



# **Handleiding OpenRefine**

## **Deel 3 — XML grouped**

Nynke Kuipers—Kubus erfgoed & digitalisering

Oktober 2024



**KUBUS.frl**  
erfgoed & digitalisering

Beeld voorpagina: Flesvormige vaas met drie shishi en een draak anoniem, Publiek domein



## Licentie

Op deze handleiding is een Creative Commons Naamsvermelding-licentie van toepassing. Deze publicatie mag worden bewerkt, doorgestuurd, geüpload, gedownload, gekopieerd en anderszins verspreid, zolang de naam van de maker vermeld wordt. De gehele licentietekst is te vinden op: <https://creativecommons.org/licenses/by/4.0/>

Als je deze handleiding wil hergebruiken kun je de volgende methode van naamsvermelding aanhouden:

*Handleiding OpenRefine - deel 3 - xml grouped (Nynke Kuipers, Kubus erfgoed & digitalisering) CC-BY*



Deze handleiding is tot stand gekomen binnen de Datawerkplaats van Landschap Erfgoed Utrecht.



# Inhoudsopgave

|                                                      |    |
|------------------------------------------------------|----|
| <b>Introductie</b> .....                             | 4  |
| <b>Project maken in OpenRefine</b> .....             | 5  |
| Het XML bestand.....                                 | 5  |
| Project aanmaken.....                                | 6  |
| Kolommen herschikken in OpenRefine.....              | 8  |
| Wat zie je? .....                                    | 10 |
| <b>Voorbereiding volgorde cellen</b> .....           | 11 |
| Edit lege cellen.....                                | 12 |
| Edit enkele lege cel .....                           | 13 |
| Resultaat voorbereiding.....                         | 14 |
| <b>Samenvoegen en verplaatsen</b> .....              | 15 |
| Kolommen samenvoegen schaalaaanduiding .....         | 15 |
| Leegmaken cellen schaalaaanduiding .....             | 18 |
| Kolommen vrije tekst en schaal samenvoegen .....     | 20 |
| <b>Data ombouwen naar Adlib-tagged formaat</b> ..... | 21 |
| Tags toevoegen aan veldinhoud.....                   | 22 |
| Herhaalbare velden samenvoegen.....                  | 24 |
| Verwijder overbodige kolommen .....                  | 25 |
| Split into several columns .....                     | 26 |
| Extra kolom toevoegen.....                           | 27 |
| De kolommen naar rijen .....                         | 28 |
| <b>Importbestand maken</b> .....                     | 30 |
| <b>Links en verwijzingen</b> .....                   | 32 |

# Introductie

XML-Grouped verwerken in OpenRefine is de manier om met herhaalbare regels (occurrences) om te gaan en om de informatie uit verschillende velden, maar per regel op de goede volgorde te houden. Deze manier van exporteren wordt ondersteund in **Adlib** en **Axiell Collections**.

In het voorbeeld van deze handleiding gaan aan de slag met de verschillende velden van de **afmeting: onderdeel, afmeting, waarde, eenheid, precisie, bijzonderheden** en **vrije tekst**. Er zijn schaal aanduidingen in als afmeting genoteerd en die gaan we naar het **afmetingen vrije tekst** veld verplaatsen, terwijl de andere afmetingen intact blijven.

Om ook zelf de stappen uit de handleiding te kunnen reproduceren is het xml-bestand te downloaden: [oefenbestand \(zipfile met xml\)](#)

**Let op:** *het opschonen gebeurt buiten je systeem. Aanpassingen die in de tussentijd door collega's gedaan worden, worden bij het importeren van de schone set **overschreven**.*

**Let op:** *importeren in Adlib / Axiell Collections is een onomkeerbare actie. Zorg dat je zeker weet dat wat je doet. Je kunt het proces ook eerst op drie records **uitproberen**: mocht er iets misgaan, kan je dit makkelijk terugdraaien / wijzigen.*

**Tip vooraf:** *Maak een **back-up** van je data voordat je ermee aan de slag gaat. Bekijk de officiële handleiding hoe dat aan te pakken. Mocht er dan toch iets fout gaan, kan je de data terugplaatsen. Adlib: via de Customer Support Portal. Axiell Collections: <https://help.collections.axiell.com/en/Topics/Home.htm>*

In deze handleiding is gewerkt met [OpenRefine Version 3.8.2](#). Bij nieuwere of oudere versies kan het er net iets anders uitzien dan getoond in de screenshots van deze handleiding.

Voor uitgebreide instructies over het gebruik van OpenRefine is er de [user manual](#) (engelstalig) van OpenRefine zelf.

# Project maken in OpenRefine

## Het XML bestand

Een XML bestand kun je openen in je internetbrowser. Het is een hele technische weergave van de data, en daardoor als bestand zelf niet erg mensen-leesbaar.

Adlib / Axiell collections maakt onderscheid tussen **xml-grouped** en **xml-unstructured**.

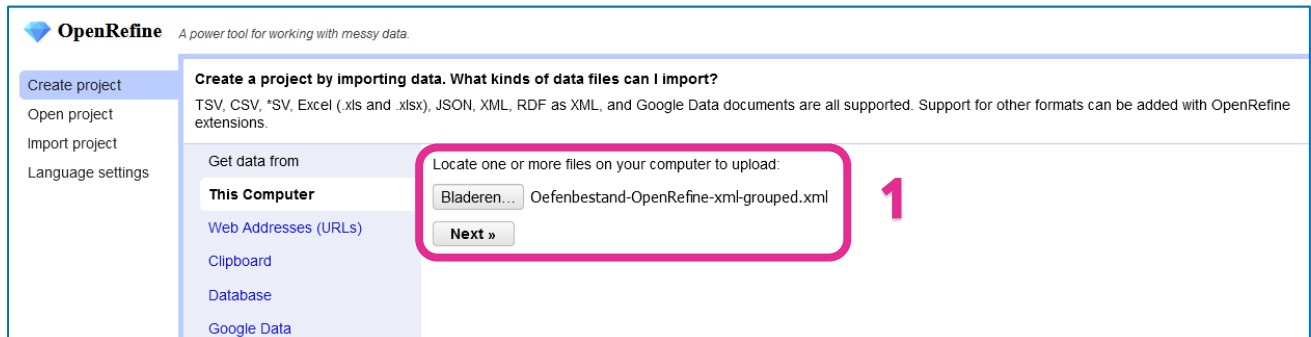
**Overeenkomst** is een hele complete weergave van je data. Het belangrijkste **verschil** is dat in een xml-grouped dat bij herhaalbare regels van velden de volgorde van regels inclusief de lege velden gehandhaafd blijft.

Zo ziet het oefenbestand eruit in een internetbrowser:

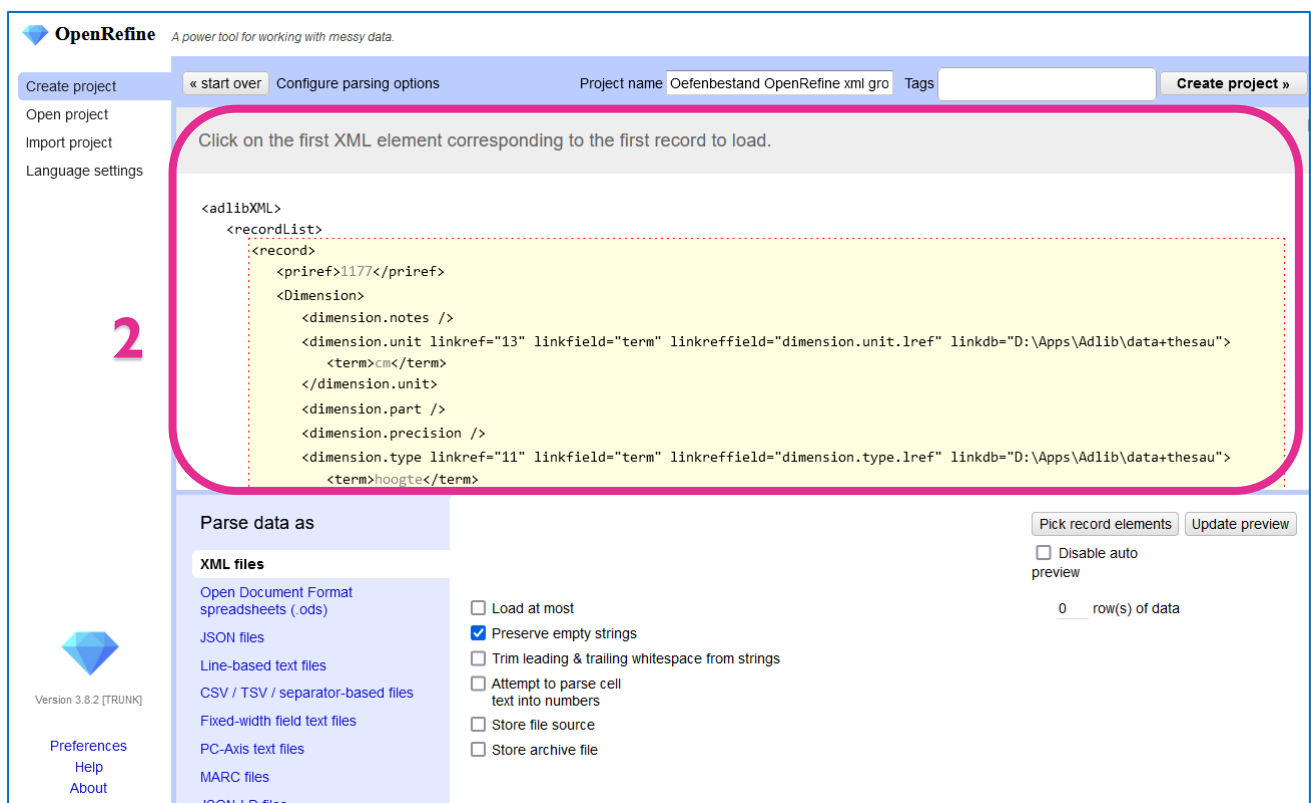
```
<recordList>
  <record>
    <preref>1177</preref>
    <Dimension>
      <dimension.notes/>
      <dimension.unit linkref="13" linkfield="term" linkreffield="dimension.unit.lref" linkdb="D:\Apps\Adlib\data+thesau">
        <term>cm</term>
      </dimension.unit>
      <dimension.part/>
      <dimension.precision/>
      <dimension.type linkref="11" linkfield="term" linkreffield="dimension.type.lref" linkdb="D:\Apps\Adlib\data+thesau">
        <term>hoogte</term>
      </dimension.type>
      <dimension.value>26.0</dimension.value>
    </Dimension>
    <Dimension>
      <dimension.notes/>
      <dimension.unit linkref="13" linkfield="term" linkreffield="dimension.unit.lref" linkdb="D:\Apps\Adlib\data+thesau">
        <term>cm</term>
      </dimension.unit>
      <dimension.part/>
      <dimension.precision/>
      <dimension.type linkref="12" linkfield="term" linkreffield="dimension.type.lref" linkdb="D:\Apps\Adlib\data+thesau">
        <term>breedte</term>
      </dimension.type>
      <dimension.value>10.0</dimension.value>
    </Dimension>
    <Dimension>
      <dimension.notes/>
      <dimension.unit linkref="13" linkfield="term" linkreffield="dimension.unit.lref" linkdb="D:\Apps\Adlib\data+thesau">
        <term>cm</term>
      </dimension.unit>
      <dimension.part/>
      <dimension.precision/>
      <dimension.type linkref="1866" linkfield="term" linkreffield="dimension.type.lref" linkdb="D:\Apps\Adlib\data+thesau">
        <term>lengte</term>
      </dimension.type>
      <dimension.value>62.0</dimension.value>
    </Dimension>
  </record>
</recordList>
```

