

EE Podcast #49 Multi-Language Sites in EE

Lea Alcantara: This is the ExpressionEngine Podcast Episode #49 where we talk about multi-language sites with our guest, Tom Jaeger. I'm your host, Lea Alcantara, and I'm joined by my co-host, Emily Lewis. This episode is sponsored by EECI 2011. EECI is up for its 5th season and this time it's returning to the United States of America the most significant conference where ExpressionEngine developers, designers and users will run from October 19th to the 21st at the Invincible Dog in Brooklyn, New York. A few tickets are still available, so check out EECIConf.com for more details.

Emily Lewis: The ExpressionEngine Podcast would also like to thank Pixel & Tonic for being our major sponsor of the year.

Lea Alcantara: Awesome, so today, a very exciting topic. When I mentioned this on Twitter, everyone just started bringing in their two cents here, and we are talking about designing and developing multi-language sites. So to help us navigate this huge topic, we have Tom Jaeger of EE Harbor. Welcome, Tom.

Tom Jaeger: Hey, thanks for having me on.

Lea Alcantara: Thanks for coming by. So yeah, multi-language sites, there are so many ways I think one could tackle this type of project. Emily, have you ever done a multi-language site?

Emily Lewis: Well, I worked for an employer that had multi-language sites and my responsibility was mostly from a front end perspective, evaluating some of the design considerations and some of the more social- cultural things that come into play when you are dealing with multi-language sites.

Lea Alcantara: So what kind of things would you say those, like, what are those cultural things that come into play?

Emily Lewis: Well, I'm not a visual designer, but some of the things I had to do research on were some things as simple as like color. Different colors can have different impacts in different cultures. Obviously, you don't rely too much on color to drive a person to action because that's usability or accessibility no-no. But you also need to be careful about colors that may dissuade users from doing something due to a cultural influence.

But then there are also things that from a front end perspective, not really on how the user experiences the site, but when I'm working in the HTML, it's just little things that you need to make sure you do like setting your language, making sure using UTF-8, and also just different areas of pages could also have different types of language within them. So it's not necessarily that you specify the language for the entire page, you can also specify language for different snippets of content on a page. So it's a huge area. I mean, it's just a few things that you could take into consideration as a front end developer.

Lea Alcantara: Well, that's pretty interesting because as a Canadian, we have official languages, which is English and French, so if you run a government website, it was interesting that you pointed out that not all multi-language sites have to be one language for that one section and another language for another section. For example, the government of Canada's website needs to have both English and French in one section or at least some French to help and guide someone to the French section of the page, so it's interesting to note that.

Tom, did you have similar findings when you first started researching and creating multi-language sites?

Tom Jaeger: The vast majority of multi-language sites that I've worked on have been a full site in a given language, not so much where parts of a page or parts of a certain section are in multiple languages at a given part, if you will.

Emily Lewis: How many multi-language sites have you worked with, Tom?

Tom Jaeger: Somewhere between 10 and 12 multi-language sites.

Emily Lewis: What kind of sites were they for? Were they for events or organizations?

Tom Jaeger: Yeah, the vast majority of them were for a large company creating their multi-language sites to be used in different nations. I also did some work for Sony Music with Steven Hambo from Hambo Development.

Emily Lewis: Where was the Sony Music site? I'm actually interested. I know that's a huge multi-national corporation. What kind of requirements was needed for that site because that's just a large corporation?

Tom Jaeger: The Sony Music site, it was actually a particular artist site. It was a very interesting one. Sony Music goes about developing multi-language sites with a kind of mix of traditional multilingual methods in ExpressionEngine as well as some homegrown solutions that they have. That was also a very interesting one because actually we took the site, and it's got a whole online community. Offline, we made a development snapshot. They continued to use the site live. I actually edited web blog fields and things like that while it was live and then we had to merge the two sites and the content back together at the end.

Emily Lewis: Wow. That sounds not fun.

Tom Jaeger: You are not kidding.

Lea Alcantara: So how long did something like that take? Do you find that because you work with multi-language sites that it has literally doubled the work?

Tom Jaeger: I don't think it necessarily doubled the work of a traditional site. The Sony Music, that one was actually a very quick turnaround time. We've launched a whole new design for that. The data merge took, I don't know, maybe 16 hours or something like that to actually kind of go and line up again and make sure the data wasn't lost and new fields were there across the board, the ones they added and the ones we added.

