HAMBY: Okay.

>> EVA CERNINAKOVA: Just two sets because we prefer having also sets for individual organization units because the records are harvested together from all libraries from the consortium, and it is complicated for the staff in the library who take for harvesting.

>> LINDA JANSOVA: And also, the OAI - PMH protocol can be used some projects we would like to participate in the future so then you have to have a different URL for harvesting which makes it a bit more difficult than it would be if the sets feature was present in the module.

>> ROGAN HAMBY: Another quick question from Sharon, what brought on the desire for the personal library function?

>> LINDA JANSOVA: Once more please?

>> ROGAN HAMBY: What was the reason for the personal library project? How did that start?

>> LINDA JANSOVA: Yes, we decided to start a project because we thought that if librarians had a chance to see Evergreen in action, and to try to use it as individuals and maybe it would persuade them to switch to Evergreen from their proprietary ILS that they are using. We thought it would be a nice way to get acquainted with Evergreen and to really see what it can, what it can do and that it is a very good software which can be used in multiple types of libraries and which can serve as system or software for the whole consortium.

>> ROGAN HAMBY: We have a few other questions but I think that would make sense at the end so we will let you get back to your slides.

>> LINDA JANSOVA: Okay. I will just introduce a topical little bit. It was back in 2008 when we started to translate Evergreen during an elective course at the Institute of information studies and librarianship at Charles University in Prague, we used Moodle at that time and disseminated the files for translation and we had an e-learning form and the results of early efforts were shared with a wider audience at a library conference in early 2009.

Back in 2009, we also encountered first issues with implementing translations as can be seen in this message sent to the open ILS general listserv.

This clearly shows that a vast majority of Evergreen is translated into Czech especially thanks to Eva who is also the global coordinator of Evergreen translations.

When we speak about launchpad I would like to mention that currently in order to find localization it is useful to search for i18n tags. Eva will provide more details on the translations but the translation is what tells the product so over to you, Eva.

>> EVA CERNINAKOVA: Localization is much more than translation. However, for use it is essential to be able to use the catalog function in the native language. I personally also believe that the catalog interface available in various languages makes libraries of more inclusive and more accessible to everyone.

In this part of the presentation, I will try to very briefly introduce how the Evergreen translation works, what translation looks like. What tools are used and what issues translators need to address.

Usually, we need Evergreen catalog through its interface. Here you can see an example of a small piece of the Evergreen interface specifically the action drop-down menu. And actions like suspend, activate etc. -- It also with the context sensitive help but appears when you hover the mouse over the drop-down menu.

Even if it doesn't look like that, there is the same interface in this slide. However in this case it is a tiny portion of the source code. You can see several things from the drop-down menu like action for selected holds. Suspend etc. As you can see the text in the source code are just in English. But of course, we need those things to be translated into Czech Spanish, French and other languages. So that is why it is necessary to mark the strings for translation in the source code. Which is usually done by the code developer.

In this slide you can see the same source code, but in this case, the strings are marked for translation for each string is enclosed in a sequence of characters like brackets, percent sign etc., that tells the Evergreen interface that there is a piece of code that is translatable and that should be displayed as translated into the selected language.

One and new Evergreen release is prepared, strings marked for translations are extracted from the source code into the so called pod files. Portable object implied. Each POD file represents a message ID. And to each message ID should be assigned a message string which is a translation. Message ideas and message strings are indicated by a prefix msgid and msgstr. the message strings are still empty. This is all right because POD files are the basis for translations. The translations are usually not typed directly into the POD files usually and an online tool or application is used for translations. And for Evergreen translations we use online tools Launchpad and PO editor. So when the strings are extracted from the source code to POD files, the POD files are uploaded to Launchpad and PO Editor and they are ready to be translated --translated into individual languages.

When the new release of Evergreen is approaching, files with the translated strings are downloaded from Launchpad and PO Editor and edited in Evergreen as PO files. PO files contain the original string in the translation. More precisely, message IDs and message strings.

Let's have a short look at Launchpad. The edger green project translations are available when you click on the translation on the top menu bar. If you want to actively translate and not only see the translations, it is necessary to have user account in Launchpad be locked in. That you can see the Evergreen project translation in Launchpad with basic information about Evergreen translation. In the list of translation for that master, the list of various languages.