# Project aanmaken

1. Ga naar **bladeren** om het gewenste bestand te selecteren > **next**



2. Record selecteren  
Geef aan welk deel van de data het **record** weergeeft



### 3. Instellingen

OpenRefine ziet zelf al wat voor soort bestand het is en maakt een **preview**. Je kunt de instellingen zodanig aanpassen totdat je preview het gewenste resultaat heeft.

Voor dit voorbeeld gebruik ik de standaard instellingen zoals hieronder aangegeven.

**OpenRefine** A power tool for working with messy data.

Create project « start over Configure parsing options Project name: Oefenbestand OpenRefine.xml.gro Tags: **4** Create project »

Open project  
Import project  
Language settings

|     | record - Dimension - dimension.unit - linkref | record - Dimension - dimension.unit - linkfield | record - Dimension - dimension.unit - linkref | record - Dimension - dimension.unit - linkref |
|-----|-----------------------------------------------|-------------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| 1.  | 13                                            | term                                            | dimension.unit.lref                           | D:\Apps\Adlib\data+thesau                     |
| 2.  | 13                                            | term                                            | dimension.unit.lref                           | D:\Apps\Adlib\data+thesau                     |
| 3.  | 13                                            | term                                            | dimension.unit.lref                           | D:\Apps\Adlib\data+thesau                     |
| 4.  | 0                                             | term                                            | dimension.unit.lref                           | D:\Apps\Adlib\data+thesau                     |
| 5.  | 13                                            | term                                            | dimension.unit.lref                           | D:\Apps\Adlib\data+thesau                     |
| 6.  | 13                                            | term                                            | dimension.unit.lref                           | D:\Apps\Adlib\data+thesau                     |
| 7.  | 13                                            | term                                            | dimension.unit.lref                           | D:\Apps\Adlib\data+thesau                     |
| 8.  | 0                                             | term                                            | dimension.unit.lref                           | D:\Apps\Adlib\data+thesau                     |
| 9.  | 13                                            | term                                            | dimension.unit.lref                           | D:\Apps\Adlib\data+thesau                     |
| 10. | 13                                            | term                                            | dimension.unit.lref                           | D:\Apps\Adlib\data+thesau                     |

**3**

**Parse data as**

**XML files**

- Open Document Format spreadsheets (.ods)
- JSON files
- Line-based text files
- CSV / TSV / separator-based files
- Fixed-width field text files
- PC-Axis text files
- MARC files

☐ Load at most

☒ Preserve empty strings

☐ Trim leading & trailing whitespace from strings

☐ Attempt to parse cell text into numbers

☐ Store file source

☐ Store archive file

☐ Disable auto preview

0 row(s) of data

Pick record elements Update preview

Version 3.8.2 [TRUNK]

Preferences  
Help  
About

### 4. Create project

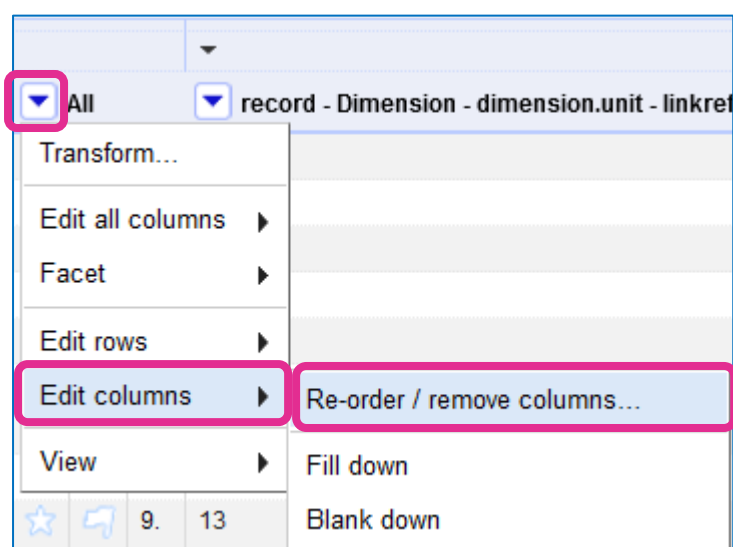
## Kolommen herschikken in OpenRefine

Met een xml export krijg je een hoop kolommen mee waar je niets mee hoeft. Deze kolommen bevatten technische data. Hierdoor is het nodig om als eerste stap de kolommen te herschikken, zodat je goed aan de slag kunt met de data. Om zeker te weten dat de recordstructuur op de juiste manier behouden blijft, verwijder ik deze technische kolommen niet, maar verplaats ik ze alleen.

De technische kolommen ontstaan omdat het veld in Adlib / Axiell Collections een gerelateerd veld is. Bijvoorbeeld een term uit de thesaurus. Hierdoor wordt meerdere metadata meegenomen in de export. Aan de hand van het veld **afmeting type** levert dit de volgende kolommen op:

- dimension.type - linkref > recordnummer van de term in de gerelateerde database
- dimension.type - linkfield > soort data dat moet worden weergegeven
- dimension.type - linkrefield > veldnaam van de link
- dimension.type - linkdb > het pad naar de gerelateerde database
- dimension.type - **term** > de daadwerkelijke inhoud van het veld (hoogte, breedte, etc)

1. Klik op het pijltje voor de kolom **All > Edit columns > Re-order / remove columns...**



2. Je kunt nu de blokjes met de kolomnamen naar de gewenste plek schuiven (drag and drop). Begin altijd met het **recordnummer (=priref)** of **objectnummer**

**Re-order / Remove columns**

Drag columns to re-order **2**

Drop columns here to remove

- record - priref
- record - object\_number
- record - Dimension - dimension.part
- record - Dimension - dimension.type - term
- record - Dimension - dimension.value
- record - Dimension - dimension.unit - term
- record - Dimension - dimension.precision
- record - Dimension - dimension.notes
- record - dimension.free
- record - Dimension - dimension.unit - linkref
- record - Dimension - dimension.unit - linkfield
- record - Dimension - dimension.unit - linkrefield
- record - Dimension - dimension unit -

Remove all Add all

OK Cancel

3. Je kolommen zijn nu geordend. Let op de weergave van rows en records. Komt jouw aantal records overeen met hoeveel records je uit je systeem hebt gehaald?

**JA:** je kan door

**NEE:** probeer nog wat aan je instellingen te rommelen en de volgorde van kolommen aan te passen

**725 records**

Show as: rows records

**2228 rows**

Show as: rows records

## Wat zie je?

Er zijn nu kolommen zichtbaar waar je wat mee wilt:

| ▼ All | ▼ record - priref | ▼ record - object_number | ▼ record - Dimension - dimension.part | ▼ record - Dimension - dimension.type - term | ▼ record - Dim |
|-------|-------------------|--------------------------|---------------------------------------|----------------------------------------------|----------------|
| ★ ↗   | 1.                | 1177                     | 00533                                 | hoogte                                       | 26.0           |
| ★ ↗   |                   |                          |                                       | breedte                                      | 10.0           |
| ★ ↗   |                   |                          |                                       | lengte                                       | 62.0           |
| ★ ↗   |                   |                          |                                       | schalen                                      | 1:100          |
| ★ ↗   | 2.                | 1360                     | 00666                                 | hoogte                                       | 110.0          |
| ★ ↗   |                   |                          | scheepsmodel                          | breedte                                      | 25.0           |
| ★ ↗   |                   |                          |                                       | lengte                                       | 107.0          |
| ★ ↗   |                   |                          |                                       | schalen                                      | 1:24           |
| ★ ↗   | 3.                | 1872                     | 01014                                 | hoogte                                       | 36.0           |
| ★ ↗   |                   |                          |                                       |                                              |                |
| ★ ↗   |                   |                          |                                       | breedte                                      | 9.5            |
| ★ ↗   |                   |                          |                                       | lengte                                       | 43.0           |
| ★ ↗   |                   |                          | romp                                  | lengte                                       | 31.0           |
| ★ ↗   |                   |                          |                                       | schalen                                      |                |
| ★ ↗   | 4.                | 1895                     | 01031                                 | hoogte                                       | 43.0           |
| ★ ↗   |                   |                          |                                       | breedte                                      | 16.5           |
| ★ ↗   |                   |                          |                                       | lengte                                       | 101.0          |
| ★ ↗   |                   |                          |                                       | schalen                                      | 1:50           |

Maar ook kolommen waar je niets mee kunt/hoeft (de technische velden die automatisch mee zijn gekomen uit je systeem):

| ▼ record - Dimension - dimension.type - linkref | ▼ record - Dimension - dimension.type - linkfield | ▼ record - Dimension - dimension.type - linkreffield | ▼ record - Dim   |
|-------------------------------------------------|---------------------------------------------------|------------------------------------------------------|------------------|
| 11                                              | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 12                                              | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 1866                                            | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 15532                                           | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 11                                              | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 12                                              | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 1866                                            | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 15532                                           | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 11                                              | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
|                                                 |                                                   |                                                      |                  |
| 12                                              | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 1866                                            | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 1866                                            | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 15532                                           | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 11                                              | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 12                                              | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 1866                                            | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |
| 15532                                           | term                                              | dimension.type.lref                                  | D:\Apps\Adlib\da |

## Vorbereiding volgorde cellen

Voordat we echt aan de slag kunnen, moeten we eerst zorgen dat de velden die meerdere regels (occurrences) hebben in de goede volgorde blijven staan en bij de goede regel blijven.

