

Ch 06a: Importing and matching references (worked example)

Ian Johnson, updated 29 May 2026

Background

This section gives a worked example of importing two sets of references (Primary and Secondary) for inscriptions from a spreadsheet derived from Zotero data. The example comes from the IDENK project (Idenk.net) based at the EFEO, courtesy the project director Arlo Griffiths [REQUEST AGREEMENT, I am sure it will not be a problem]

Basic structure

- *Inscription* records contain two record pointer fields (*Primary references* and *Secondary references*)
ec988974-2ede-4a79-a58a-98a433475e9b.jpg
- These point to *Bibliographic reference* records
1e939358-73d7-4c84-afae-35b25613ecc5.jpg
- *Bibliographic reference* records contain
 - a record pointer field to a bibliographic record (Book, Chapter, Journal Article, Thesis etc.) derived from the Zotero library
 - pagination information (a text field containing page numbers, illustration or other information about sections of the document)

- *Bibliography records* are of various types (Book, Chapter, Journal Article, Thesis etc.)

c01a9cd9-1508-4a84-aa09-47f4660a1fce.jpg

The bibliography records are identified by strings such as Adams1912_01 (not shown in the view above). These are identifiers which have been filled in the Zotero Short Title field in a large Zotero library (21,000 references). These are referred to as ZSTs.

One could also use the Zotero key field, which is automatically populated with an 8 alphanumeric hash key which is statistically unique and cannot be edited - it is this key that we use to link our internal bibliographic records back to their Zotero origin records.

The steps are as follows:

- The first batch were already in the Heurist database so these identifiers are exported to a spreadsheet. Later batches will be imported from a spreadsheet used to collect data 'in the field'
- If not already separate in the spreadsheet, use Libre Office *Data > To columns* to split the ZST and the pagination reference (page numbers and or illustrations)
- Deduplicate the rows in the spreadsheet (Data > Duplicates) based on the ZST and pagination. Each row will then be a reference to a particular place in a particular bibliographic record.
- Import into Bibliographic reference records
- Split the original spreadsheet (before deduplication) into two CSV files, one for Primary and one for Secondary references
- Import each of these files into the Inscriptions table, the first into the primary reference field, the second into the secondary reference field.

These fields are record pointer fields pointing to Bibliographic reference records, so the textual ZST + pagination value is first matched with the values in the database and this generates the ID of the Bibliographic reference records created in the previous steps, and we use this ID which is inserted into

the record pointer fields.

- After updating the bibliographic records (Book, Chapter, Journal Article, |Thesis etc.) from the Zotero library (Populate > *Zotero Bibliography sync* we need to relate the *Bibliographic reference records* to the bibliographic records by matching the ZST in the first with the Zotero Short Title field in the second, using *Recode > Foreign Key match*.

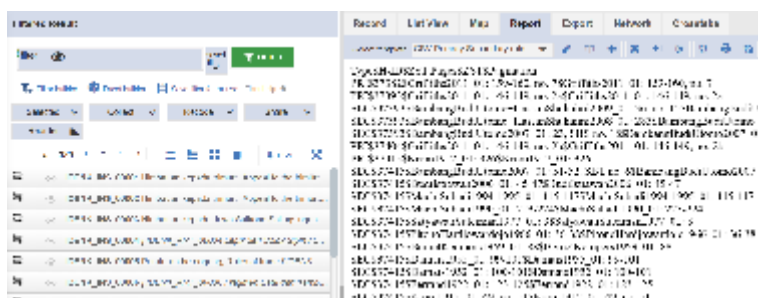
From scratch

Delete all Bibliographic references. This also deletes the pointers to them from Inscriptions.

Preparing the spreadsheet

Batches 2 and 3 will already have their much more comprehensive spreadsheet, see later.

Export all the existing ZST references for Primary and Secondary refs (not Surrogates) to CSV using *CSV Primary Secondary refs* custom report:



Open in Libre Office (the delimiter is tab, not \$ as shown)

Rather than child records we will point multiple inscriptions to common Biblio Reference records which include a page range. Note that editing these records can corrupt other Inscription entries which point to it if the change is such as to change the reference, since the Bibliographic reference records are to a specific place in a particular bibliographic entity.

