|  |  |  |
| --- | --- | --- |
| **Date** | **Multilingual Questions and Issues for Nesstar 4** |  |
|  |  |  |
| March 8, 2012 –  Issue #1 | QUESTION:  Do lists such as variable groups have to be identical matches in terms of order and subgroups? |  |
|  | INFO:   * The purpose of variable groups (and subgroups) is to facilitate the finding of relevant variables. * Workflow: Alex notes that variable groups will probably have to be unclicked when importing FR XML, March 19, 2012.   Testing results:   * Alex also noted that after importing, FR variables are found in variable groups (and subgroups if defined in EN) in the order defined in the default language – there is no matching on actual variable groups, only on their EN instances and order. And any additional En language variable groups (that is to say if there are more default (EN) language variable groups than in the sublanguage (FR) variable groups), the variable groups will display in the default language in EN. * There were two observations also made on the matter of the Odesi subject/theme tree, which is now a bilingual (EN lists first, then FR), April 23 … * First the subject tree would have to be translated as is it in EN, whether there are surveys/aggregated data or not. The FR subject tree will then be in the order of the EN subject tree and therefore difficult to browse; * NSD, Norwegian Social Science Data Services <http://www.nsd.uib.no/and> and <http://www.nsd.uib.no/nsd/english/index.html> have translated their subject themes into EN, but the English subject themes remain in the alphabetical order of the “Norwegian” subject themes, see: <http://nsddata.nsd.uib.no/webview/>. |  |
|  | ANSWER -  Yes, unfortunately, assuming no changes to Publisher.  There are a couple of technical solutions. The first one is to translate name of variable groups (and subjects) in French but keep same order and content, April 12, 2012.  In other words, this could be approached by doing the grouping only on the EN side and then just import without groups and translate the EN variable groups and subgroups into FR afterwards.  However, as there are two problems to the first solution. 1) find-ability of variables in FR is compromised – the principle on the FR side is findability and because of that we edit STC’s variable groups and then sort them alphabetically, and 2) extremely time-consuming modifications would be required for adopting EN-side variable groupings without an automated multilingual validator of groupings at the variable level. The reason for this is because variables are not matched, only the group tree outline (and number of variables under that tree), for the default and sublanguage(s) is matched, a shortcoming for Publisher 4.0.  There are other alternatives to being in lockstep to the English variable groupings which can be seen by NSD. NSD has only translated the Norwegian study description fields into EN and has not translated any variable level fields into EN, leaving the Norwegian variable groups and variable-level metadata to display. | RESOLVED?  Undesirable 🡪 Contact Nesstar |
|  |  |  |
| March 8, 2012 – Issue #2 | QUESTION:  How do we ensure that French SPSS system file metadata gets imported when we are starting with a new French file (not yet published in Odesi)? |  |
|  | INFO: Alexandreconfirmed this works, April 13, 2012.  It is the choice of language in the Publisher that allows for the default metadata language, so that can be either EN or FR.  Testing results:  The results are the same whether starting in French or English as the default language. In other words, it is problematic in exactly the same way.  Method:  He first selecting French as the default language in Publisher, (making FR the default metadata language for the record) and so published first in French and then selecting French as the added language (making EN the metadata “sublanguage” for the record) and published in English last. And then the reverse was tested, making EN the default language and FR the sublanguage.  Note: the only variable is which language the problems occur in as complete line-ups are not possible.  Conclusion: Whether FR or EN, the default language overwrites the sublanguage where there are blanks or fewer instances in the sublanguage metadata compared to the default metadata. |  |
|  | ANSWER:  There is no problem technically with choice of default language in Publisher. The problem is rather one of manageability of the bleeding of default language metadata into the sublanguage metadata display for the FR user in Webview.  This is a policy issue on one level.   * Which is the lesser of two undesirable options - whether it is better for English metadata to display in French or vice versa;   This is also a Quality Control issue:   * Can we afford the level of QC time required to search out all the mis-matches between the records and reduce them to a sufficient minimal level, that is the bleeding of EN metadata into the FR metadata (as the default language will likely be EN in Publisher as on the Server)? | RESOLVED?  Not in 4.0 as is |
|  |  |  |
| March 8, 2012 – Issue #3 | ISSUE:  Any blank FR fields (or fewer numbers of instances than in EN), e.g., a) keywords, b) other documentation, c) authoring entity at the document description level) will be filled in in English.  Note: Issue #4 below (pre- and post-question text) is a further example of this. |  |
|  | Publishing Issues Alexandre confirmed this as a problem.  This is problematic on two levels, first (problem 3.1) the number of instances has to match and second (problem 3.2), any deleted information on the French side will also be deleted in the English.  Example: where there seven keywords in EN (default language) and five in FR (sublanguage), the five FR keywords will display in the FR Webview followed by the 6th and 7th EN keywords.  This issue is about different practices in two cases, (keywords and other documentation) and a decision to be made in the third case (authoring entity), and perhaps in the first two other cases.   1. The keywords in EN and FR could be very different. In FR shorter keyword lists are preferred (such as the IMDB Subjects lists, [example ”Sujets”)](http://www23.statcan.gc.ca:81/imdb/p2SV_f.pl?Function=getSurvey&SDDS=3226&lang=en&db=imdb&adm=8&dis=2#a1) 2. Originally the other documentation practices were very similar in EN and FR. Now command files are being provided in EN, resulting in –differing practices re: other documentation and considerable user confusion at the U of O, where users, including TA’s in methodology classes, think this is the data file. 3. Authoring entity – even though the authors (e.g., 8 authors for the EN GSS 1 alone) for the English record, may not have been involved in the French record, they need to display in French, followed by the FR authors because of the problem where EN information cannot be deleted on the French side without the information also being deleted on the English side. |  |