Denk hier bijvoorbeeld aan de afmetingen in Adlib. Hier wordt de informatie in herhaalbare regels bij elkaar geplaatst. En deze volgorde en sortering zijn dan ook erg belangrijk om zo goed mogelijk intact te houden.

Weergave in Adlib:

| Afmetingen<br>Onderdeel | Afmeting                | Waarde | Eenh               | Precisie | Bijzonderheden    |
|-------------------------|-------------------------|--------|--------------------|----------|-------------------|
| scheepsmodel            | <a href="#">hoogte</a>  | 36.0   | <a href="#">cm</a> |          | 1014a             |
|                         | <a href="#">breedte</a> | 9.5    | <a href="#">cm</a> |          |                   |
|                         | <a href="#">lengte</a>  | 43.0   | <a href="#">cm</a> |          | lengte over alles |
| romp                    | <a href="#">lengte</a>  | 31.0   | <a href="#">cm</a> |          |                   |

Weergave in OpenRefine:

|       |              |         |      |    |
|-------|--------------|---------|------|----|
| 01014 | scheepsmodel | hoogte  | 36.0 | cm |
|       |              |         |      |    |
|       |              | breedte | 9.5  | cm |
|       |              | lengte  | 43.0 | cm |
|       | romp         | lengte  | 31.0 | cm |
|       |              | schalen |      |    |

Om dit te doen moeten we de lege velden vullen bij de velden met herhaalbare regels. In dit voorbeeld: afmeting **onderdeel**, **afmeting**, **waarde**, **eenheid**, **precisie** en **bijzonderheden**.

Bij de overige velden hoef je niets aan te passen, want daar is geen gerelateerde volgorde.

Er zijn meerdere manieren van aanpak mogelijk. We nemen er twee door:

### 1. Edit lege cellen

Hier gaan meerdere lege cellen in één keer vullen

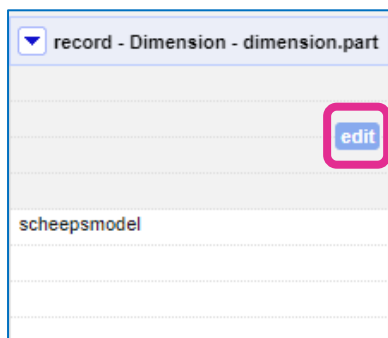
### 2. Edit enkele lege cel

Hier gaan we eerst zoeken naar een lege cel om die vervolgens te vullen

## Edit lege cellen

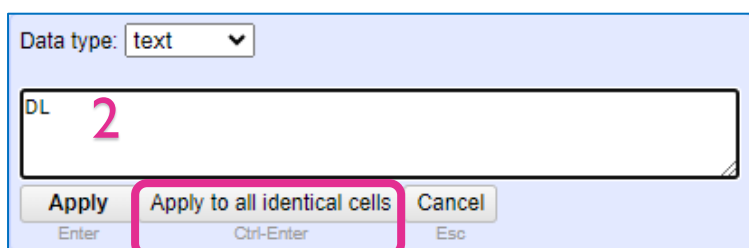
In dit voorbeeld verander ik de inhoud van de lege velden in de *velddtag* bij kolom afmeting onderdeel.

1. Beweeg over een leeg veld in de kolom die je wilt aanpassen > klik op **edit**



| record - Dimension - dimension.part |
|-------------------------------------|
|                                     |
|                                     |
|                                     |
| scheepsmodel                        |
|                                     |
|                                     |

2. Typ de gewenste tekst in en druk op **Apply to all identical cells**



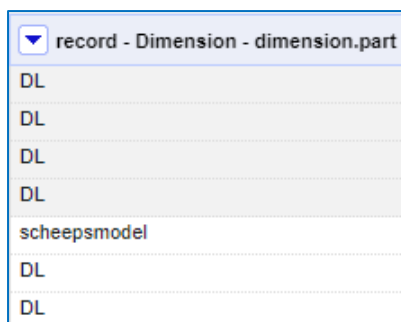
Data type: text

DL

Apply Apply to all identical cells Cancel

Enter Ctrl-Enter Esc

3. Het resultaat:

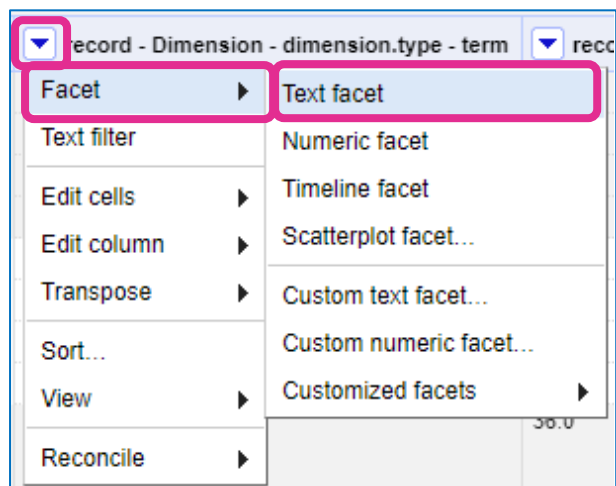


| record - Dimension - dimension.part |
|-------------------------------------|
| DL                                  |
| DL                                  |
| DL                                  |
| DL                                  |
| scheepsmodel                        |
| DL                                  |
| DL                                  |

## Edit enkele lege cel

Soms zie je bij de eerste records geen lege cellen staan, maar zijn ze er misschien wel.

1. Klik op het pijltje voor de kolomnaam > **Facet** > **Tekst facet**



2. Zoek **blank** en klik erop om de records met een lege cel te tonen (staat altijd helemaal onderaan)



3. Klik op **edit** (in dit geval vul ik een afmeting in, en verander ik *alleen deze cel* met **Apply**)

| record - Dimension - dimension.type - term | record - Dimension - dimension.value | record - Dimension - dimension.unit - term |
|--------------------------------------------|--------------------------------------|--------------------------------------------|
| hoogte                                     | 59.0                                 | cm                                         |
|                                            | 47.0                                 | cm                                         |
| schalen                                    | 1:10                                 |                                            |

Data type: text

breedte

**Apply** Apply to all identical cells Cancel

Enter Ctrl-Enter Esc

## Resultaat voorbereiding

**Herhaal** de stap die het beste past bij de velden **waarde**, **eenheid**, **precisie** en **bijzonderheden**.

Nu ziet je Tabel er ongeveer zo uit: Bij de velden die onderdeel zijn van de herhaalbare regel, zijn alle cellen gevuld en zo blijft de volgorde van de informatie altijd weer bij de juiste regel te plaatsen.

| record - object_number | record - Dimension - dimension.part | record - Dimension - dimension.type - term | record - Dimension - dimension.value | record - Dimension - dimension.unit |
|------------------------|-------------------------------------|--------------------------------------------|--------------------------------------|-------------------------------------|
| 00533                  | DL                                  | hoogte                                     | 26.0                                 | cm                                  |
|                        | DL                                  | breedte                                    | 10.0                                 | cm                                  |
|                        | DL                                  | lengte                                     | 62.0                                 | cm                                  |
|                        | DL                                  | schalen                                    | 1:100                                | EE                                  |
| 00666                  | scheepsmodel                        | hoogte                                     | 110.0                                | cm                                  |
|                        | DL                                  | breedte                                    | 25.0                                 | cm                                  |
|                        | DL                                  | lengte                                     | 107.0                                | cm                                  |
|                        | DL                                  | schalen                                    | 1:24                                 | EE                                  |
| 01014                  | scheepsmodel                        | hoogte                                     | 36.0                                 | cm                                  |
|                        | DL                                  | breedte                                    | 9.5                                  | cm                                  |
|                        | DL                                  | lengte                                     | 43.0                                 | cm                                  |
|                        | romp                                | lengte                                     | 31.0                                 | cm                                  |
|                        | DL                                  | schalen                                    | W/A                                  | EE                                  |

Je bent er klaar voor om aan de slag te gaan met het samenvoegen en verplaatsen van de schaal aanduidingen!

# Samenvoegen en verplaatsen

Als voorbeeld wil ik de **schaalaanduidingen**, die nu als aparte afmetingen genoteerd staan, verplaatsen naar **afmeting (vrij veld)**. Hiervoor moet de data uit cellen van verschillende kolommen worden samengevoegd, maar dan alleen voor de schaal aanduidingen, de rest blijft als het is. We gaan stap voor stap daar naartoe werken.

