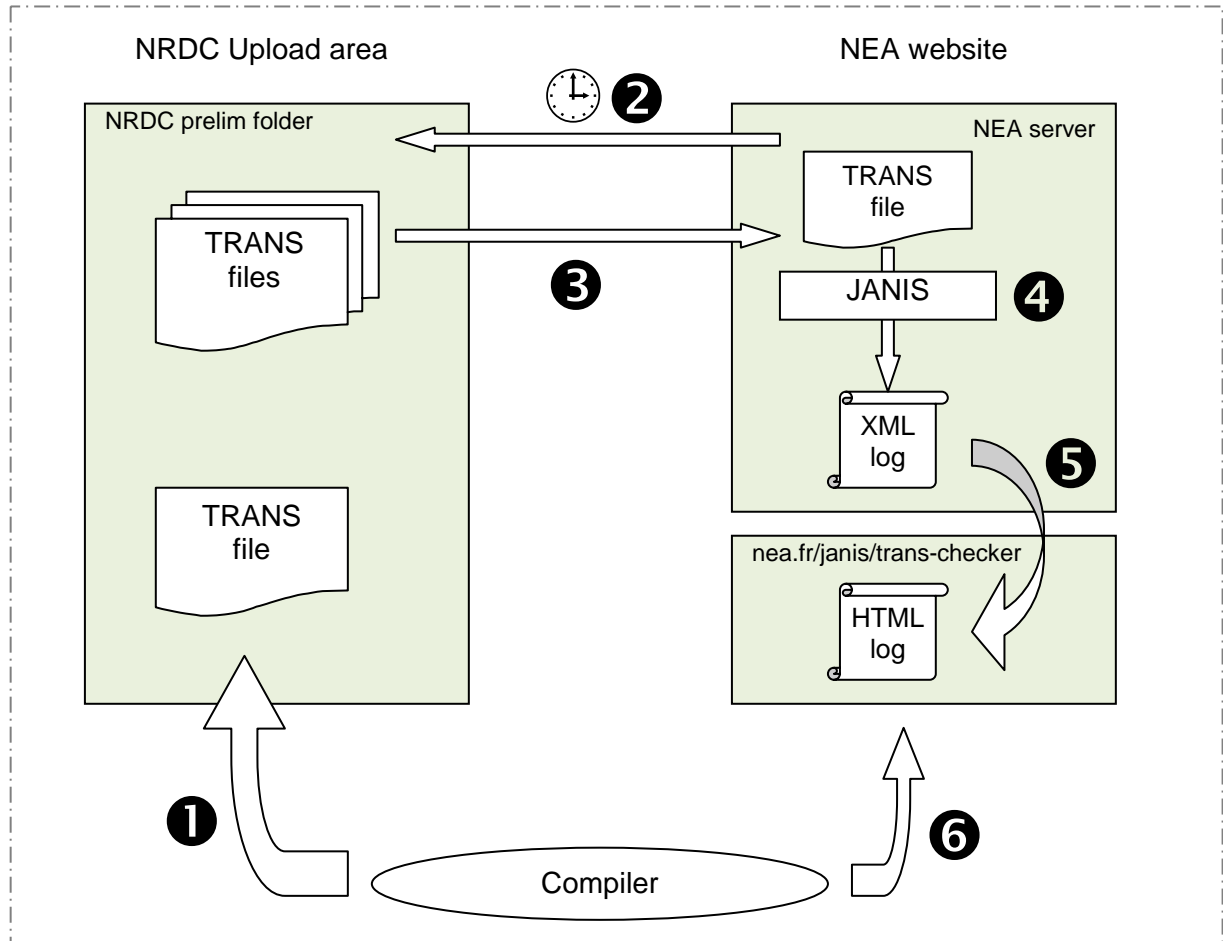


JANIS EXFOR TRANS Checker

www.nea.fr/janis/trans-checker

Principle



Steps:

1. The EXFOR compiler upload his/her prelim file to NRDC upload area
2. NEA server periodically check if new TRANS files have been uploaded
3. If a new file is detected it is downloaded to NEA server
4. A "JANIS import" of this file is done, and this produces an XML log file of errors/warnings
5. This XML log file is transformed into an HTML file, which is accessible from Internet
6. The EXFOR compiler can then review the HTML log file

Steps 3-5 take only a few seconds so the time to wait for obtaining the HTML log after having uploaded a TRANS file to NRDC upload area file depends on the check frequency chosen for step 2. The check frequency is currently set to "every 3 minutes".

