

lesSDRF

Welcome to an overview of lesSDRF – a user-centric tool designed to streamline the creation of SDRF files.

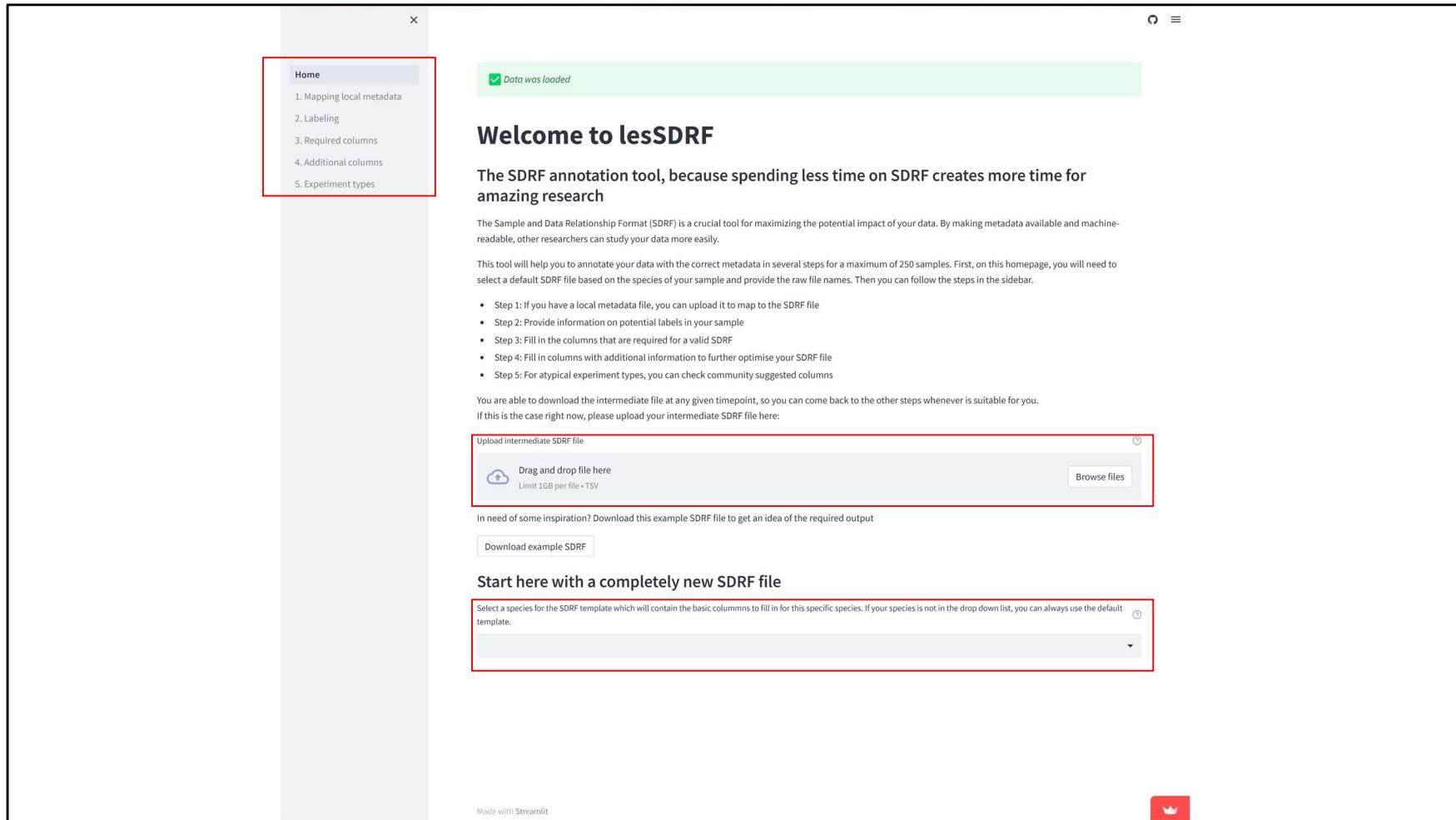
Throughout this manual, we'll delve into its features, functionalities and how it simplifies the complex task of proteomics metadata annotation.

You can access the tool via: <https://lessdrf.streamlit.app/>

Home: Select SDRF template and input the raw file names

The home screen invites users to select an SDRF template, which is derived from the original SDRF initiative and is species-specific. By choosing this template, users can be assured that they will include the required columns for a valid SDRF.