## Kolommen samenvoegen schaal aanduiding

De rijen waar de schaal aanduiding in genoemd is wil ik graag samenvoegen

1. Zoek eerst alle records waar '**schalen**' is ingevoerd

record - Dimension - dimension.type - term

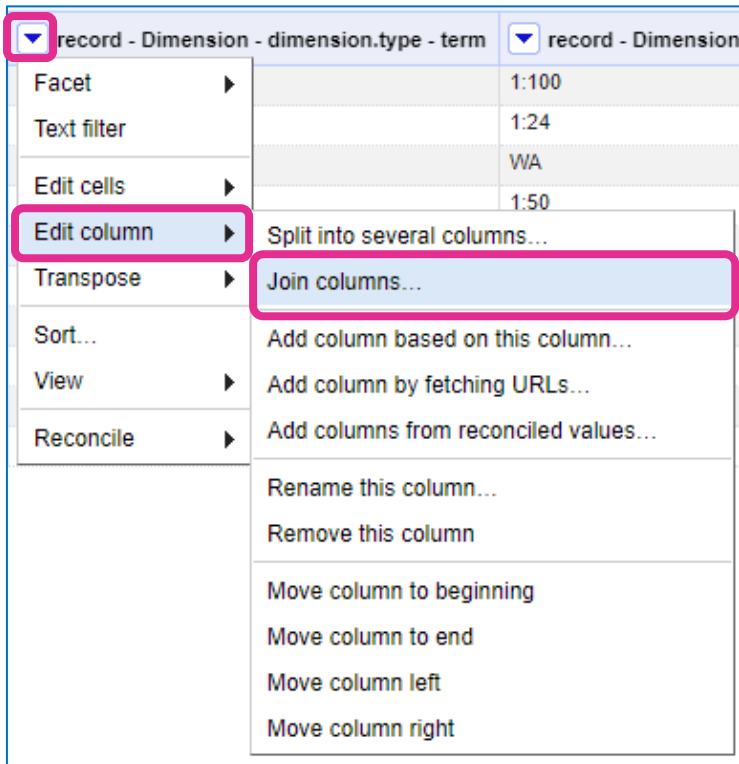
7 choices Sort by: name count Cluster

breedte 723  
DI 1  
hoogte 725  
lengte 52  
plaat 1  
schaal (afm) 1  
**schalen 725** exclude  
Facet by choice counts

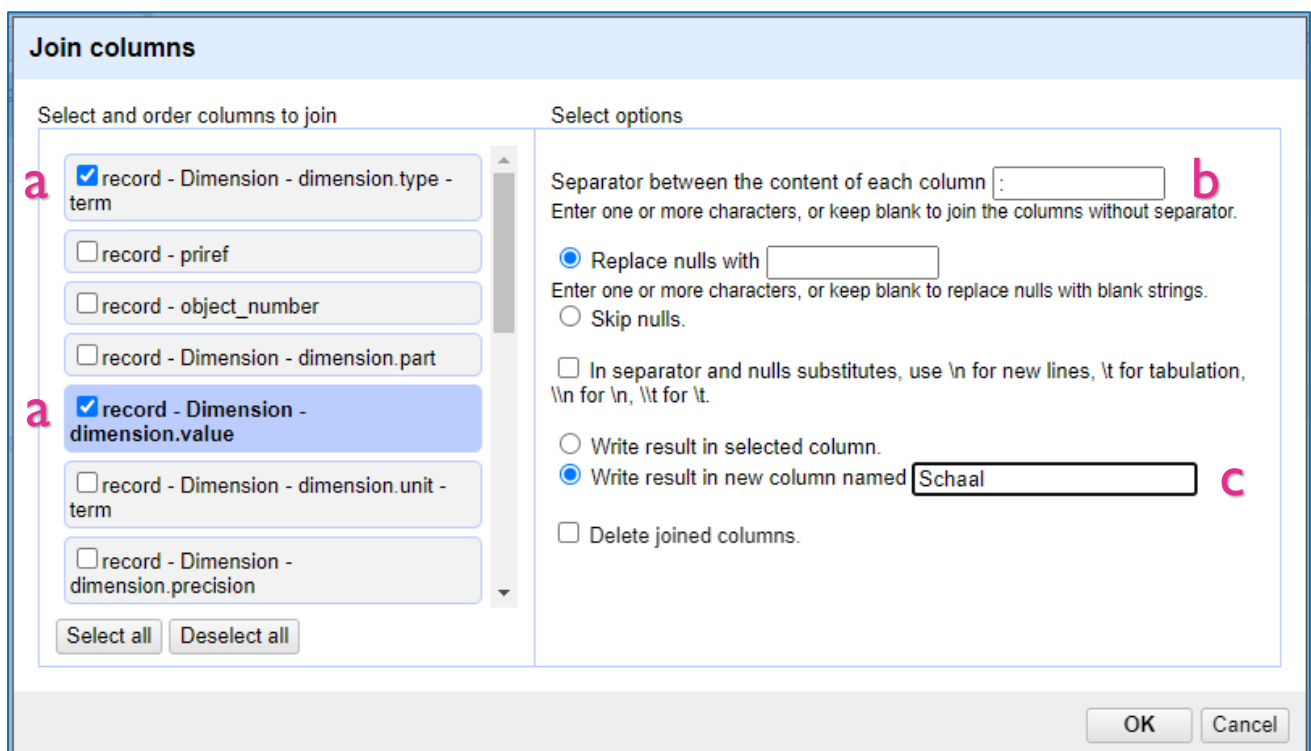
2. Verander de Records-weergave in de **Rows**-weergave

| Show as: <b>rows</b> records        |                                            | Show: 5 10 25 50 100 500 1000 rows   |                                      | « first                              | « previous                           | 1                                    |
|-------------------------------------|--------------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| record - Dimension - dimension.part | record - Dimension - dimension.type - term | record - Dimension - dimension.value | record - Dimension - dimension.value | record - Dimension - dimension.value | record - Dimension - dimension.value | record - Dimension - dimension.value |
| DL                                  | schalen                                    | 1:100                                | EE                                   |                                      |                                      |                                      |
| DL                                  | schalen                                    | 1:24                                 | EE                                   |                                      |                                      |                                      |
| DL                                  | schalen                                    | WA                                   | EE                                   |                                      |                                      |                                      |
| DL                                  | schalen                                    | 1:50                                 | EE                                   |                                      |                                      |                                      |
| DL                                  | schalen                                    | 1:20                                 | EE                                   |                                      |                                      |                                      |
| DL                                  | schalen                                    | 1:20                                 | EE                                   |                                      |                                      |                                      |

3. Klik op het pijltje voor de eerste kolom die je wilt samenvoegen > **Edit column** > **Join columns**



4. Pop-up scherm: kolommen samenvoegen
- Selecteer de velden: **afmeting** (type) en **waarde**
  - Seperator “: ” (dubbele punt + spatie) (of een scheidingsteken naar wens)
  - We willen de samengevoegde kolommen graag in een nieuwe kolom samenvoegen



## 5. Resultaat

Je ziet in de record-weergave heel goed dat alleen de cellen uit de selectie van de rijen met afmeting type schalen is samengevoegd. De rijen met andere soorten afmetingen zijn niet samengevoegd.

*In de Rows-weergave:*

| Show as: <b>rows</b> records Show: 5 10 25 50 100 500 1000 rows « first |                   |                                      |                                     |
|-------------------------------------------------------------------------|-------------------|--------------------------------------|-------------------------------------|
| record - Dimension - dimension.type - term                              | Schaal            | record - Dimension - dimension.value | record - Dimension - dimension.unit |
| schalen                                                                 | schalen:<br>1:100 | 1:100                                | EE                                  |
| schalen                                                                 | schalen:<br>1:24  | 1:24                                 | EE                                  |
| schalen                                                                 | schalen:<br>WA    | WA                                   | EE                                  |
| schalen                                                                 | schalen:<br>1:50  | 1:50                                 | EE                                  |
| schalen                                                                 | schalen:<br>1:20  | 1:20                                 | EE                                  |
| schalen                                                                 | schalen:<br>1:20  | 1:20                                 | EE                                  |

*En in de Record-weergave:*

| Show as: rows <b>records</b> Show: 5 10 25 50 100 500 1000 records « first |                   |                                      |                                     |
|----------------------------------------------------------------------------|-------------------|--------------------------------------|-------------------------------------|
| record - Dimension - dimension.type - term                                 | Schaal            | record - Dimension - dimension.value | record - Dimension - dimension.unit |
| hoogte                                                                     |                   | 26.0                                 | cm                                  |
| breedte                                                                    |                   | 10.0                                 | cm                                  |
| lengte                                                                     |                   | 62.0                                 | cm                                  |
| schalen                                                                    | schalen:<br>1:100 | 1:100                                | EE                                  |
| hoogte                                                                     |                   | 110.0                                | cm                                  |
| breedte                                                                    |                   | 25.0                                 | cm                                  |
| lengte                                                                     |                   | 107.0                                | cm                                  |
| schalen                                                                    | schalen:<br>1:24  | 1:24                                 | EE                                  |

## Leegmaken cellen schaalaaanduiding

Nu is het nodig om de cellen in de kolommen **afmeting (type)** en **waarde** waar een schaalaaanduiding staat, leeg te maken. Deze data is nu samengevoegd in een nieuwe kolom, dus deze data kan verwijderd worden uit de oorspronkelijke kolommen.