The language name link opens a list of individual translation templates for the particular language. The name of the template, then it works is a link opening the interface for translation. Useful is also the number of untranslated strings that opens the interface for translation but with only untranslated strings displayed.

This is the translation interface itself. When the translation on the pages are completed, the translations are saved by clicking the save and continue button.

There are several useful functions in Launchpad and I will mention for example such as generated from translation used for the same word or phrase in Evergreen, or even in other software projects available in Launchpad. The possibility to copy the entire phrase into the translation box using the wheel with the blue arrow next to the English stream. There are of course many other tools including searching and all translated or untranslated strings etc.

The other tool POEditor. POEditor.com/projects. your account must be approved by the Evergreen admin. It is possible to ask for approval on the Evergreen General mailing list.

There are several useful tools available in POEditor for example to search or display strings or only untranslated strings. In the POEditor there's a possibility to select strings by tags by Evergreen editors and I find this useful it is possible to display strings for the Evergreen release, and POEditor are translations are saved automatically so it's not necessary to save them.

When translation in Launchpad or POEditor, when translating and Launchpad or POEditor it is good to know that in some cases the part of string in Launchpad or PoEditor should not be translated. Data placeholders and variable names, at an HTML/XML, this contains specific characters so when you have these percent signs and brackets etc. you could be unaligned. If you are not sure what to translate you can ask the question on the Evergreen mailing list. You can see examples from Launchpad 5 there's tags and variables at should not be translated that are marked yellow. You can see angular strings in POEditor, the angular text which are not intended for translations.

In order to translate Evergreen, you need more than translation tools. There are many challenges to be faced when translating. There are many synonyms and multiple meanings --also some strings in Evergreen only appear after a specific action is taken etc. is taken etc. For appropriate translation it is essential to understand the background of the strength, this means Evergreen software and its processes, the library terminology, the ways libraries work in the context of local conditions and much much more.

There are many ways to determine context of the particular strengths, useful is the Evergreen server itself. For testing you can use your own installation of Evergreen or Evergreen demo servers.

Useful could be the information and Launchpad or POEditor in which part the software string is located. In the example from Launchpad on the side you can see below the translation box that the strings check, it is located in the files with the name containing the billing and payment so it is clear that the meaning of the string has something to do with billing and payments and not for example with monitoring something.

You can find useful informational and comments are context information. An example from Launchpad you can see the variable used in the strength of this case the percent sign 1 represents an organization unit name.

Some information about the context of particular strings can be found in Evergreen source code, you can find the strength it's possible to search the GitHub pot repository using grep set or browsing but the browsing requires some basic structure of Evergreen. It is possible to search downloaded source code files using the tools like Total Commander or Midnight Commander.

I want to show you a few examples illustrating well how the context is important -- and how it can be misleading translations. Most of the translations are rated by students who translate Evergreen at the beginning. Just because they didn't have necessary knowledge of the context. So copy was translated in the meaning of reproduce. The phrase copy location order used in Evergreen for sorting we in locations was translated make a copy of location voucher. Or place hold was translated as place the handle. Etc. You can see such translation.

important also is the terminology, terms used in the library --in other languages may be based on completely different concept so it is important not to translate literally but to use terminology, to the language in the given country. In Czechia we have a database of library terminology useful for translation paid I recommend using it.

It is important to note that Evergreen standardized terms are used. In MARC format, such as field and subfield names or value list for codes. in Evergreen we use also terms from SQL protocol and formatting names and names of countries under languages etc. For translation of such terms the form should always be used.

It is important that the translation is not much longer than the original strings because long string could cause issues in Evergreen interface. Which applies only to translations. I recommend using as short expression as possible, the short string in the interface is always easier than displaying the long one. However sometimes the translation is simply longer and there's nothing to do about it. I just unnecessary to be creative and in the side you can see an example of search results sorting but the Czech translation did not fit. Resolve this by arranging the words and phrases and omitting their repeating word.