The alternative is the use of child records and significant duplication (1 in 4 approx). Using independent bibliography references is altogether simpler to deal with apart from the slight drawback above.

Preparing the batch X spreadsheet

To document when I have the final spreadsheet

Loading the bibliographic references and linking

Load into Heurist with Populate > CSV.

Select INScriptions for H-ID as these records relate to data in the Inscriptons

[illegible]

However, first choose Bibliographic record pointer as we want to create bibliography records and then reference them in Inscriptions.

[illegible]

Note: after deduplication we have 1114 Biblio references, these examples were pre deduplication

[illegible]

Target entity type: Bibliographic Reference [View entity type](#)

Bibliographic Reference

Workflow instruction below

☐ Create new records ☐ Update existing records ☐ Import existing records

Existing: 0 New: 1474

Line value	Unique value	Column	Index to update	Index to update
☐	1474	PR1	Primary	Primary
☒	1114	28	Page	Page
☒	200	28	Page	Page
☒	270	Page	Page	Page
☒	4	Page	Page	Page

Column name: + Add

Value:

We now have 1114 bibliographic records like this:

ZST Pages - text reference in INScripton

ZST (Zotero short title) used as identifier

Page/illustration references

Bibliographic pointer

ibangBudiUtomo+HassanShuhaimi2008_01: 110-111

BambangBudiUtomo+HassanShuhaimi2008_

110-111

selected: Book, Book chapter, Journal, Journ...

Now split the incoming original spreadsheet into Primary (n=398) and Secondary (n=1076) references based on PRI and SEC in the first column.

Save as two CSV files. Load each in turn.

Importing from: PriSecZSTs Batch 1 SECONDARY.csv

Specify identifier and date columns

Identifier columns are those that contain a fluidat record ID. They MUST contain "H ID" somewhere in the column name.

Column	Type	Fluidat Identifier	Fluidat Identifier
Type	Integer	☒	Fluidat Identifier
H ID	Integer	☒	Fluidat Identifier
ZST Page	Integer	☐	Fluidat Identifier
Page	Text	☐	Fluidat Identifier

Fluidat Identifier:

Please preview. First 1000 rows

Should show clearly rendered data in a table with column names in first row. If not, try changing choice of separator characters and entering

Type	H ID	ZST Page	ZST	Page
SEC	37008	BambangBudiUtomo+HassanShuhaimi2008_01: 110-111	BambangBudiUtomo+HassanShuhaimi2008_01	110-111
SEC	37008	BambangBudiUtomo+HassanShuhaimi2008_01: 110-111	BambangBudiUtomo+HassanShuhaimi2008_01	110-111
SEC	37008	BambangBudiUtomo+HassanShuhaimi2008_01: 110-111	BambangBudiUtomo+HassanShuhaimi2008_01	110-111
SEC	37008	BambangBudiUtomo+HassanShuhaimi2008_01: 110-111	BambangBudiUtomo+HassanShuhaimi2008_01	110-111

[illegible]

Workflow instruction below

☒ Match on columns ☐ Skip matching (all new records)

If new records are added to the database, you will match **Bibliographic Reference** in the incoming data against records already in the database. You do not add columns containing multiple values for inserting, as this will generate multiple records per input line. To force creation of a new record, select columns and map to a field which will never match.

New records to add will be created. Matching will find 1:1 field for new records to insert where the new record all new records for unmatched rows.

Unique values Expand	Column	Column to 1:1 with mapping (inserted match)
MATCHING	Choose only the fields you want to match. Note: ON Y fields suitable for matching are shown in bold.	
☐ 001	Primary author ID	New column which enables found IDs
☐ 1	Type	
☒ 305	ZST Pages	ZST Pages - text reference: ... + Add link
☐ 131	ZSI	
☐ 265	Publication	

Now select the Inscription records which are to be updated.

Target entity type: Inscription ☐ Check this box to ignore record type. Target record type can be left blank to query all records. Also select one of field. Matching and pointers will be applied across all records by default with the two field mapping.