Furthermore, the home screen offers an option to upload an intermediate SDRF file. This feature enables users to pause their annotation process and resume at a later time. Since it's a browser-based tool, the SDRF file isn't saved upon closing the browser. This capability also streamlines the annotation process for similar projects.



This is the home screen. On the left, the various stages of the tool are listed.

Currently displayed on the home screen is an upload feature for intermediate SDRF files. This facilitates the uploading of previously created SDRF files to be further refined using lesSDRF. These could be SDRF files from an earlier lesSDRF session or partial SDRFs exported from other software.

For those initiating a new SDRF, it's essential to choose a species from the dropdown menu. This action will select an SDRF template, ensuring it encompasses all the required columns for a valid SDRF.

You are able to download the intermediate file at any given timepoint, so you can come back to the other steps whenever is suitable for you.
If this is the case right now, please upload your intermediate SDRF file here:

Upload intermediate SDRF file

Drag and drop file here
Limit 1GB per file - TSV

Browse files

In need of some inspiration? Download this example SDRF file to get an idea of the required output

Download example SDRF

Start here with a completely new SDRF file

Select a species for the SDRF template which will contain the basic columns to fill in for this specific species. If your species is not in the drop down list, you can always use the default template.

human

Input raw file names as a comma or tab separated list

Prosser_1004.raw Prosser_1085.raw Prosser_1086.raw Prosser_1119.raw Prosser_1136.raw Prosser_1146.raw Prosser_1195.raw Prosser_1198.raw Prosser_1200.raw Prosser_1205.raw Prosser_1209.raw Prosser_1228.raw Prosser_1273.raw Prosser_1288.raw Prosser_1290.raw Prosser_1304.raw Prosser_1311.raw Prosser_1320.raw Prosser_1329.raw Prosser_1356.raw Prosser_1366.raw Prosser_1375.raw Prosser_1379.raw Prosser_1383.raw Prosser_1425.raw Prosser_1427.raw Prosser_1462.raw Prosser_1467.raw Prosser_1472.raw Prosser_1510.raw Prosser_1535.raw Prosser_1602.raw Prosser_1605.raw

replicate	technology type	assay name	comment[technical replicate]	comment[data file]	comment[fraction identifier]	comment[[label]]	comment[instrument]	comment[notes]
0	None	None	None	Prosser_1004.raw	None	None	None	None
1	None	None	None	Prosser_1085.raw	None	None	None	None
2	None	None	None	Prosser_1086.raw	None	None	None	None
3	None	None	None	Prosser_1119.raw	None	None	None	None
4	None	None	None	Prosser_1136.raw	None	None	None	None
5	None	None	None	Prosser_1146.raw	None	None	None	None
6	None	None	None	Prosser_1195.raw	None	None	None	None
7	None	None	None	Prosser_1198.raw	None	None	None	None
8	None	None	None	Prosser_1200.raw	None	None	None	None
9	None	None	None	Prosser_1205.raw	None	None	None	None

Made with Streamlit

After choosing an SDRF template, the user must enter the raw file names. These can be provided in either comma-separated or tab-separated formats, allowing for direct copying from a local Excel file, for example. Once entered, the raw file names are promptly populated in the "comment[data file]" column, as illustrated below.

Currently the number of raw files is restricted to prevent potential lagging of the tool.

Step 1: Map local metadata file to SDRF columns

Recognizing that users frequently have some local metadata gathered during their experiments, the initial step in lesSDRF facilitates the uploading and mapping of this local metadata to the SDRF columns established on the Home screen. Once mapped, the system verifies the local metadata for ontology compatibility.

1. Map local metadata to SDRF

If you have a local metadata file available, you can use this file to map the data to the required SDRF information.

