

2022.07.15

Translation based on the Hungarian document, 2022.07.15

The launch of a new MTMT feature: duplicate suspects

With launch of a new MTMT feature: duplicate suspects, the processes for searching, handling and entering (import and manual) duplicate records have become more efficient. The existing rules for searching and detecting duplicate publications in the database have been modified, and a number of external identifiers (WOS, SCOPUS, DOI, PubMed) have been given a prominent role in the search process.

For records imported from WOS, SCOPUS, PubMed, the efficiency of duplicate detection has been significantly improved. The new rules and conditions are also applied to manual data entry, so that the efficiency of duplicate searches when non-public records are made public or when a duplicate search is launched from a specific record or list has also increased.

The new search and marking process distinguishes between a true duplicate and a suspected duplicate. Real duplicates and their suspected duplicate records are placed in a duplicate group. In the group, real duplicates retain their duplicate marker, while suspected duplicate records are given a suspected duplicate marker (?). Only two suspected duplicate records can be marked as duplicates at the same time. A record marked as a suspected duplicate becomes a true duplicate when it is actually marked as a duplicate by its designated pair in the current duplicate management interface. It can then be treated as a duplicate. A record marked as a suspected duplicate may be marked as a non-duplicate if justified, but loses its suspected duplicate marking afterwards. Only public records will continue to be involved in the duplicate search process. A search can still be initiated on a non-public record. A new feature is that, unlike before, the search is not limited to records of the same type (e.g. journal article / chapter in a book can also be a candidate for suspected duplicate).

When searching for a duplicate, the following rules apply:

- if two external identifiers (e.g. WOS, DOI) are the same and there are no mismatched pairs (e.g. both have SCOPUS identifiers but they are different), the records are marked as duplicates.
- if the record contains more than one external identifier of the same type (e.g. 2 WOS identifiers) it can only receive a duplicate suspect mark.
- where at least one pair of identifiers is identical (e.g. same DOI), regardless of type it will be marked as duplicate suspect
- regardless of type for a title match (the duplicate search for a currently valid title algorithm), a suspected duplicate will be assigned to the records concerned.
- in the case where only suspect records are found during the duplicate search, all elements of the duplicate group are marked as suspect.

Duplicate suspicion handling for publication (Core, Citing Articles) import: Ikfélasd – kléké v

During the import from WOS, SCOPUS file, the system "collects" and adds the missing (WoS, Scopus, PubMed) external identifiers to the records via APIs (e.g. WOS identifier addition for SCOPUS import), the duplicate checking process is only run afterwards. If the program finds a duplicate during the import, the already public record detected as a duplicate will be added to the import list. In the case where only a suspected duplicate record is found, the record is not public, it is placed in the import list

with a suspected duplicate flag. In the case of WOS XML automatic Citation imports, records with a suspected duplicate flag are assigned the status "Work in progress, Not Published". Whether from file or by automatic WOS-XML import, records with a non-public duplicate flag should be made public after the duplicate flag has been checked.

After selecting a non-public record marked as a suspected duplicate during the check, the records to be processed can be displayed by selecting Duplicates / Duplicate search / Show from the top menu.

If the records are indeed duplicates of each other, they can be marked as duplicates by clicking on the *Mark as duplicate* button and, if authorised, the duplicates can be merged. In the case where the selected records are not duplicates of each other, clicking on the *Mark as non duplicate* button will unmark the suspected duplicate and make the citation not yet public.

Other important rules for handling records with duplicate flags:

- when making public: the record of a new publication that is not yet public, if the duplicate check finds "only" a suspected duplicate record, it may be made public.
- records marked as suspected duplicates are not involved in the duplicate handling, neither in manual nor in automatically triggered merges.

A duplicate suspect message record may be searched by a duplicate search from a record, but may also be searched on selected or all elements of a list where the above rules apply.

In the *Publication* datatype, by specifying the Duplicate/Suspected duplicate conditions, the duplicate suspect records marked as duplicates can be searched and listed.

Duplicate

	Type	ID	ID	ID	ID
1st rekord	Article	DOI	WOS	–	PubMed
2nd rekord	Article	DOI	–	Scopus	PubMed
Match	matches	matches	–	–	matches

The type matches and there are at least 2 pairs of matches from the prominent identifiers (DOI, WOS, Scopus, PubMed)

Duplicate Suspects

	Type	ID	ID	ID
1st rekord	Article	DOI	WOS	–
2nd rekord	Article	DOI	–	Scopus
Match	matches	matches	–	–

There is 1 pair of matches, but not at least 2 pairs of matches from the prominent identifiers

	Type	ID	ID	ID	ID
1st rekord	Article	DOI	WOS	–	PubMed
2nd rekord	Study	DOI	–	Scopus	PubMed
Match	does'n match	matches	–	–	matches

If the type does not match, it cannot be a duplicate even if it has more than 1 matching ID pair

	Type	ID	ID	ID	ID
1st rekord	Article	DOI	WOS	Scopus	PubMed
2nd rekord	Article	DOI	WOS	Scopus	PubMed
Match	matches	matches	matches	does'n match	matches

One pair of prominent identifiers does not match

	Type	ID	ID	ID	ID
1st rekord	Article	DOI	WOS	WOS	PubMed
2nd rekord	Article	DOI	WOS	–	PubMed
Match	matches	matches	matches	–	matches

More than 1 of the same prominent identifier type belonging to the same publication