Emily Lewis: Now, the process for the other multilingual sites that you work on, but not Sony. I happen to know that was for Pitney Bowes Business Insight, right?

Tom Jaeger: Yes, that's correct.

Emily Lewis: So I actually used to work for them, so I'm sort of familiar with how the process was in terms of establishing the need for the international sites, but they were basically like they are public facing like marketing style site that then had to be translated into all the different languages for their different offices in different countries. Now, if I recall correctly, it's been a while since I worked for them, but the visual front end remained pretty much the same, but the bulk of the work was to translate not only the language but the ExpressionEngine templates and everything else. Can you talk about the process of working with those sites?

Tom Jaeger: Sure, absolutely. So when we went about those sites, you kind of covered the background pretty well, I think, but we basically went with a multi-site manager approach to creating these sites, largely because different staffs would be dealing with the content updates for each site according to their country. Along with that, we did use the Pages Module, which represents a few difficulties in itself. As you mentioned, there is the template portion to address in those sites as well as the content portion.

Pitney Bowes has a great staff, I mean, obviously, come on. And then that helped a lot right there. I think that combined with people in those countries speaking those languages really made things very smooth on a development end. So when we were dealing with these sites, which were rather large, a couple of hundred pages each, we started by looking at hard-coated template items, whether that's URLs, whether that's words, phrases, paragraphs, things like that, and getting those addressed right off the bat. There is some turnaround time needed for the actual translation there.

For that, we started looking at URL segments, particularly. URL segments, categories, category URLs, URL titles. In ExpressionEngine, you have certain limitations with each of those fields. Some special characters are allowed in the Pages Module with URL field that aren't allowed in the template field or maybe not allowed in the URL title field for a given entry. So we kind of started looking at those requirements, "Okay, you know, which fields is for which characters and can we standardize this to like a universal allowed characters, if you will." And we kind of established a procedure for that going forward. We got the content translated. We moved the templates over to the appropriate names and removing special characters when needed. It's the same with category URLs, things along those lines, and we essentially got everything translated and we did a massive, massive template overhaul on the site.

Emily Lewis: Now, is that where your Transcribe module came into play and evolved? Because I was talking to Ian Pitts, who was the web team manager on the project, and he was telling me about how Transcribe kind of really made the heavy lifting a lot easier and streamlined the process.

Tom Jaeger: Yeah, that's exactly right. Transcribe kind of emerged from conversations with Ian. Now, Ian is a very sharp EE developer and we were kind of talking about going about these things and these different character sets, with the hard-coated entries, the templates, the URL segments and things like that. How can we streamline this in a sense, and that's largely how Transcribe was born. We kind of theorize the system with taking these hard-coated items and essentially

creating these variables in the back end, if ever, they are easy to manage, update, be able to do imports and things like that to sort of be able to massively cut down the time needed to roll out new languages by creating new translations of these variables and phrases and then be able to just kind of use our module to flip the language if you will.

Lea Alcantara: So Transcribe available for EE2?

Tom Jaeger: Transcribe is not available for EE2 right now. It is in development, though.

Lea Alcantara: Okay, do you have any ETA for us?

Tom Jaeger: I would say the ETA will be at the end of the month here.

Lea Alcantara: Oh.

Emily Lewis: Oh.

Lea Alcantara: That's coming up. So you talked a little bit about the planning and organization part of it. What were the challenges when you are planning the implementation of the translated languages, and were there any tactical or client-related challenges in bringing multi-language sites into EE?

Tom Jaeger: I will start with the client because that's an easy one. Particularly working with Pitney Bowes, they were a phenomenal staff. They are very competent. They knew what they are doing. They have their act together. So from that angle, no, we didn't. I think lessons we quickly learned were the URL issues and the special characters are contained or not contained within them. Another quick lesson we learned obviously is that automated translations with Google Translate or things like that, they just really don't cut it.

Lea Alcantara: Yeah.

Emily Lewis: Are they just not accurate enough in terms of their translation?