Important: you can upload the file in csv, tsv or xlsx format.
The order of your raw file names should match the order in which you inputted them in the previous step

This is your current SDRF file.

	source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry category]	characteristics[age]	characteristic
0	None	None	None	None	None	None	None
1	None	None	None	None	None	None	None
2	None	None	None	None	None	None	None
3	None	None	None	None	None	None	None
4	None	None	None	None	None	None	None
5	None	None	None	None	None	None	None
6	None	None	None	None	None	None	None
7	None	None	None	None	None	None	None
8	None	None	None	None	None	None	None
9	None	None	None	None	None	None	None

Upload your local metadata file (.csv, .tsv or .xlsx)

Drag and drop file here
Limit 1GB per file • CSV, TSV, XLSX

Browse files

local_metadata.xlsx 9.0KB

Your metadata file:

	source name	Organism	part	sample
0	Sample 1	Homo sapiens	heart LV	Prosser_1004.raw
1	Sample 2	Homo sapiens	heart LV	Prosser_1085.raw
2	Sample 3	Homo sapiens	heart LV	Prosser_1086.raw
3	Sample 4	Homo sapiens	heart LV	Prosser_1119.raw
4	Sample 5	Homo sapiens	heart LV	Prosser_1136.raw
5	Sample 6	Homo sapiens	heart LV	Prosser_1146.raw
6	Sample 7	Homo sapiens	heart LV	Prosser_1195.raw
7	Sample 8	Homo sapiens	heart LV	Prosser_1198.raw
8	Sample 9	Homo sapiens	heart LV	Prosser_1200.raw

Upload local metadata in csv, tsv or xlsx format. Upon upload the data will be visualised below.

The screenshot shows a Streamlit application window titled "1. Mapping local metadata". On the left, a sidebar lists steps: Home, 1. Mapping local metadata (which is active), 2. Labeling, 3. Required columns, 4. Additional columns, and 5. Experiment types. Below this is a button "Press to download SDRF file".

The main area has a header "Upload your local metadata file (.csv, .tsv or .xlsx)". It includes a "Drag and drop file here" field with a 1GB limit for CSV, TSV, and XLSX files, and a "Browse files" button. A file named "local_metadata.xlsx" (9.6KB) is listed.

Below this is a table titled "Your metadata file:" showing 10 rows of data:

	source name	Organism	part	sample
0	Sample 1	Homo sapiens	heart LV	Prosser_1004.raw
1	Sample 2	Homo sapiens	heart LV	Prosser_1085.raw
2	Sample 3	Homo sapiens	heart LV	Prosser_1086.raw
3	Sample 4	Homo sapiens	heart LV	Prosser_1119.raw
4	Sample 5	Homo sapiens	heart LV	Prosser_1136.raw
5	Sample 6	Homo sapiens	heart LV	Prosser_1146.raw
6	Sample 7	Homo sapiens	heart LV	Prosser_1195.raw
7	Sample 8	Homo sapiens	heart LV	Prosser_1198.raw
8	Sample 9	Homo sapiens	heart LV	Prosser_1200.raw
9	Sample 10	Homo sapiens	heart LV	Prosser_1205.raw

Below the table is a red-bordered section titled "Select columns containing data that will be used in the SDRF data". It contains three dropdown menus: "source name", "Organism", and "part", all currently selected. To the right is a checkbox "Ready to match?" which is checked.

Below this section are three dropdown menus for matching columns:

- Select column 1 to match in your metadata file: (empty dropdown)
- Select column 2 to match in your metadata file: (empty dropdown)
- Select column 3 to match in your metadata file: (empty dropdown)

At the bottom left is a "Made with Streamlit" footer, and at the bottom right is a small red icon with a white crown.

On this page, users can initiate by choosing the local metadata columns they wish to incorporate into their SDRF file. This choice prompts a subsequent dropdown menu where these columns can be selected.

The screenshot shows the lesSDRF application interface. On the left, a sidebar lists steps: Home, 1. Mapping local metadata (selected), 2. Labeling, 3. Required columns, 4. Additional columns, and 5. Experiment types. Below this is a button to 'Press to download SDRF file'. The main area has a header 'Upload your local metadata file (.csv, .tsv or .xlsx)'. A 'Drag and drop file here' section shows a file named 'local_metadata.xlsx' (9.6KB). Below it, a table titled 'Your metadata file:' lists 10 samples from 'Sample 1' to 'Sample 10', each with 'Homo sapiens' as the organism and 'heart LV' as the part. A message 'Ready to match?' with a checked checkbox is shown above a matching interface. The interface includes dropdowns for 'source name', 'Organism', and 'part', all with 'x' buttons. To the right, a red box highlights a dropdown for 'Organism' with options: 'characteristics[organism]', 'characteristics[disease]', 'characteristics[individual]', and 'characteristics[organism part]'. Another red box highlights a dropdown for 'part' with options: 'characteristics[organism]', 'characteristics[sex]', 'characteristics[enrichment pro...]', and 'characteristics[compound]'. To the right of these dropdowns, two checkboxes are shown: 'Match and check ontology' (checked) and 'Match and check ontology'. A green box with a checkmark and the text 'Great! The local metadata values are valid terms and are mapped to the SDRF file.' is visible.