|  | Quality Control Issues  Extremely difficult to manage without automated validation tools to help identify the blanks are in the sublanguage. | RESOLVED?  Not in 4.0 |
|  |  |  |
| March 9, 2012 – Issue #4 | QUESTION/ISSUE: Pre and post question text missing in French will display English pre and post questions (GSS1) |  |
|  | INFO: Alexandre confirmed this.  All instances have to match.  Only solution at present is to add (that is: translate the EN) the FR metadata, and adjust the French version to match the English. It is impossible to simply erase the English information in the French version, the English will simply re-appear instantly. |  |
|  | AN*SWER*:  Quality control would be very challenging in both searching out & identifying all the problems and then correcting them. Not all EN records have pre- and post-question text.  Because this is one of a long list of multilingual questions, it should go with the list of questions to Nesstar for their feedback on possible tools for automating validation (and with it, unlinking the EN and FR). | RESOLVED?  Not in 4.0 |
|  |  |  |
| March 19, 2012 – Issue #5 | QUESTION/ ISSUE: Values not described or declared in French will display English value labels in the frequencies (e..g., FR GSS 1 does not have values of “0” and “98” declared for EDYEARS. |  |
|  | INFO: Alexandre confirmed this April 13. The value labels also have to be a perfect match. |  |
|  | ANSWER: | RESOLVED? |
|  |  |  |
| March 19, 2012  Issue #6 | QUESTION / ISSUE: French browsing – bi-lingual tree – How to in Publisher? |  |
|  | INFO: French subject headings can be added so that there is browsing in French. Users will have to change the language to French and the subject headings and names of files will be French.  Alex tested how to translate the English subjects into FR for her Nesstar 4.0 DLI presentation in April on the Odesi test server at <http://142.150.191.250:88/>, these is her FR display compared to the EN:   |  |  | | --- | --- | | EN | FR | |  |
|  | ANSWER: In the Manage Server box, change language to French and enter the French names. You will have to erase the English names first.  Note: The longer the list of subjects/themes, the more difficult to navigate a non-alphabetical subject/theme list. | RESOLVED? yes |
|  |  |  |
| March 27  Issue #6 | QUESTION or ISSUE: Downloading XML – seems to only download English even if in French side of metadata. |  |
|  | INFO: downloading French SPSS will get French .sav file with English metadata. When a bi-lingual file is published, two files are created on the server. E.g. publish a bi-lingual file with ID# gss-12M0001-E-1985-c-1-test creates two xml files -  gss-12M0001-E-1985-c-1-test.xml gss-12M0001-E-1985-c-1-test\_fr.xml Nesstar adds the “\_fr”.  Question - how to download both English xml file and French xml file? | |  | | --- | |  | |
|  | ANSWER: Nesstar Support says this issue was fixed in Server release 4.0.3 or higher. Will have server upgraded to see if this fixes the problem. | RESOLVED? downloading codebooks in french and english work with the new server (4.0.4) |
|  |  |  |
| March 20  Issue #7 | QUESTION:  Are there real Nesstar 4.0 sites out there that use Nesstar multilingual data down to variable browsing and variable description? |  |
|  | FOR INFORMATION – multilingual sites  **NSD’s Collection of Norwegian Surveys** – Nesstar 4.0 <http://nsddata.nsd.uib.no/webview/>  Notes:  This site is not using the multi-lingual metadata function like we were planning to.  The variable descriptions and variable groupings are in Norwegian only. In addition, they don’t use a subject/theme tree.  Uses multi-lingual metadata function at ?? levels:   * Ability to change language of Interface between Norwegian and English * Study description is Norwegian and English   ANSWER: No  **Centre for Socio-Political Data (CDSP)** - Nesstar 4.0 <http://nesstar.sciences-po.fr:81/webview/>  This site is also not using the multi-lingual metadata function as we were planning to. Most of the study description is in one default language (varies by study series) and the full variable description and the variable groupings are in the one default language of the study as well. In addition, they don’t use a subject/theme tree.  Uses multi-lingual metadata function at two levels:   * Ability to switch interface * Very occasionally (e.g., the General Election Study Belgium 2003), the abstract in the study description is provided in both English and French   ANSWER: No  **Local Government Data Unit Wales** – Nesstar 3.5 with some 4.0 patches  Cubes w/ .svg maps (3.5) and the 4.0 multilingual patch  <http://dissemination.dataunitwales.gov.uk/webview/>  This site (cubes only, no survey microdata) is using the multi-lingual metadata function as we were planning to. The subject/theme tree (with sub-themes) is multilingual (Welsh and English). The study description and other documentation is fully Welsh and EN. The detailed dimensions and cells descriptions for the cells are fully Welsh and EN.  ANSWER: Yes (as above) and no (they don’t use microdata, and we don’t use cubes) |  |
|  |  |  |
|  |  |  |
| April 23 – Issue #8 | QUESTION / ISSUE:  The <p> codes for breaking paragraphs in lengthy abstracts import into 4.0 as text and not as commands |  |
|  | INFO:  As there are limited abstracts, which are easy to monitor during markup, these paragraph break codes can be deleted at time of mark-up whether in EN or FR.  Note: this is a general issue, not specific to multilingual metadata. |  |
|  | ANSWER:  This seems to be a manageable Quality Control issue. | RE |
|  |  |  |
|  | QUESTION / ISSUE: |  |
|  | INFO: |  |
|  | ANSWER: |  |
|  |  |  |