Tom Jaeger: That is the opinion I would get. Every now and then, we would have a little snippet of content and we post it up and say, "Hey, you know, could we get this translated?" And somebody from their Spain location would come on and review the content we translated or we would review what they put in against Google Translate and it's often vastly different. So I do think it is kind of some little locality type things are different between Spain Spanish and Central American Spanish and things like that.

Emily Lewis: Right.

Tom Jaeger: With different dialect, I suppose.

Lea Alcantara: So...

Tom Jaeger: I'm sorry, go ahead.

Lea Alcantara: I just wanted to clarify that most of the multi-language sites that you've worked on have been for Latin-based languages?

Tom Jaeger: No, not completely.

Lea Alcantara: Okay.

Tom Jaeger: We've done a lot of them. We've done Portuguese, Finnish, Danish, Spanish, Latin American Spanish, as well as Spain Spanish. I don't even know how many we are in. The Sony build, there is a ton in there.

Lea Alcantara: Did you end up ever having to work with a language that you say non-Roman based lettering, let's say...

Emily Lewis: Kanji.

Lea Alcantara: Like, yeah, Kanji, Chinese or like Chinese, Kanji, Japanese, Hiragana or Indian Sanskrit, et cetera.

Tom Jaeger: I actually haven't had to work with any 2-byte character sets or anything like that, with characters that read right to left or things like that. My understanding is that there is some limitations within ExpressionEngine when you start dealing with 2-byte character sets particularly.

Lea Alcantara: What kind of challenges are those?

Tom Jaeger: Yeah, so having that hit them, my understanding is some of the way the titles are set up.

Lea Alcantara: Yeah.

Tom Jaeger: The way the database titles are set up, you have to actually tweak that field in the database. I'm guessing there are other things along those lines. It's kind of thing that I'm sure somebody who has hit it could explain a lot better.

Lea Alcantara: Okay.

Emily Lewis: Well, actually, we are planning on having someone. We will be revisiting this topic of multilingual sites in a future episode focusing on those 2-byte character set based sites.

Tom Jaeger: I mean, that would be very valuable.

Lea Alcantara: So during your journey in multi-language sites with ExpressionEngine, did you use any other add-ons or tutorials, or what helped you get these sites out beyond, say, Transcribe?

Tom Jaeger: Yep, the Pages Module, which is what we are using. The site was originally created with prior to Structure's existence.

Lea Alcantara: Yeah.

Tom Jaeger: So the Pages Module, that represents some difficulties in itself when you replicate a site. The page's URL structure does not copy over.

Lea Alcantara: Interesting.

Tom Jaeger: So you kind of just have a site with thousands of entries sitting there.

Lea Alcantara: So should that be a feature request if we are going to be making MSM sites that page structure should also be replicated?

Tom Jaeger: I think it would be nice. A lot of it comes back to how the sites are being built. Again, going back to Pitney Bowes, even across their sites, they don't necessarily have the same site architecture.

Lea Alcantara: Okay.

Tom Jaeger: So we might remove some first level nav items, a lot of second levels, third and so forth and on. So going back to feature request, I think with that one going either way. It would definitely be helpful even just to have that base to start from and then rip stuff out.

Lea Alcantara: So where there any other add-ons besides the Pages Module? Or what were most of the things that you had to do was able to be done natively with ExpressionEngine?

Tom Jaeger: Yeah, I mean, pretty much everything was able to be addressed natively with ExpressionEngine. We used a lot of the standard modules with Wygwam, Player, Matrix.

Lea Alcantara: Interesting. So Wygwam, were there any tweaks that needed to be made to accommodate extra alphabetic characters or language files at all?

Tom Jaeger: I don't believe so. I think Wygwam is pretty smooth.

Lea Alcantara: Interesting.

Emily Lewis: Well, that actually brings up something we had. One of our listeners wrote in, Jorge Barbosa. He had wanted us to bring up the topic of add-ons being multi-language friendly. For example, he was referring to the Favorites add-on from Solspace where you can configure your preferences to have a custom message when someone selects one of the actions with Favorite. But the message is always in English, there is no option of adding a different language response to those preferences. Both Lea and Tom, what are your thoughts or experiences with add-ons that are restricted to a certain language and then you need to utilize something like that for another type of site, another international site?