Once the local metadata columns are paired with the SDRF columns, a "match and check ontology" checkbox becomes visible. Activating this checkbox prompts lesSDRF to conduct an ontology verification. If the ontology check passes, a green box is displayed.

The screenshot shows the 'Mapping local metadata' step of the lesSDRF application. On the left, a sidebar lists steps: Home, 1. Mapping local metadata (selected), 2. Labeling, 3. Required columns, 4. Additional columns, 5. Experiment types, and a link to download the SDRF file. The main area has a title 'Upload your local metadata file (.csv, .tsv or .xlsx)' with a 'Browse files' button and a 'local_metadata.xlsx 9.6kB' file listed. Below is a table titled 'Your metadata file:' showing 10 rows of data:

	source name	Organism	part	sample
0	Sample 1	Homo sapiens	heart LV	Prosser_1004.raw
1	Sample 2	Homo sapiens	heart LV	Prosser_1085.raw
2	Sample 3	Homo sapiens	heart LV	Prosser_1086.raw
3	Sample 4	Homo sapiens	heart LV	Prosser_1119.raw
4	Sample 5	Homo sapiens	heart LV	Prosser_1136.raw
5	Sample 6	Homo sapiens	heart LV	Prosser_1146.raw
6	Sample 7	Homo sapiens	heart LV	Prosser_1195.raw
7	Sample 8	Homo sapiens	heart LV	Prosser_1198.raw
8	Sample 9	Homo sapiens	heart LV	Prosser_1200.raw
9	Sample 10	Homo sapiens	heart LV	Prosser_1205.raw

Below the table, a section titled 'Select columns containing data that will be used in the SDRF data' shows three dropdowns: 'source name', 'Organism', and 'part'. To the right is a checkbox labeled 'Ready to match?' which is checked. Further down, there are three pairs of dropdowns for matching columns:

- source name (local) ↔ source name (SDRF)
- Organism (local) ↔ characteristics[organism] (SDRF)
- part (local) ↔ characteristics[organism part] (SDRF)

Each pair has a checkbox labeled 'Match and check ontology' next to it. To the right of these dropdowns is a red-bordered box containing three status messages:

- A green box: 'Great! The local metadata values are valid terms and are mapped to the SDRF file.'
- A green box: 'Great! The local metadata values are valid terms and are mapped to the SDRF file.'
- A red box: '✗ ('heart LV') are not ontology terms. Select the correct terms in the next steps directly from the ontology'

At the bottom left is a 'Made with Streamlit' footer.

Once the local metadata columns are paired with the SDRF columns, a "match and check ontology" checkbox becomes visible. Activating this checkbox prompts lesSDRF to conduct an ontology verification. If the ontology check passes, a green box is displayed. However, if there's a discrepancy, a red box surfaces, highlighting the mismatched term and signaling to the user the specific term that needs correction.

Step 2: Select labeling approach

Labeling is an essential component of metadata. In this section, we'll illustrate both a label-free method and a TMT-labeled experiment using the same dataset to showcase the differential impact on the SDRF file structure.

When the label-free option is chosen, it's simply appended to the "comment[label]" column. In contrast, when a labeling method like TMT is selected, users must map each raw file to the appropriate label. Given that a single file can encompass multiple labels, this necessitates the generation of a new row in the SDRF file for each distinct file-label pairing.

2. Labeling

If a raw file contains multiple labels, every label will need to be annotated on a different row. Here you can map label information to your raw files. As a result, the raw file information will be duplicated with the correct label filled in.