Bibliographic Reference ☐ Existing (New: 0) ☐ New (New: 0)

Workflow instruction below

step 1: MATCHING

☒ Match on column (s) ☐ Use H-ID ☐ Match on column (s) (Use H-ID for matching)

Please select one or more columns on which to match inscription in the importing data with the records already in the database. Note: you cannot use columns containing variable values for matching, as this will generate records from the imported data. As more records are added, values will change and this will affect the match. Only one column should be selected.

The selected column will be used for matching. The selected field will be used for matching records and allow the record set of new records for imported rows.

Use value	Unique values	Column	Column to Field mapping (record match)	Values in row 1
☒	191	H-ID	Column to field mapping being selected here	210003
☐	305	Primary refs H-ID		00682
☐	1	Type		194
☐	305	2ST Pages		Colins2011 01 100 100 row 7
☐	191	2ST		Colins2011 01
☐	282	Expansion		194-194, row 7

Click on Use H-ID (this was in the original file and referenced the INScriptons)

Existing: 191 New: 0 tells us that there are 191 Inscriptions (of the 398) which have Primary bibliographical data (for the Secondary references it's **Existing: 282 New: 0**)

We import the Primary references H-ID into the *Primary refs* > record pointer field (later, the Secondary references H-ID into the Secondary refs > record pointer field):

Primary references:

Target entity type: Inscription ☐ Check this box to ignore record type. Target record type can be left blank to query all records. Also select one of field. Matching and pointers will be applied across all records by default with the two field mapping.

Bibliographic Reference ☐ Existing (New: 0) ☐ New (New: 0)

Workflow instruction below

step 2: FIELDS TO IMPORT

☐ Create new records ☒ Update existing records ☐ Ignore no data - Add data later?

Please select the columns which you wish to import into the inscription record when being imported. Note that changing the match on the database will require a full refresh of data.

Existing: 191 New: 0

Use value	Unique values	Column	Fields to update	Values in row 1
☒	191	H-ID	Field to update being selected here	210003
☐	305	Primary refs H-ID	Primary refs > [Record pointer]	00682
☐	1	Type		194
☐	305	2ST Pages		Colins2011 01 100 100 row 7
☐	191	2ST		Colins2011 01
☐	282	Expansion		194-194, row 7

Secondary references:

Drop the table into the 'Inscription' tablespace

Importing the table into the 'Inscription' tablespace

Workflow restriction table

Step 2: FIELDS TO IMPORT

Fields to update

Use value	Unique value	Column	Fields to update	Values in row 1
☒	1070	1070	1070	1070
☐	1070	1070	1070	1070
☐	1070	1070	1070	1070
☐	1070	1070	1070	1070
☐	1070	1070	1070	1070

Prepare, then Start Update:

Import of Inscription compl...

Import of Inscription compl...

References:

and all looks good:

BIBLIOGRAPHY

Primary refs > [G000052011_01 \[140-149, no. 2a \]](#)

Primary refs (DHARMA Short Title) [G000052011_01 \[140-149, no. 2a \]](#)

Secondary refs [BambangBudiUtomo+HassanShuhaimi2008_01 \[38, no. 15 \]](#)

[BambangBudiUtomo+HassanShuhaimi2008_01 \[283 \]](#)

[BambangBudiUtomo2007_01 \[23, SBS no. 15 \]](#)

Secondary refs (DHARMA Short Title) [BambangBudiUtomo+HassanShuhaimi2008_01 \[38, no. 15 \]](#)

[BambangBudiUtomo+HassanShuhaimi2008_01 \[283 \]](#)

[BambangBudiUtomo2007_01 \[23, SBS no. 15 \]](#)

Connecting Bibliographic reference records with bibliographic entities

Now we have to connect our Bibliographic reference records with the appropriate bibliographic records imported from Zotero. We must do this for each of the reference types used since we cannot match across multiple tables.

For books:

Revision #3

Created 2026-07-01 11:57:35 UTC by IanJohnson

Updated 2026-07-02 12:51:34 UTC by IanJohnson