Tom Jaeger: Sure, a lot of add-ons are built with a language file from the ground up to actually support the functionality of that module. That's kind of the first place I look to support and to actually use that module in different language. Beyond that, with something like Favorites, I'm thinking tags off the top of my head. Maybe the ability to share tags between different languages, MSM or same site, things like that. That's something I think Solspace would have to look at, if we are talking about Solspace's Tag Module here.

Emily Lewis: It's probably something that developers should keep in mind, though.

Lea Alcantara: Yeah, just in general, yeah.

Tom Jaeger: Absolutely, absolutely.

Emily Lewis: I mean, just even like Lea was speaking in Canada, two languages are predominant. I mean, I don't think you can get away with relying on just American English for anything anymore.

Tom Jaeger: I agree.

Lea Alcantara: Yeah. And I think that's an interesting thing you brought up with the language files with add-ons. It's not just third-party add-ons, like in order to edit like certain system messages even in English easily in ExpressionEngine without hacking ExpressionEngine native files, you have to either buy an add-on or well, you actually have to hack the PHP file and manually change that which increases problems when you are upgrading ExpressionEngine and you don't document exactly what you did at every single site. So maybe that's something to be addressed. It's hard to say because, I mean, the entire world does use the Internet and English is a very predominant language, but we have to be sensitive to the fact that not everyone can or will want to work primarily with English-dominant sites or English-dominant control panels. But the challenges regarding that, I think just to make it easier without having to hack native files to do that would be something that I think EllisLab and third-party developers should keep in mind when they are developing.

And I mean a lot of ExpressionEngine developers aren't even from North America. Actually, the majority, well, I wouldn't say the majority, but for example, Low lives in Europe, obviously, and speaks a completely different language or native tongue. And when I tweeted that, I was going, "We are going to do the podcast on multi-language sites," a lot of Belgians were saying, "You need to have a Belgian here because the first language we have to deal with is Belgian. We don't even deal with English right away."

Tom Jaeger: Interesting.

Lea Alcantara: Yeah, so it's interesting where there is a ton of sites where the first priority isn't even English. Their first priority is their native language.

Emily Lewis: Tom, I was wondering that with your experience working on the different multilingual sites, you talked about some of your technical challenges. But I'm sort of wondering or sort of a hindsight perspective of having completed so many, if you could maybe talk about the tips for people who maybe they have to do a multilingual site for the first time to get them started from what you've learned and what you've seen through your time working with these different kinds of sites with some quick getting started tips.

Tom Jaeger: Sure. Now, I think the first thing would be to educate yourself on the different approaches to building multilingual sites with ExpressionEngine. Each one of them really does have their own pros and cons to it. If a site is going to have different content editors for different languages, you probably don't want to go the route of having each language in the same web log entry as a quick example.

Lea Alcantara: Interesting.

Tom Jaeger: The other thing I would keep an eye on when dealing with multi-language in general is the actual number of fields you have in your site. One of the issues I ran into early on while building multilingual sites was hitting a maximum MySQL row limitation with the XP Web log data channel or channel entries.

Lea Alcantara: Interesting.

Emily Lewis: Oh, I remember this when this happened at Pitney Bowes.

Tom Jaeger: Yes.

Lea Alcantara: And so how was that dealt with?

Tom Jaeger: Yep, so the way we dealt with it, thankfully right off the bat we had it backed up, so we rolled back the database. It happened on a site replication where you are replicating that, all the web log or channel fields for a given site into the new one at that time. So the first we did is we rolled back the database and brought the system up. Just to go back, the issues actually with MySQL are 1118, and I'll never forget that one.

Lea Alcantara: It's very specific.

Tom Jaeger: Yeah.

Lea Alcantara: Okay.

Tom Jaeger: I would never forget that one, unfortunately. Basically, it's when your column count in a given table exceeds the maximum byte allotment for a table. So there is no given magic number of fields. It's not when you hit 1,600 fields in a table that you get this error. Each data type or field type in ExpressionEngine requires a different amount of data towards that column count. So the ways that we got around that where we were able to minimize that is by switching to a lot of Matrix fields where it kind of uses less data in the channels data or web log data table through the convert a lot to Matrix. We were able to get rid of fields that maybe weren't really needed, if you will.

Lea Alcantara: Okay.