This is your current SDRF file.

	source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry category]	characteristics[age]	characteristic
0	Sample 1	Homo sapiens	None	None	None	None	None
1	Sample 2	Homo sapiens	None	None	None	None	None
2	Sample 3	Homo sapiens	None	None	None	None	None
3	Sample 4	Homo sapiens	None	None	None	None	None
4	Sample 5	Homo sapiens	None	None	None	None	None
5	Sample 6	Homo sapiens	None	None	None	None	None
6	Sample 7	Homo sapiens	None	None	None	None	None
7	Sample 8	Homo sapiens	None	None	None	None	None
8	Sample 9	Homo sapiens	None	None	None	None	None
9	Sample 10	Homo sapiens	None	None	None	None	None

Input the label that was used in your experiment. If no label was added, indicate this using *label free sample*.

Select your matching ontology term using this autocomplete function

Choose an option

Or follow the ontology based drop down menu below

label free sample
➤ TMT
➤ iTRAQ
➤ SILAC
➤ ICAT
➤ metabolic label
➤ Stable isotope dimethyl labeling

Submit selection

Match the filenames to the labels you selected above. Select ALL if the label is found in all raw files.

Ready?

Here users can select the type of labeling present in their dataset using an autocomplete search function or a drop down menu containing the ontology terms.

Label free experiment

2. LABELING

If a raw file contains multiple labels, every label will need to be annotated on a different row. Here you can map label information to your raw files. As a result, the raw file information will be duplicated with the correct label filled in.

This is your current SDRF file.

	source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry category]	characteristics[age]	characteristic
0	Sample 1	Homo sapiens	None	None	None	None	None
1	Sample 2	Homo sapiens	None	None	None	None	None
2	Sample 3	Homo sapiens	None	None	None	None	None
3	Sample 4	Homo sapiens	None	None	None	None	None
4	Sample 5	Homo sapiens	None	None	None	None	None
5	Sample 6	Homo sapiens	None	None	None	None	None
6	Sample 7	Homo sapiens	None	None	None	None	None
7	Sample 8	Homo sapiens	None	None	None	None	None
8	Sample 9	Homo sapiens	None	None	None	None	None
9	Sample 10	Homo sapiens	None	None	None	None	None

Input the label that was used in your experiment. If no label was added, indicate this using *label free sample*.

Select your matching ontology term using this autocomplete function
Or follow the ontology based drop down menu below

Choose an option

label free sample

> IMI

> iTRAQ

> SILAC

ALL

Prosser_1004.raw
Prosser_1085.raw
Prosser_1086.raw
Prosser_1119.raw
Prosser_1136.raw
Prosser_1146.raw
Drosophila_1106.raw

Choose an option

Ready?

For a label-free experiment, users should opt for 'label free sample'. Subsequently, they can pair this choice with specific raw files. However, if all files share this label (or lack thereof), the 'ALL' option can be chosen from the dropdown menu.

The screenshot shows a Streamlit application for labeling mass spectrometry raw files. The sidebar on the left lists steps: Home, 1. Mapping local metadata, 2. Labeling (selected), 3. Required columns, 4. Additional columns, 5. Experiment types, and a button to Press to download SDRF file.

In the main panel, there is a table with one row:

9	Sample 10	Homo sapiens	None	None	None	None	None
---	-----------	--------------	------	------	------	------	------

Below the table, instructions say: "Input the label that was used in your experiment. If no label was added, indicate this using *label free sample*." A dropdown menu says "Choose an option" and has a checked checkbox for "label free sample". Other options include TMT, iTRAQ, SILAC, ICAT, metabolic label, and Stable isotope dimethyl labeling. A "Submit selection" button is present.

Below this, a section titled "Match the filenames to the labels you selected above. Select ALL if the label is found in all raw files." contains a red box around a dropdown menu labeled "Select files labeled with label free sample." It shows "ALL" and a checked "Ready?" checkbox.

At the bottom, a table titled "SDRF file with label information" shows two sets of data. The first set (left) has columns: ics[biological replicate], technology type, assay name, and comment[technical replicate]. The second set (right) has columns: comment[data file], comment[fraction identifier], comment[label], and comment[in]. Both sets have rows from 0 to 9. The "comment[label]" column for the right set is highlighted with a red box and contains "label free sample" for all entries.

At the bottom left is a "Made with Streamlit" logo, and at the bottom right is a small red icon.

When the "Ready?" checkbox is selected, "label free sample" will be populated in the "comment[label]" column for each raw file.