Initially when it is necessary to use standardized terms you can see that in a few examples of standardized terms from the MARC code list, the length of translation can be almost 3 times longer than the original string. Also abbreviations may be used differently in various languages. For instance, in Czech some expressions do not have standard abbreviations such as date of birth abbreviated as DOB in English or abbreviation for fixed field used in MARC grid and Evergreen cataloging interface and appreciation for a literary form, data status etc.

when the phrase no item with barcode, it is important that the word "barcode" is placed at the end of the sentence because it is followed by the barcode of specific item. But in Czech, placing the translation of the word barcode at the end of the sentence is not ideal. In such a case, you may opt for a less ideal translation or better phrase.

to summarize, basic reclamation for translators it would be something like this. Translations must be understandable, it is necessary to adapt them to the language in local context. It is necessary to be consistent in used expressions. Using different terms for the same thing is confusing for both patrons and staff. It is necessary to understand what is translated. It is necessary to leave the string untranslated until you are absolutely sure about the meaning of the string. Misleading translations is worse than no translation.

It is good to read the translation again, typing ever can happen to anyone. And it is really important to be as brief as possible untranslated strings should be as short as possible. You can find more details about translation workflow and some more recommendations and Evergreen.

I am at the end but however I will mention one more topic, sometimes you need to translate customized links specific for your library or consortium. If the strings are not available in Evergreen software in general. You can see an example of customized template. Particularly it is customized homepage of the Jabok library catalog containing many links to searches for Jabok students and you can see the pages available both in Czech and English.

If the string is supposed to be translated, it must be marked for translation in the source code file just like any other translation string in the source code. In this case, we see the OPAC template. And you can see customized names of lists have been marked for translation. Then it is still necessary to but the translation to a PO file containing the translated strings in your desired language. Our consortium uncommon catalog I usually add the translation to the PO file manually and it is good practice to comment the edited translation for example, in the library under which part of the template etc. Is being used.

I think it would be possible to talk about translation for a few more hours. But I will finish at this point as we do not have so much time. If you want to see what the translations of Evergreen into Czech looks like, you can use the Czech demo installation of Evergreen the link is also available on Evergreen docu wiki.

>> ROGAN HAMBY: We have a few questions in the chat. I will ask these but hopefully we can move along quickly. One is from Andrea, in regards to the currency designators not showing correctly for localization do you know if that is consistent in all the grids?

>> EVA CERNINAKOVA: I am sorry, I did not catch the first part of the question.

>> ROGAN HAMBY: Can you change, the inability to change the currency designators to a local, is that broken across all the grids are only a problem in some?

>> EVA CERNINAKOVA: It is a problem in all of the grids. We try to set the angular -- but there was a problem that it showed Czk, which was not desirable for us. So we decided to do it like this. .

>> ROGAN HAMBY: Hopefully that's an underlying angular thing that can be fixed in a single place. Another question was for Diane, how many Evergreen consortia are there in the Czech Republic?

>> LINDA JANSOVA: We have three, the biggest is the common catalog which is about at a library. Involved or participating in the there's a smaller consortium of two libraries. and then we have the personal library consortium of librarians where we have a couple of librarians in part.

>> EVA CERNINAKOVA: I would add that Czech Republic having a consortia of independent library -- having such consortia in that Czech Republic.

>> ROGAN HAMBY: We have one last question and this is potentially too bigger question to answer right now but Diane asked, what is it about Evergreen that made it a better option to translate it and do something that is already available in Czech are there options in Czech that you thought were viable or did you have to do translating no matter what?

>> LINDA JANSOVA: I can answer that question. Actually, we started being involved in Evergreen and translation before, more than a decade ago when there were no other open source ILS available in the Czech Republic we decided to change the situation --and if you have look at that number of libraries that use Evergreen, it is not exactly hundreds or thousands but some have been courageous enough to go for open source and it has persuaded libraries not to go to open source integrated library systems, but be more interested in other open source and now I think far more libraries use open source in their daily work that they did a decade or so ago.

>> ROGAN HAMBY: Thank you for that. People are welcome to continue talking in the chat and I want to thank you for the presentation, you certainly made Evergreen stronger in your community.

>> EVA CERNINAKOVA: It is a pleasure to be with you.

>> LINDA JANSOVA: Thank you, Rogan.