1. Zoek eerst alle records waar '**schalen**' is ingevoerd > selecteer deze
2. Zorg dat je in de **rows**-weergave zit
3. Kolom **Waarde** > **Edit cells** > **Common transforms** > **To null**

The screenshot shows the OpenRefine interface with a table of data. The first column is 'record - Dimension - dimension.value', the second is 'record - Dimension - dimension.unit - term', and the third is 'record - Dimension - dimension.value'. The first row has values 'EE' and 'PR'. The second row has values 'EE' and 'PR'. The third row has values 'ca.' and 'ca.'. The 'Edit cells' menu is open, showing options like 'Transform...', 'Common transforms', 'Fill down', 'Blank down', 'Split multi-valued cells...', 'Join multi-valued cells...', 'Cluster and edit...', and 'Replace...'. The 'Common transforms' submenu is also open, showing options like 'Trim leading and trailing whitespace', 'Collapse consecutive whitespace', 'Unescape HTML entities', 'Replace smart quotes with ASCII', 'To titlecase', 'To uppercase', 'To lowercase', 'To number', 'To date', 'To text', 'To null', and 'To empty string'. The 'To null' option is highlighted with a red box, and a large red number '3' is next to it.

| record - Dimension - dimension.value | record - Dimension - dimension.unit - term | record - Dimension - dimension.value |
|--------------------------------------|--------------------------------------------|--------------------------------------|
| EE                                   | PR                                         |                                      |
| EE                                   | PR                                         |                                      |
| ca.                                  | ca.                                        |                                      |

4. Nu gaan we hetzelfde doen voor **afmeting (type)** > > **Edit cells** > **Common transforms** > **To null**

## 5. Resultaat van de opschoning:

| Show as: <b>rows</b> <b>records</b> Show: 5 10 25 50 100 500 1000 records « first < pr |                   |                                        |                        |
|----------------------------------------------------------------------------------------|-------------------|----------------------------------------|------------------------|
| ▼ record - Dimension - dimension.type - term                                           | ▼ Schaal          | ▼ record - Dimension - dimension.value | ▼ record - Dimension - |
| hoogte                                                                                 |                   | 26.0                                   | cm                     |
| breedte                                                                                |                   | 10.0                                   | cm                     |
| lengte                                                                                 |                   | 62.0                                   | cm                     |
|                                                                                        | schalen:<br>1:100 |                                        | EE                     |
| hoogte                                                                                 |                   | 110.0                                  | cm                     |
| breedte                                                                                |                   | 25.0                                   | cm                     |
| lengte                                                                                 |                   | 107.0                                  | cm                     |
|                                                                                        | schalen:<br>1:24  |                                        | EE                     |
| hoogte                                                                                 |                   | 36.0                                   | cm                     |
|                                                                                        |                   |                                        |                        |
| breedte                                                                                |                   | 9.5                                    | cm                     |
| lengte                                                                                 |                   | 43.0                                   | cm                     |
| lengte                                                                                 |                   | 31.0                                   | cm                     |
|                                                                                        | schalen:<br>WA    |                                        | EE                     |

## Kolommen vrije tekst en schaal samenvoegen

De schaal aanduidingen moeten nu nog samengevoegd worden met eventueel bestaande inhoud in het veld **afmeting vrije tekst**.

1. Klik op het pijltje voor de kolom **dimension free** > **Edit column** > **Join columns...**
2. Pop-up scherm: kolommen samenvoegen
  - a. Selecteer de kolommen **dimension.free** en **Schaal**
  - b. Als separator kies ik “;” [puntkomma + spatie]
  - c. Lege velden slaan we over > dit voorkomt onnodige ; .
  - d. We willen de samengevoegde kolommen graag in een nieuwe kolom samenvoegen
  - e. Tot slot mogen de oude kolommen meteen worden verwijderd

**Join columns**

Select and order columns to join

a ☒ record - dimension.free

☐ record - priref

☐ record - object\_number

☐ record - Dimension - dimension.part

☐ record - Dimension - dimension.type - term

a ☒ **Schaal**

☐ record - Dimension - dimension.value

☐ record - Dimension - dimension.unit - term

☐ record - Dimension - dimension.precision

Select all Deselect all

Select options

Separator between the content of each column ; b  
Enter one or more characters, or keep blank to join the columns without separator.

☐ Replace nulls with   
Enter one or more characters, or keep blank to replace nulls with blank strings.

☒ Skip nulls. c

☐ In separator and nulls substitutes, use \n for new lines, \t for tabulation, \ for \n for \n, \\t for \t.

☐ Write result in selected column.

☒ Write result in new column named d

☒ Delete joined columns. e

OK Cancel

## Data ombouwen naar Adlib-tagged formaat

De data is opgeschoond en aangepast. Nu moeten de resultaten weer teruggegeven worden Adlib. In dit voorbeeld is de werkwijze met Adlib-tagged bestand uitgewerkt. Er zijn meer mogelijkheden voor soorten importbestanden, die laten we hier even buiten beschouwing.

Een **Adlib-tag** is opgebouwd uit twee tekens (bijv. **DI**) met een [spatie] en dan de veldinhoud. De tekens van de tags zijn hoofdlettergevoelig. Er komen ook cijfers voor in de tags. Alleen het recordnummer heeft een speciaal teken (%0 [procent + nul]).

De tags van de velden die we willen importeren:

| Technische naam     | Veldnaam                 | Tag |
|---------------------|--------------------------|-----|
| Priref              | Recordnummer             | %0  |
| Dimension.part      | Onderdeel                | DL  |
| Dimension.type      | Afmeting                 | DI  |
| Dimension.value     | Waarde                   | WA  |
| Dimension.unit      | Eenheid                  | EE  |
| Dimension.precision | Precisie                 | PR  |
| Dimension.notes     | Bijzonderheden           | DN  |
| Dimension.free      | Afmetingen (vrije tekst) | DA  |

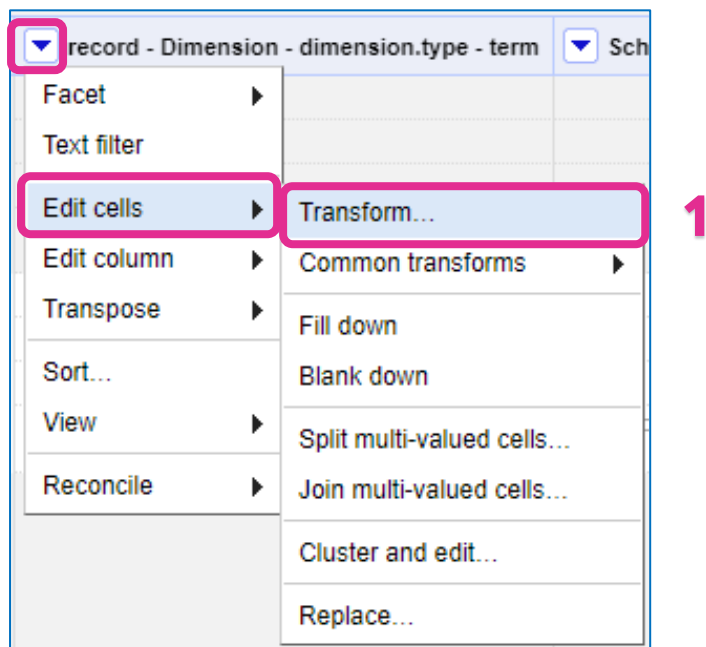
Een voorbeeld van een Adlib-tagged bestand:

```
%0 343
DA Plafondschildering direct op pleisterlaag aangebracht.
DI hoogte
DI breedte
DN Globale afmetingen gemeten vanaf de vloer.
EE cm
EE cm
IN S0003
OB allegorie (symboliek)
WA 525.0
WA 290.0
**
%0 344
DI hoogte
DI breedte
DN Afmetingen gemeten zonder lijst.
EE cm
EE cm
IN S0004
OB scheepsportretten
**
```

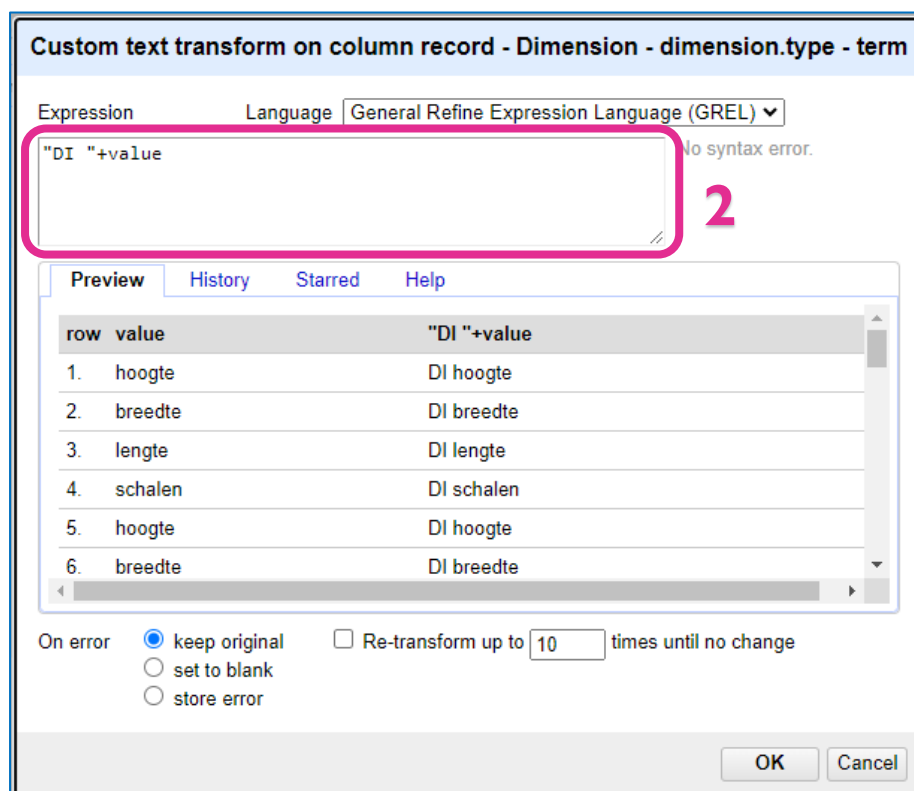
## Tags toevoegen aan veldinhoud

In deze stap gaan we de tags toevoegen aan de veldinhoud.

1. Klik op het pijltje voor de kolom **dimension.type** > **Edit cells** > **Transform**



2. Voer in **"DI "+value** (dubbele haak open tag (hoofdlettergevoelig) spatie dubbele haak sluiten plus value) (DI is de tag voor afmeting type)



3. Herhaal dit voor de velden **onderdeel, waarde, eenheid, precisie, bijzonderheden** en **vrije tekst (nieuwe kolom)**.

De overige velden gaan we niet meenemen naar het importbestand, dus daar is het op dit moment niet nodig om een actie op uit te voeren.