Tom Jaeger: So we're just creating fields, new field groups, where we had to use duplicates. So it was, "Hey, we had these ten fields and then in this new channel or web log, we are only using like four or five of them, let's not do that." So actually, as soon as we set that, it's actually not true, because when you use this field group, it does use the same fields in the table.

Lea Alcantara: Interesting.

Emily Lewis: So you really have to take a close look at your data itself as well?

Tom Jaeger: Absolutely.

Emily Lewis: You mentioned that it was relatively a smooth process working with Pitney Bowes' staff in part because they already spoke English. There wasn't a language barrier that you had to overcome. Have you worked on a site where you did have to deal with someone who spoke the native tongue, and was there anything you had to account for in that sort of situation?

Tom Jaeger: Good question. Do you mean is their first language was like Danish or something like that?

Emily Lewis: Right. I mean, I've worked with people who English isn't the first language, and I know they are communication issues, but in terms of not just communication but communicating how they needed the systems to work, I guess.

Tom Jaeger: Sure. I actually have run into that with multilingual sites. I've run into issue similar to that with just site builds or honestly, add-on support, in general.

Lea Alcantara: Interesting.

Emily Lewis: Just out of curiosity, how do you help alleviate any issues like that when you are dealing with a client or customer who their experience is going to be different than yours, not just because of language, but perhaps their expectation of functionality is going to be different?

Tom Jaeger: Yeah, sure. So the first thing, the expectation of functionality is probably the biggest one I see. If somebody is coming to our site with English as their second, maybe third language, I don't know, but the first thing I do is try to set the expectations of what the functionality or feature set is or what's built out in ExpressionEngine, how this is intended to work and how it's currently working.

Lea Alcantara: That's cool.

Emily Lewis: Oh, sorry, Lea. I have one last question. I was just curious. Have you done multi-language sites with any other kind of system other than ExpressionEngine as a point of comparison?

Tom Jaeger: Static HTML.

Lea Alcantara: Oh, wow.

Tom Jaeger: Yeah. Woof.

Lea Alcantara: So did you have any other choice besides ExpressionEngine? So why did you choose ExpressionEngine to work with, or was ExpressionEngine already chosen and then you were just brought in afterwards?

Tom Jaeger: Yeah, so most of the sites I built, I've been brought into after ExpressionEngine has been chosen. Oftentimes, the site is already built and they say, "Hey, it's on ExpressionEngine and we want to add another language now. Well, what's the best way to go about that?"

Lea Alcantara: Okay. Okay, interesting. So I think that's about all the time we have for today. So thank you so much, Tom, for joining us.

Tom Jaeger: Absolutely, my pleasure.

Emily Lewis: Yes, thanks, Tom. We really appreciate it. In case, our listeners have any questions or wanted to follow up with you, where could they find you online?

Tom Jaeger: They can go to EEHarbor or email me at tom@eeharbor.com.

Emily Lewis: Are you at Twitter?

Tom Jaeger: I'm sure at Twitter.

Emily Lewis: What's your handle?

Tom Jaeger: Yeah, it's [@tomjaeger](https://twitter.com/tomjaeger).

Emily Lewis: Okay.

Lea Alcantara: That's T-O-M-J-A-E-G-E-R?

Tom Jaeger: Correct, yeah.

Emily Lewis: Yes.

Lea Alcantara: You've got to spell it out when you are podcasting. So most of today's discussion focused on considerations for Roman alphabet, Latin-based languages, which is why we are planning another episode on multi-language development focusing on the nuances of Asian languages with Nicolas Bottari. So be sure to check our schedule for that upcoming episode.

Emily Lewis: And in the meantime, if you have any ideas, resources or questions you would like us to discuss during that upcoming podcast, please be sure to email us at feedback@ee-podcast.com.

Lea Alcantara: So before we go, we would like to thank our sponsors for this podcast, EECl 2011 and Pixel & Tonic.

Emily Lewis: We would also like to thank our partners, EllisLab, EngineHosting and Devot:ee.

Lea Alcantara: And thanks to you for listening. If you want to know more about the podcast and upcoming episodes, please follow up on Twitter [@ee-podcast](https://twitter.com/ee-podcast) or visit our website at ee-podcast.com. This is Lea Alcantara.

Emily Lewis: And Emily Lewis.

Lea Alcantara: Signing off for the ExpressionEngine Podcast. See you next time.

Emily Lewis: Cheers