Labeled experiment

Home
1. Mapping local metadata
2. Labeling
3. Required columns
4. Additional columns
5. Experiment types
Press to download SDRF file

#	Sample	Species	Label	Label	Label	Label
1	Sample 11	Homo sapiens	None	None	None	None
2	Sample 12	Homo sapiens	None	None	None	None
3	Sample 14	Homo sapiens	None	None	None	None
4	Sample 18	Homo sapiens	None	None	None	None
5	Sample 24	Homo sapiens	None	None	None	None
6	Sample 26	Homo sapiens	None	None	None	None
7	Sample 28	Homo sapiens	None	None	None	None
8	Sample 31	Homo sapiens	None	None	None	None
9	Sample 33	Homo sapiens	None	None	None	None

Input the label that was used in your experiment. If no label was added, indicate this using *label free sample*.

Select your matching ontology term using this autocomplete function
Or follow the ontology based drop down menu below

Choose an option

label free sample

TMT
 TMT126
 TMT127
 TMT127C
 TMT127N
 TMT128
 TMT128C
 TMT128N
 TMT129
 TMT129C
 TMT129N
 TMT130
 TMT130C
 TMT130N
 TMT131
 TMT131N
 TMT131C
 TMT132N
 TMT132C
 TMT133N
 TMT133C
 TMT134N
➤ iTRAQ
➤ SILAC
➤ ICAT
➤ metabolic label
➤ Stable isotope dimethyl labeling

For a labeled experiment, users can choose the specific labels they've used. For instance, they might select TMT126, TMT127, and TMT128.

Home

1. Mapping local metadata

2. Labeling

3. Required columns

4. Additional columns

5. Experiment types

Press to download SDRF file

TMT130
TMT130C
TMT130N
TMT131
TMT131N
TMT131C
TMT132
TMT132C
TMT133
TMT133N
TMT133C
TMT134N

ITRAQ
SILAC
ICAT
metabolic label
Stable isotope dimethyl labeling

Submit selection

Match the filenames to the labels you selected above. Select ALL if the label is found in all raw files.

Select files labeled with TMT126.

Prosser_1205.raw x

Select files labeled with TMT127.

Prosser_1205.raw x

Select files labeled with TMT128.

Prosser_1205.raw x

Ready?

SDRF file with label information

characteristics[individual]	assay name	technology type	comment[data file]	comment[[label]]	comment[technical replicate]	comment[fraction i]
0 205	run 10	proteomic profiling by mass spectrometry	Prosser_1205.raw	TMT126	1	
1 205	run 10	proteomic profiling by mass spectrometry	Prosser_1205.raw	TMT127	1	
2 205	run 10	proteomic profiling by mass spectrometry	Prosser_1205.raw	TMT128	1	

Made with Streamlit

Following this, the user is prompted to specify which raw files contain each selected label. Given that a single raw file can have multiple labels, this will lead to the creation of multiple rows in the final SDRF file. If the SDRF file already has data from the raw file (for instance, if it's an intermediate SDRF file), the existing information will be replicated.

Step 3: Input required columns

Based on the SDRF template selected on the Home screen, users are provided with a list of required columns. Filling out these columns is crucial to generate a valid SDRF file.

Many columns require inputs that adhere to precise SDRF guidelines and formatting. To simplify this for users and to minimize errors, lesSDRF has integrated these requirements directly within the tool. This eliminates the need for users to sift through the extensive SDRF documentation. Additionally, the corresponding ontologies are embedded within lesSDRF, offering an easy navigation experience.

The following columns are empty and required:

- start
- characteristics[organism part]
- characteristics[cell type]
- characteristics[ancestry category]
- characteristics[age]
- characteristics[sex]
- characteristics[disease]
- characteristics[individual]
- characteristics[biological replicate]
- technology type
- assay name
- comment[technical replicate]
- comment[fraction identifier]
- comment[instrument]
- comment[cleavage agent details]
- comment[modification parameters]
- comment[fragment mass tolerance]
- comment[precursor mass tolerance]

Press to download SDRF file

This is your current SDRF file.

	source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry category]	characteristics[age]	characteristics[sex]
0	Sample 1	Homo sapiens	None	None	None	None	None
1	Sample 2	Homo sapiens	None	None	None	None	None
2	Sample 3	Homo sapiens	None	None	None	None	None
3	Sample 4	Homo sapiens	None	None	None	None	None
4	Sample 5	Homo sapiens	None	None	None	None	None
5	Sample 6	Homo sapiens	None	None	None	None	None
6	Sample 7	Homo sapiens	None	None	None	None	None
7	Sample 8	Homo sapiens	None	None	None	None	None
8	Sample 9	Homo sapiens	None	None	None	None	None
9	Sample 10	Homo sapiens	None	None	None	None	None