4. Het resultaat, maar we zijn er nog niet...

| record - Dimension - dimension.type - term | record - Dimension - dimension.value | record - Dimension - dimension.unit - term |
|--------------------------------------------|--------------------------------------|--------------------------------------------|
| DI hoogte                                  | WA 26.0                              | EE cm                                      |
| DI breedte                                 | WA 10.0                              | EE cm                                      |
| DI lengte                                  | WA 62.0                              | EE cm                                      |
|                                            |                                      | EE EE                                      |
| DI hoogte                                  | WA 110.0                             | EE cm                                      |
| DI breedte                                 | WA 25.0                              | EE cm                                      |
| DI lengte                                  | WA 107.0                             | EE cm                                      |
|                                            |                                      | EE EE                                      |

5. Zorg dat alle lege cellen ook een tag krijgen. Beweeg over een lege cel > **Edit** > *pas aan* > **Apply to all identical cells**. *Tip: gebruik Facet zoeken om lege velden op te sporen.*
6. Zorg dat eventueel dubbele tags worden opgelost: in dit voorbeeld is bijvoorbeeld “EE EE “ ontstaan, dit moet “EE “ zijn. Beweeg over een betreffende cel > **Edit** > *pas aan* > **Apply to all identical cells**. *Tip: gebruik Facet zoeken om deze dubbele tags op te sporen.*

7. Het resultaat:

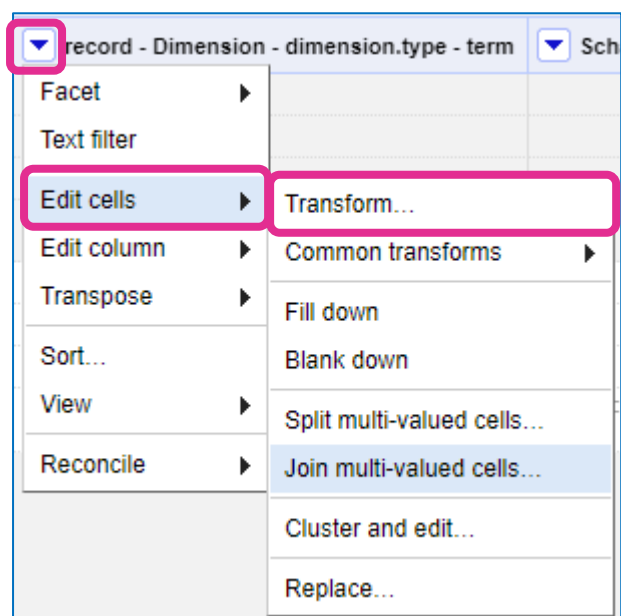
| record - Dimension - dimension.type - term | record - Dimension - dimension.value | record - Dimension - dimension.unit - term |
|--------------------------------------------|--------------------------------------|--------------------------------------------|
| DI hoogte                                  | WA 26.0                              | EE cm                                      |
| DI breedte                                 | WA 10.0                              | EE cm                                      |
| DI lengte                                  | WA 62.0                              | EE cm                                      |
| DI                                         | WA                                   | EE                                         |
| DI hoogte                                  | WA 110.0                             | EE cm                                      |
| DI breedte                                 | WA 25.0                              | EE cm                                      |
| DI lengte                                  | WA 107.0                             | EE cm                                      |
| DI                                         | WA                                   | EE                                         |

*NB. Meestal kan je systeem prima omgaan met lege regels, mocht je dat niet willen is er nog een actie nodig om de regels die nu leeg zijn eerst te verwijderen. Gebruik hiervoor facetten op elke kolom die van toepassing is en dan combineer je de facetten.*

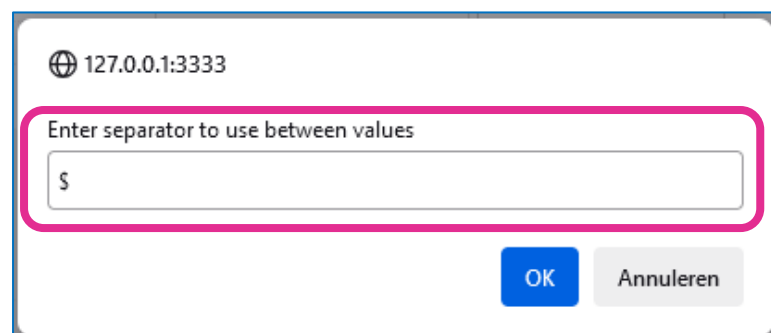
## Herhaalbare velden samenvoegen

De Record weergave van de XML-grouped is heel erg prettig omdat je alle ingevoerde velden kunt nalopen. Voor het terugplaatsen van je opgeschoonde data is het nodig om de gegevens eerst om te bouwen: alle informatie over 1 object verplaatsen naar 1 rij.

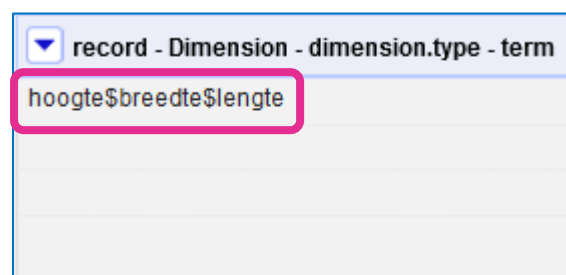
1. Klik op het pijltje voor de kolom > **Edit cells** > **Join multi-valued cells...**



2. Kies een scheidingsteken om de cellen van elkaar te onderscheiden. Let op dat het gebruikte teken niet ook al in de invoer gebruikt wordt. In dit voorbeeld gebruik ik de "\$"



3. Resultaat van de samenvoeging:

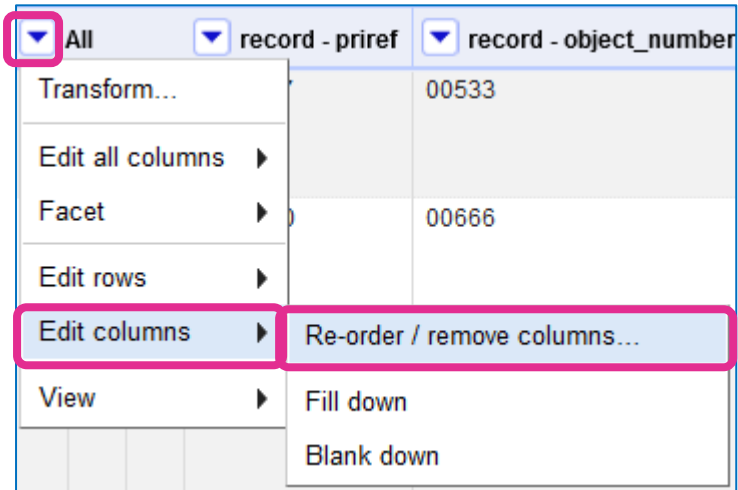


4. **Herhaal** deze stap op alle kolommen totdat je **Rows** en **Records** gelijk zijn. Dus ook de technische velden

