

Data Marketplace - Step-by-step guidelines for data sellers

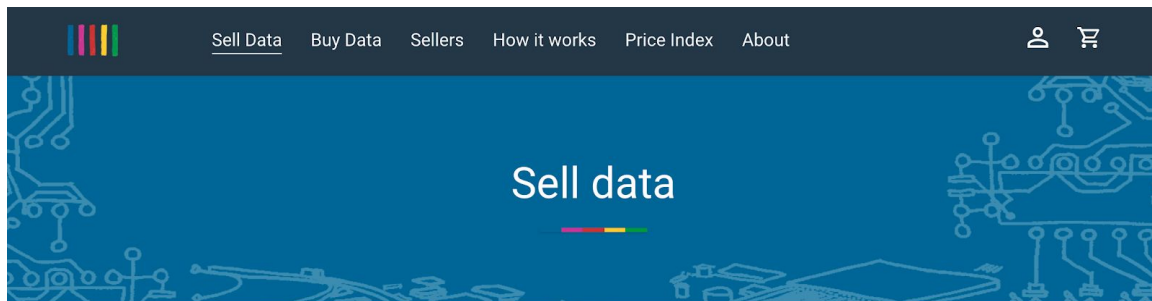
Step 1. SELL DATA

Go to the [Sell Data section](#) of the Data Marketplace

Step 2. DOCUMENT UPLOAD

Upload a TMX file by dragging the file or by clicking on the *browse* button. The file format should follow the specifications below:

- The TMX document should be **bilingual**, i.e. all <tu/> tag should contain exactly 2 variants (<tuv/>).
- The TMX document should not define the value ***all*** for its <header/> tag **srclang** attribute.
- The TMX document should not have <tu/> tags with srclang that differs from the <header/>.
- The file should be **UTF-8** encoded
- The TMX file should be conformant with the [TMX Gala specification TMX 1.4b](#)