How to understand output log files

Sample output for PRELIM.D067

JANIS - TRANS Check Log		
0 Errors		
10 Warnings		
Subent	Message	Details <i>line number</i>
D0507.001	[REFERENCE] : Bad Abstract/Conf code, invalid Conference code	2008INTLAK <i>line 50</i>
D0524.001	[AUTHOR] : Blanks are not permitted following initials	H. Johansson <i>line 158</i>
D0524.002	Missing AUTHOR keyword	<i>line 201</i>
D0524.003	Missing AUTHOR keyword	<i>line 222</i>
D0552.001	[METHOD] : Unknown Code	TROJA <i>line 2212</i>
D0568.001	[REFERENCE] : Bad Journal code, empty Page subfield, comma should be omitted	RCA,, <i>line 8327</i>
D0572.001	[REFERENCE] : Bad Journal code, empty Page subfield, comma should be omitted	NIM/B,, <i>line 9128</i>
D0573.001	[AUTHOR] : Author name too short (comma used instead of dot ?)	R.F.Seller,D.F.Herring,T.,K.W.Jones <i>line 9223</i>
D0573.002	Missing AUTHOR keyword	<i>line 9765</i>
D0573.003	Missing AUTHOR keyword	<i>line 10022</i>

JANIS - TRANS Check Log

The HTML log files starts with a summary of errors and warnings found.

This is followed by a 3-columns table were yellow background means warning and red background means error.

- The 1st column indicates the subentry were the error/warning occurs (in some case this can be the subentry preceding the location of error, e.g. for an error in the SUBENT system identifier)

- The 2nd column indicates the type of the error/warning. The list of possible messages here is limited.
- The 3rd column gives details about the error/warning. Usually it's here that you will find the offending values. The list of values, which can appear depends on the input files. The line number is given in italics. (Like 1st column value this can also be the line before or the line after the real problem)

Analysis of the example:

The 1st message means that 2008INTLAK is a code that is not present in dictionaries. This could in some cases be due to the fact that the new code has recently been proposed in a memo, but has not yet been incorporated in the dictionaries.

The 2nd message indicates that the AUTHOR keyword does not follow EXFOR rules. The space between 'H.' and 'Johansson' should be removed.

The 3rd and 4th messages are consequences of the 2nd message above. The space in the AUTHOR coded value has prevented JANIS from parsing, and keeping, this author name. So the check 'presence of required keywords', which is done after parsing each subentry of an EXFOR TRANS file reports that AUTHOR keyword is missing. This is the most difficult part to understand because in reality this keyword is not missing, but rather was not understood by JANIS. We may try, in future versions of JANIS, to distinguish a missing keyword from a misinterpreted keyword, but for the time being this makes no difference in our EXFOR reading code.

The 5th message, "[METHOD]: Unknown Code" "TROJA line 2212", means that 'TROJA' is not a valid 'METHOD' code (In current dictionaries it appears that it's in fact a valid 'ANALYSIS' code).

The 6th and 7th messages indicate that, as the journal codes do not mention the page numbers, the commas should be omitted. Note that the values displayed in the 3rd column are only the part of the EXFOR codes inside the parenthesis, up to the reference.

The 8th message indicates another potential error in AUTHOR coding.

The 9th and 10th messages are consequences of the 8th message, similar to the 3rd and 4th messages above.

Known issues

HTML log files accumulate in the www.nea.fr/janis/trans-checker folder. New PRELIM files are detected, but removed PRELIM files do not yet trigger a delete of the corresponding HTML log file.

Messages were initially not intended to be understood by everybody, so they are probably sometimes difficult to understand.

JANIS skips any keyword that doesn't follow the EXFOR rules. For the time being we have no code which tries to "repair" wrong data. This may not seem to be difficult to do, but in practice can be very dangerous. For example it may not be obvious how to correct the 8th message in the example above.

Contact

We understand that it is not easy at all to “decipher” these logs so do not hesitate to mail us if you have any questions: soppera@nea.fr, bossant@nea.fr