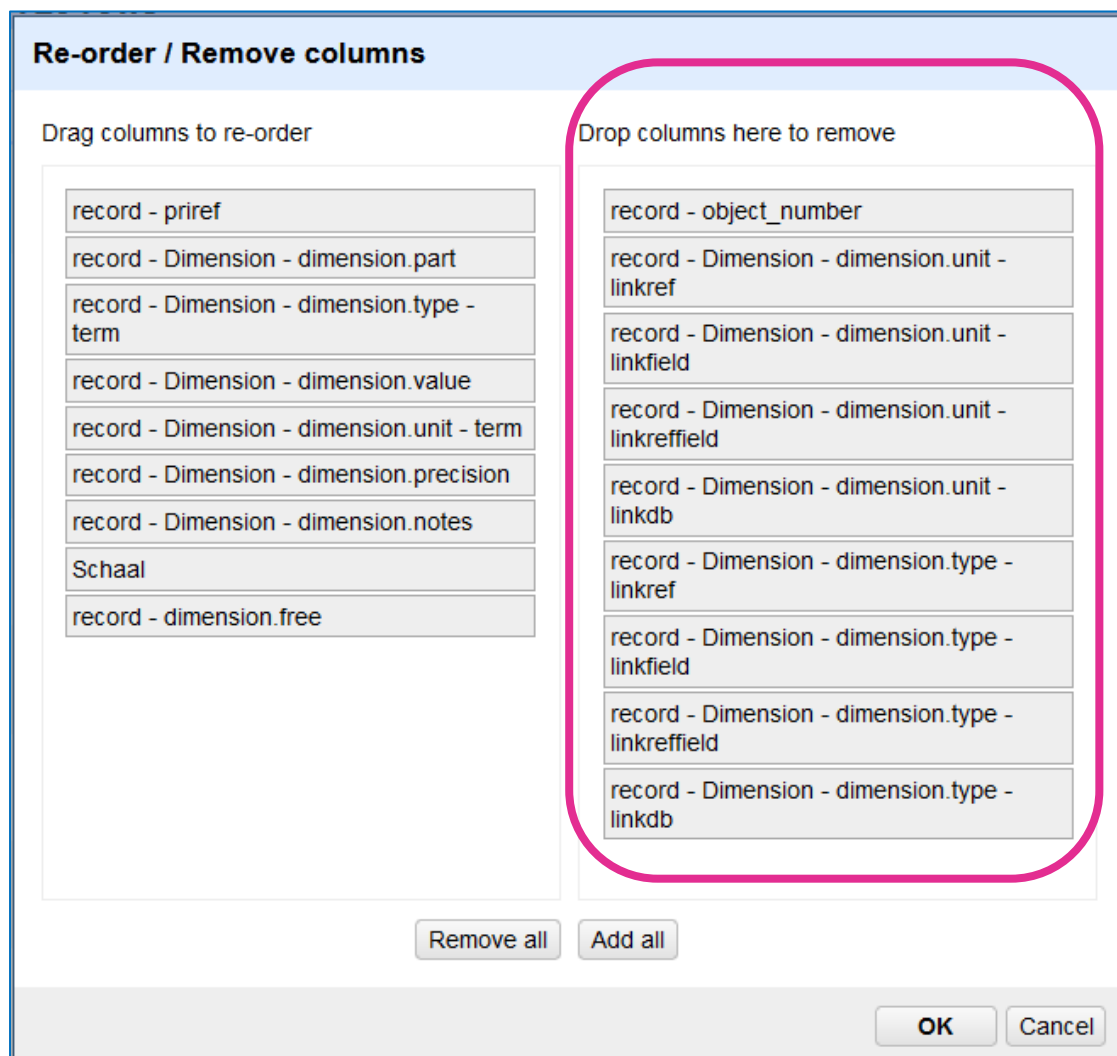
## Verwijder overbodige kolommen

Nu is het tijd om de overbodige kolommen te verwijderen

1. Klik op het pijltje voor de kolom **All** > **Edit columns** > **Re-order / remove columns...**



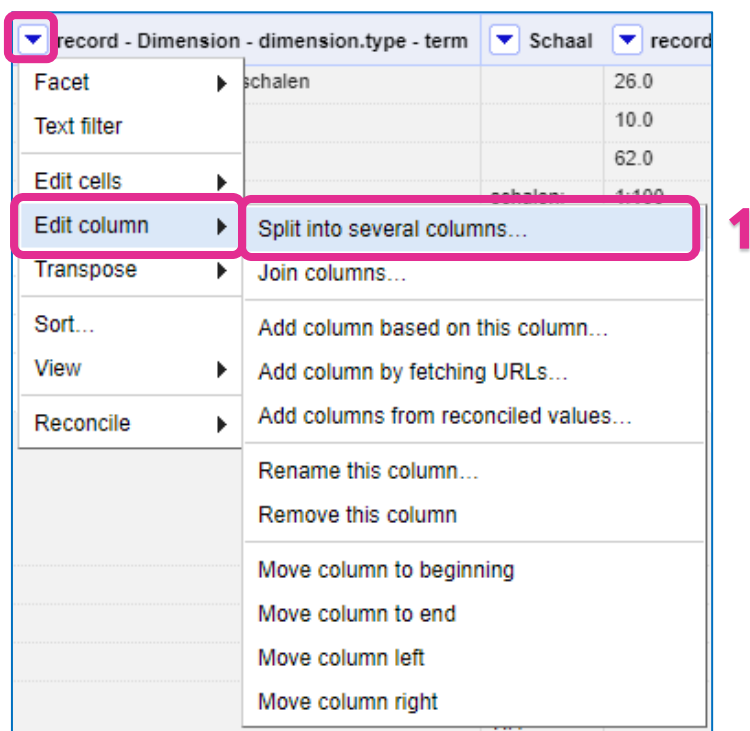
2. Schuif de kolomnamen naar de rechterhelft (drag and drop)



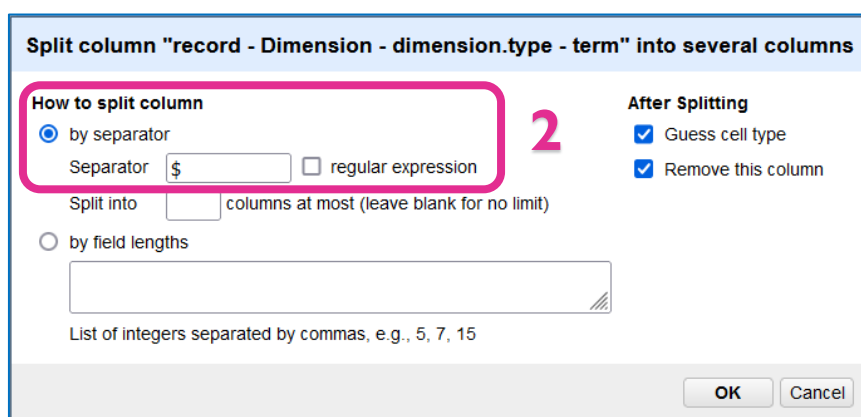
## Split into several columns

Hierna willen we graag de samengevoegde cel splitsen over meerdere kolommen zodat elke waarde weer een eigen cel krijgt, maar nu wel op dezelfde rij.

1. Klik op het pijltje voor de kolom > **Edit column** > **Split into several columns**



2. Vul het scheidingsteken weer in (in dit geval dus "\$")  
Ik kies er ook voor om de huidige kolom meteen te verwijderen



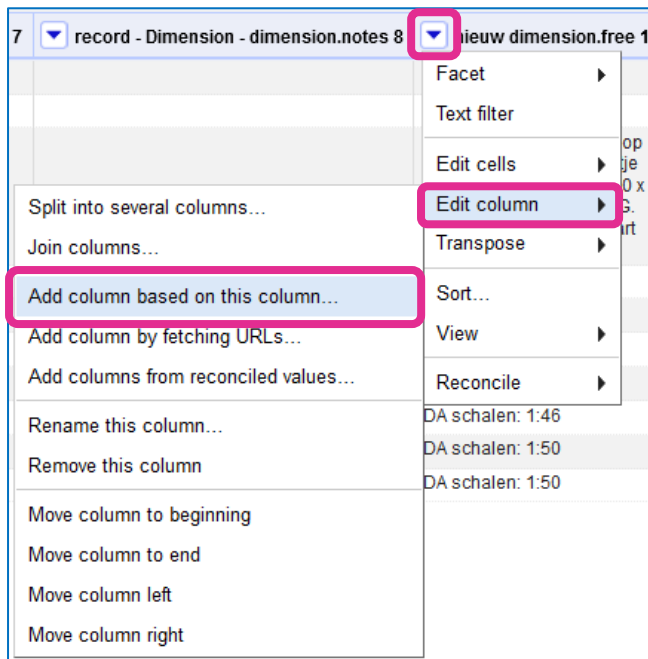
3. Resultaat > er zijn meerdere nieuwe kolommen verschenen

| record - Dimension - dimension.type - term 1 | record - Dimension - dimension.type - term 2 | record - Dimension - dimension.type - term 3 |
|----------------------------------------------|----------------------------------------------|----------------------------------------------|
| DI hoogte                                    | DI breedte                                   | DI lengte                                    |
| DI hoogte                                    | DI breedte                                   | DI lengte                                    |

## Extra kolom toevoegen

Om het einde van een record te markeren wordt \*\* (twee asterisken) gebruikt.

1. Ga naar de laatste kolom in je bestand
2. Klik op het pijltje bij de kolom > Edit column > Add column based on this column



3. Instellingen
  - a. Voeg een kolomnaam toe
  - b. Vul in plaats van *value* in “\*\*” > druk op OK

**Add column based on column nieuw dimension.free 1**

New column name:  **a**

On error: ☒ set to blank ☐ store error ☐ copy value from original column

Expression:  **b** Language:  No syntax error.

**Preview** History Starred Help

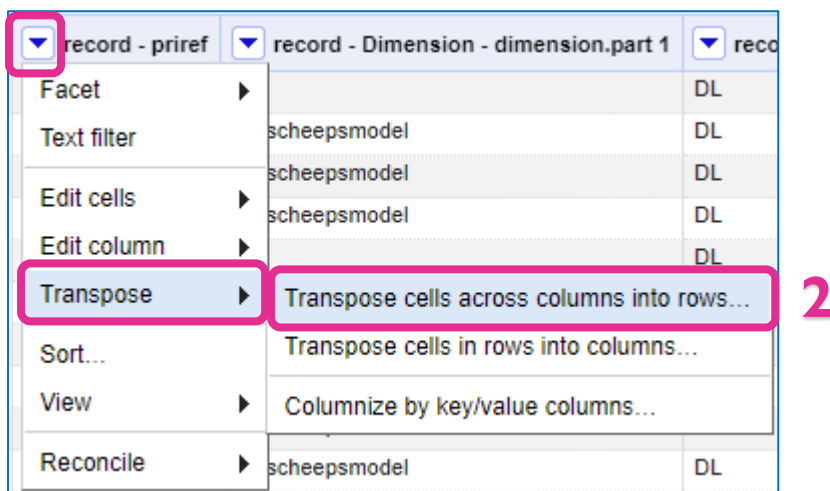
| row | value                                                                                                                                              | **** |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 1.  | DA schalen: 1:100                                                                                                                                  | **   |
| 2.  | DA schalen: 1:24                                                                                                                                   | **   |
| 3.  | DA Model verpakt in een op maat gemaakte kist. Kistje heeft een afmeting van 50 x 40 cm. Gemaakt door A.G. Hendriks in februari/maart 2013.. kist; | **   |
| 4.  | DA schalen: 1:50                                                                                                                                   | **   |
| 5.  | DA schalen: 1:20                                                                                                                                   | **   |
| 6.  | DA schalen: 1:20                                                                                                                                   | **   |