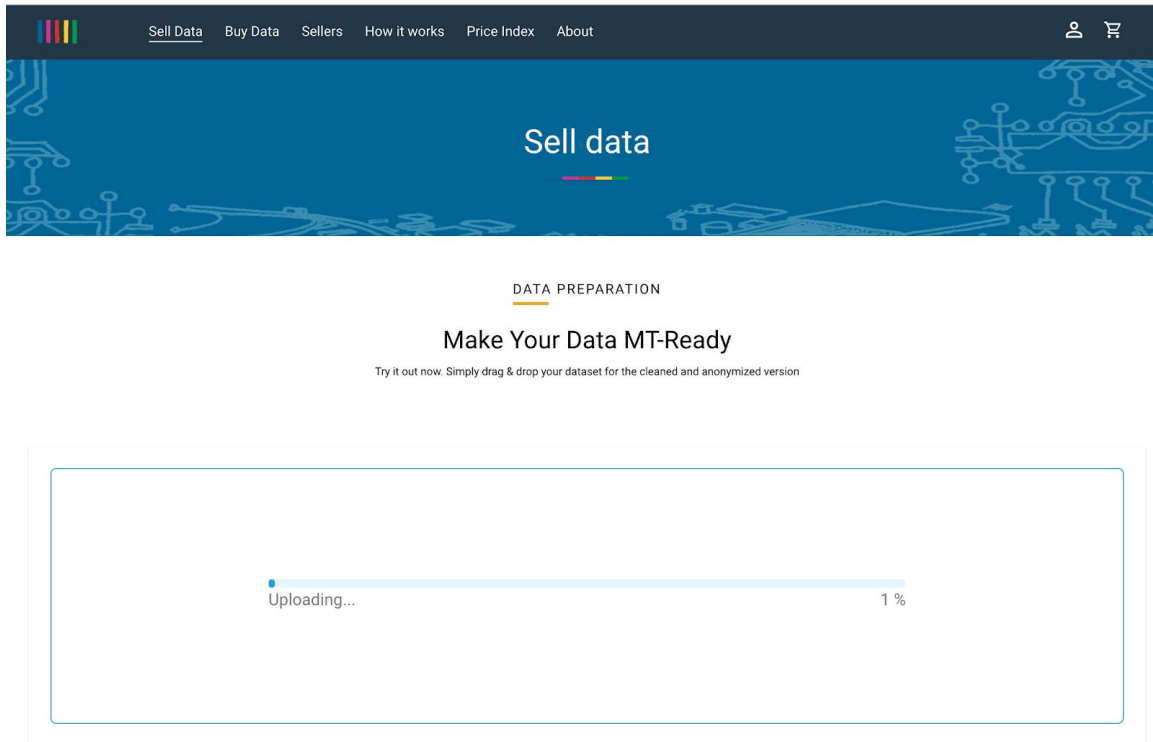
DATA PREPARATION

Make Your Data MT-Ready

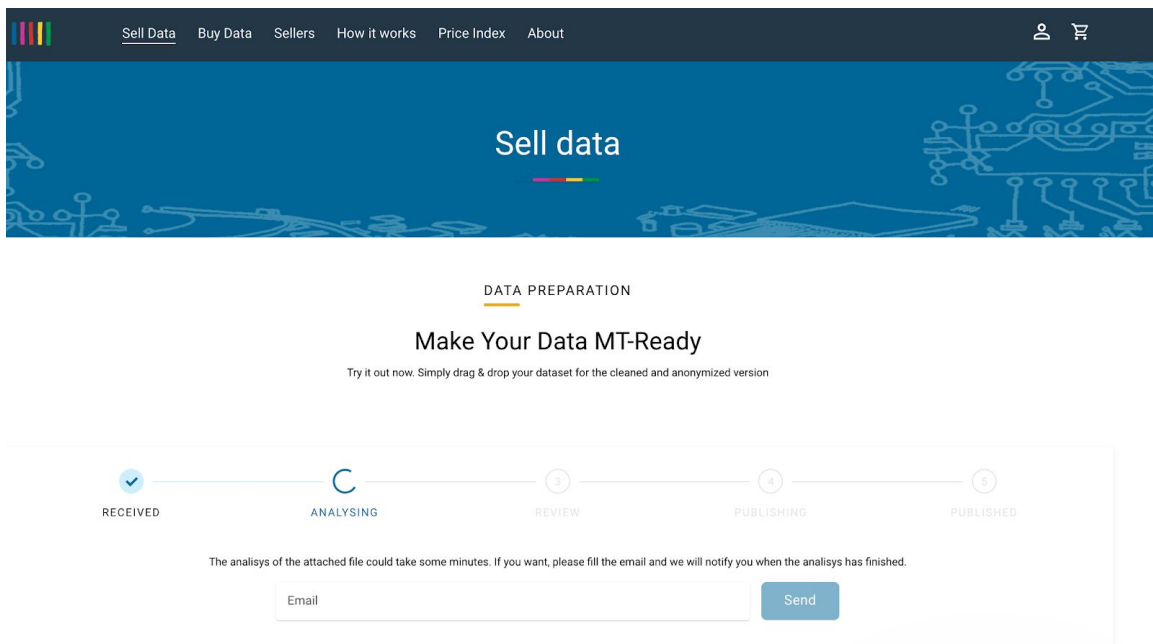
Try it out now. Simply drag & drop your dataset for the cleaned and anonymized version

A screenshot of the TMX upload interface and data preparation services. On the left, there is a blue box with a white dashed border containing a document icon with 'TMX' on it and an upward arrow. Below the icon, it says 'Upload a document (*.TMX)' and 'Drag bilingual TMX format file here or [browse](#)'. On the right, there is a white box with a blue border containing three service items, each with a yellow checkmark: 'CLEAN' (Prepare your data with free automatic smart deduplication, language inconsistency detection and identification of fluency or adequacy errors.), 'ANONYMIZE' (Use our free anonymization service to protect privacy by filtering out named entities and personal information in your data.), and 'PUBLISH' (Upload to see the clean and anonymised version of your data. Publish the data to get these services for free. Uploaded data is not stored unless published.).

Note: Please note that you can only upload one file at a time. For uploading multiple TMX files at the same time, please open the upload page on multiple tabs in your browser and initiate the upload in each browser tab.



If the upload takes longer, you'll have the option to leave your email and be notified when the upload has completed.



Step 3. ANALYSIS & REVIEW

In this step you can see the counts and the quality of the data uploaded, as follows:

- Low-quality - Segments that haven't passed the filtering
- Replica - Segments that already exist in the Data Marketplace
- High-quality - Unique segments that passed the filtering

DATA PREPARATION

Make Your Data MT-Ready

Try it out now. Simply drag & drop your dataset for the cleaned and anonymized version

The screenshot displays a progress bar with five stages: RECEIVED, ANALYSING, REVIEW (current), PUBLISHING, and PUBLISHED. Below the progress bar, the interface shows the source languages: Italian (Italy) and Bulgarian (Bulgaria). A file named 'LISA-TMX_意大利语 (意大利)_保加利亚语 (保加利亚) ...' with a size of 28.47 Mb is listed. The data analysis results are summarized in a donut chart and text: Total segments: 35114 (954077 it-IT words, 961060 bg-BG words). The breakdown is: 32204 High-quality segments, 5 Replica segments, and 2905 Low-quality segments. A 'View sample' link is provided. A 'Next' button is located at the bottom.