OK Cancel

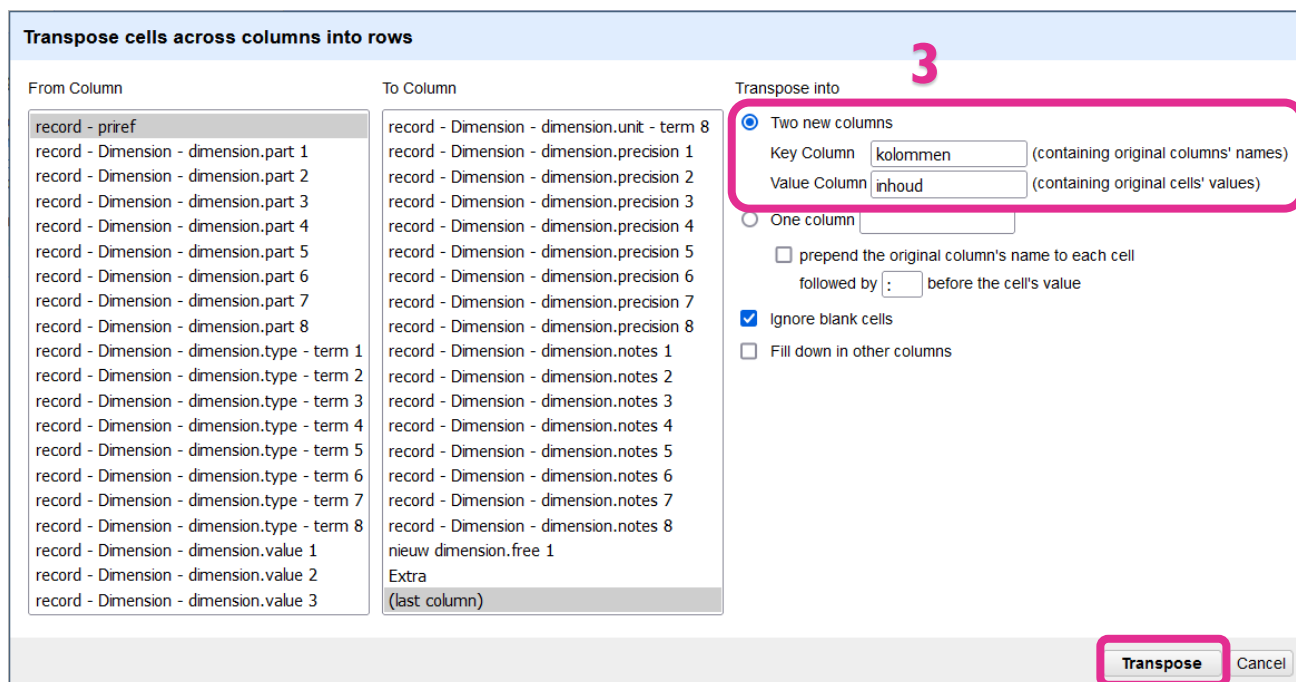
## De kolommen naar rijen

Adlib Tagged is een “plat” tekstbestand waar alle informatie onder elkaar staat. We gaan dus de kolommen omzetten naar rijen.

1. Ga naar de eerste kolom in je bestand (meestal zal dit recordnummer/priref zijn)
2. Klik op het pijltje voor de kolom > **Transpose** > **Transpose cells across columns into rows...**



3. Geef de twee nieuwe kolommen een naam



#### 4. Resultaat > er zijn nu twee kolommen

| ▼ All | ▼ kolommen | ▼ inhoud                                     |
|-------|------------|----------------------------------------------|
| ☆     | 1.         | record - priref                              |
| ☆     | 2.         | record - Dimension - dimension.part 1        |
| ☆     | 3.         | record - Dimension - dimension.part 2        |
| ☆     | 4.         | record - Dimension - dimension.part 3        |
| ☆     | 5.         | record - Dimension - dimension.part 4        |
| ☆     | 6.         | record - Dimension - dimension.type - term 1 |
| ☆     | 7.         | record - Dimension - dimension.type - term 2 |
| ☆     | 8.         | record - Dimension - dimension.type - term 3 |
| ☆     | 9.         | record - Dimension - dimension.type - term 4 |
| ☆     | 10.        | record - Dimension - dimension.value 1       |
|       |            | WA 26.0                                      |

#### 5. Verwijder de eerste kolom > **Edit column** > **Remove this column**

| ▼ kolommen         | ▼ inhoud                              |
|--------------------|---------------------------------------|
| Facet              | %0 1177                               |
| Text filter        | ension.part 1                         |
| Edit cells         | ension.part 2                         |
| <b>Edit column</b> | ension.part 3                         |
| Transpose          |                                       |
| Sort...            | Add column based on this column...    |
| View               | Add column by fetching URLs...        |
| Reconcile          | Add columns from reconciled values... |
|                    | Rename this column...                 |
|                    | <b>Remove this column</b>             |
|                    | Move column to beginning              |
|                    | Move column to end                    |
|                    | Move column left                      |
|                    | Move column right                     |

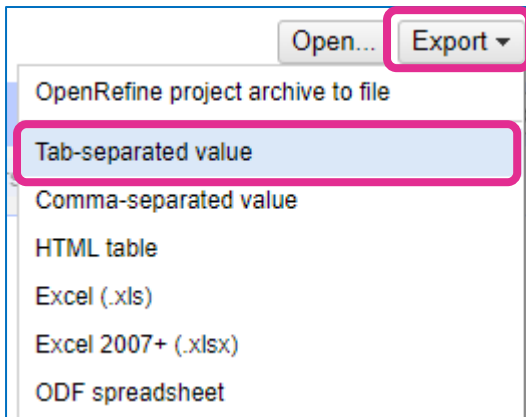
5

#### 6. Je hebt nu alle relevante gegevens onder elkaar staan in 1 kolom

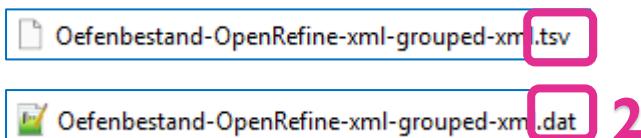
| ▼ All | ▼ inhoud |
|-------|----------|
| ☆     | 1.       |
| ☆     | 2.       |
| ☆     | 3.       |
| ☆     | 4.       |
| ☆     | 5.       |
| ☆     | 6.       |
| ☆     | 7.       |
| ☆     | 8.       |
| ☆     | 9.       |
| ☆     | 10.      |

# Importbestand maken

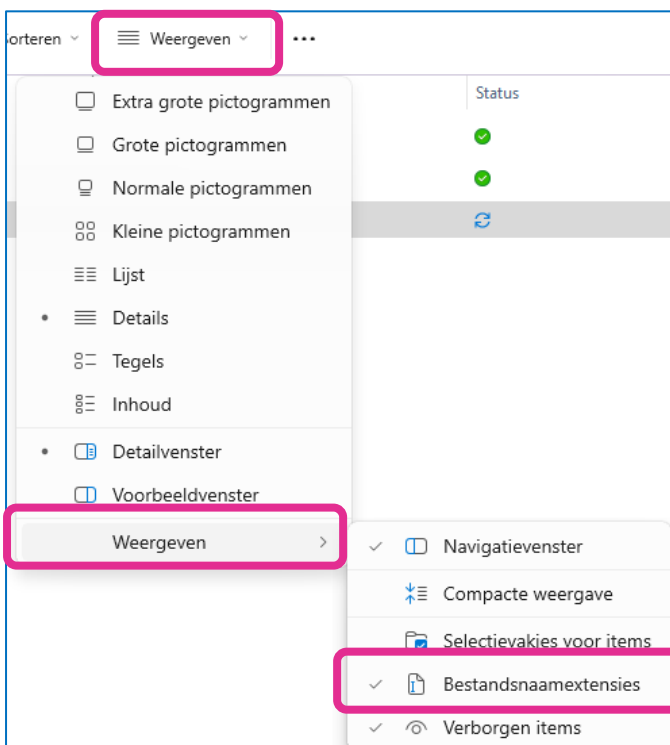
1. Ga rechtsboven in je scherm naar **Export** > **Tab-seperated value**



2. Verander de bestandsextensie handmatig van **.tsv** naar **.dat** (het bestandsformaat van een Adlib-tagged bestand)

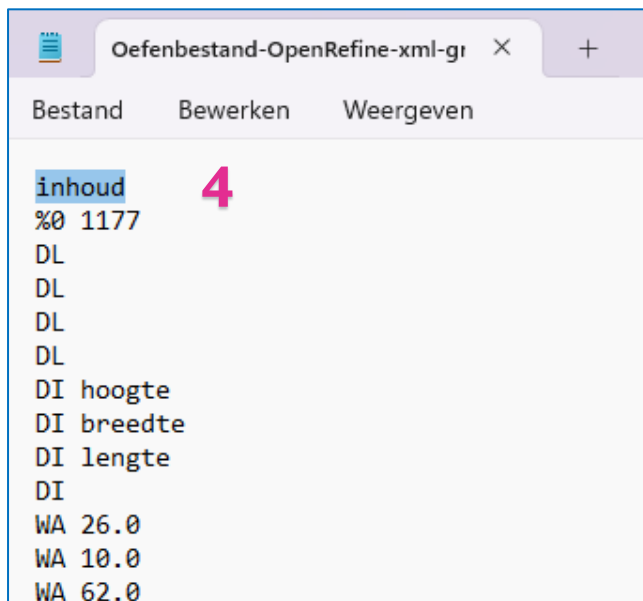


*Zie je jouw bestandsnaamextensies niet: in Windows 11 is deze instelling hier te vinden in je verkenner:*



3. Open het **.dat** bestand in **kladblok**

4. De eerste regel is je kolomnaam (in dit voorbeeld “inhoud”) > **verwijder** deze regel



```
Oefenbestand-OpenRefine-xml-gr X +
Bestand Bewerken Weergeven
inhoud
%0 1177
DL
DL
DL
DL
DI hoogte
DI breedte
DI lengte
DI
WA 26.0
WA 10.0
WA 62.0
```

5. Sla het bestand op (CTRL+S) en sluit je kladblok

6. Je Importbestand is nu klaar om te importeren!

## Links en verwijzingen

- OpenRefine user manual: <https://openrefine.org/docs>
- OpenRefine forum: <https://forum.openrefine.org/>
- OpenRefine github: <https://github.com/OpenRefine/>
- Introductie cursus OpenRefine: <https://digitaal erfgoed coach.online/online-lernen-posts/openrefine/>
- Kennisbank Meemoo: <https://kennisbank.meemoo.be/toolbox/open-refine-handleiding-voor-cultureel-erfgoed-collecties>
- Netwerk Digitaal Erfgoed: <https://netwerkdigitaalerfgoed.nl/>
- Datawerkplaatsen: <https://netwerkdigitaalerfgoed.nl/activiteiten/datawerkplaatsen/>
- Landschap Erfgoed Utrecht: <https://www.landschaperfgoedutrecht.nl/>
- <https://www.landschaperfgoedutrecht.nl/erfgoed/digitalisering/>

---

Alle handleidingen zijn te vinden op: [www.kubus.frl/openrefine](http://www.kubus.frl/openrefine)