Category	Count
Total segments	35114
High-quality segments	32204
Replica segments	5
Low-quality segments	2905

Step 4. BUY BACK or PUBLISH

In this step you can decide whether to *Buy back* the cleaned data or *Publish* it to the Data Marketplace. The option to buy back the data is not available yet.

RECEIVED ANALYSING **3** REVIEW PUBLISHING PUBLISHED

Italian (Italy) Bulgarian (Bulgaria)

LISA-TMX_意大利语 (意大利)_保加利亚语 (保加利亚) ...
28.47 Mb

35114
Total segments

954077 it-IT words
961060 bg-BG words

- 32204 High-quality segments
- 5 Replica segments
- 2905 Low-quality segments

[View sample](#)

1

Would you like to purchase the cleaned and anonymized (coming soon) version of your dataset or publish it for sale on the Data Marketplace?

Data Sellers

You can purchase the cleaned version and anonymized (coming soon) of your dataset directly without sharing it with a third party or publishing it for sale.

[Buy back](#)

Data Publishers

If you publish the cleaned and anonymized version of your dataset for sale on the Data Marketplace, you can download it for your own use for free.

[Publish](#)

Step 5. DOMAIN & CONTENT TYPE

By clicking on *Publish*, you will be asked to select domain and content type:

RECEIVED ANALYSING **3** REVIEW PUBLISHING PUBLISHED

Italian (Italy) Bulgarian (Bulgaria)

LISA-TMX_意大利语 (意大利)_保加利亚语 (保加利亚) ...
28.47 Mb

35114
Total segments

954077 it-IT words
961060 bg-BG words

- 32204 High-quality segments
- 5 Replica segments
- 2905 Low-quality segments

[View sample](#)

1

Please select the domain and the content type of your dataset for a more accurate price suggestion.

Domain

- Aerospace/Aviation
- Automotive
- Business Services
- Chemicals

Content Type

Next

Step 6. PRICE DEFINITION

Price suggestion comes from our algorithms, and you can see the examples of calculations on the [Price Index Table](#). You are free to define your own price, for as long as it is between 0.0001-0.1 price/word.

Marketplace suggested price

€ 0.005 [Use this price](#)

Suggested price / word

Common

35114 segments 954077 it-IT words
961060 bg-BG words

The suggested price is based on the [Price Index Table](#)

Total document price

Price per word: € 0.1

Service fee (15%): 15%

Note: Fee is not charged until 31 October 2021.

Total Price: € 93,067.50

Step 6. ACCOUNT CREATION

In this step, you can create an account or *Sign in* if you already have an account.

Italian (Italy) Bulgarian (Bulgaria)

LISA-TMX_意大利语 (意大利)_保加利亚语 (保加利亚) ...
28.47 Mb

Automotive Keywords/SEO

€ 0.10
Price per word

35114
Total segments
954077 it-IT words
961060 bg-BG words

- 32204 High-quality segments
- 5 Replica segments
- 2905 Low-quality segments

[View sample](#)

SIGN UP

Do you have an account? [Sign in](#)

Prefix First name

Last name

Email

Password

Retype password

Country

- I agree to the [Terms of Use](#) *
- I agree to the [Privacy Policy](#) *

Sign up

Step 7. DOCUMENT DESCRIPTION

In this step, you can add a short description of the document to help potential buyers better understand its contents. You can also *Skip* this step.


Italian (Italy)
Italy

Bulgarian (Bulgaria)
Bulgaria

LISA-TMX_意大利语 (意大利) _保加利亚语 (保加利亚) ...
28.47 Mb

Automotive Keywords/SEO


€ 0.10
Price per word



35114
Total segments
954077 it-IT words
961060 bg-BG words

- 32204 High-quality segments
- 5 Replica segments
- 2905 Low-quality segments

[View sample](#)

 Add a description of your dataset to make it stand out more in the Data Marketplace.

Type the document description

[Skip](#)

Step 8. PUBLISHING

Publish by clicking on the *Publish* button.

RECEIVED ANALYSING **3** REVIEW PUBLISHING PUBLISHED


Italian (Italy)
Italy

Bulgarian (Bulgaria)
Bulgaria

LISA-TMX_意大利语 (意大利) _保加利亚语 (保加利亚) ...
28.47 Mb

Business Services Audiovisual

€ 0.10
Price per word



35114
Total segments
954077 it-IT words
961060 bg-BG words

- 32204 High-quality segments
- 5 Replica segments
- 2905 Low-quality segments

[View sample](#)

[Publish](#)

With this step the publishing is completed. Your document will be uploaded in the Data Marketplace and will appear for sale when the Data Marketplace is launched.