In the sidebar, all the empty columns that are required for you SDRF file are listed. Select the one you want to annotate first.
When a column is filled in, it will disappear from the sidebar so you can keep track on which columns still require input.
When you want to reannotate a previously filled in column, you can do so by clicking on the **undo column** button in the sidebar.
Every page will start with your current SDRF file and at the bottom of the page there is a button to download the intermediate SDRF file

On the left side, a to-do list of columns is displayed. By selecting any of these column names, users can proceed to input the relevant information. If an error is made during entry, the "undo column" option allows the user to clear the contents of that column.

Additionally, users have the flexibility to download their intermediate SDRF file at any point by clicking the "Download" button.

3. Required columns

This is your current SDRF file.

source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry category]	characteristics[age]	characteristic
0 Sample 1	Homo sapiens	None	None	None	None	None
1 Sample 2	Homo sapiens	None	None	None	None	None
2 Sample 3	Homo sapiens	None	None	None	None	None
3 Sample 4	Homo sapiens	None	None	None	None	None
4 Sample 5	Homo sapiens	None	None	None	None	None
5 Sample 6	Homo sapiens	None	None	None	None	None
6 Sample 7	Homo sapiens	None	None	None	None	None
7 Sample 8	Homo sapiens	None	None	None	None	None
8 Sample 9	Homo sapiens	None	None	None	None	None
9 Sample 10	Homo sapiens	None	None	None	None	None

Select the part of the organism that is present in your sample

Are there multiple characteristics[organism part] in your data?

No
 Yes

Select your matching ontology term using this autocomplete function
Or follow the ontology based drop down menu below

heart l

heart left ventricle
heart right ventricle
endothelial cell of respiratory system lymphatic vessel
hepatopancreatic ampulla

If answers are currently incorrect click here to update column

Update

source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry cat...]	characteristics
Sample 1	Homo sapiens	empty	empty	empty	empty
Sample 2	Homo sapiens	empty	empty	empty	empty
Sample 3	Homo sapiens	empty	empty	empty	empty
Sample 4	Homo sapiens	empty	empty	empty	empty
Sample 5	Homo sapiens	empty	empty	empty	empty
Sample 6	Homo sapiens	empty	empty	empty	empty
Sample 7	Homo sapiens	empty	empty	empty	empty

Selecting a column, like "characteristics[organism part]," for instance, will first bring up a set of questions aimed at determining the structure of the SDRF file.

Subsequently, users can choose the appropriate ontology terms. This can be done through an autocomplete search function or by utilizing a dropdown menu. If a single term is picked, it's automatically placed into the column.

However, if multiple terms are chosen, a dropdown menu is generated within the SDRF file. Users can then pinpoint the correct ontology term for each corresponding row using this embedded menu.

3. Required columns

This is your current SDRF file.

	source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry category]	characteristics[age]	characteristics[biological replicate]
0	Sample 1	Homo sapiens	heart left ventricle	None	None	None	
1	Sample 2	Homo sapiens	heart left ventricle	None	None	None	
2	Sample 3	Homo sapiens	heart left ventricle	None	None	None	
3	Sample 4	Homo sapiens	heart left ventricle	None	None	None	
4	Sample 5	Homo sapiens	heart left ventricle	None	None	None	
5	Sample 6	Homo sapiens	heart left ventricle	None	None	None	
6	Sample 7	Homo sapiens	heart left ventricle	None	None	None	
7	Sample 8	Homo sapiens	heart left ventricle	None	None	None	
8	Sample 9	Homo sapiens	heart left ventricle	None	None	None	
9	Sample 10	Homo sapiens	heart left ventricle	None	None	None	

In the sidebar, all the empty columns that are required for you SDRF file are listed. Select the one you want to annotate first.
When a column is filled in, it will disappear from the sidebar so you can keep track on which columns still require input.
When you want to reannotate a previously filled in column, you can do so by clicking on the **undo column** button in the sidebar.

Every page will start with your current SDRF file and at the bottom of the page there is a button to download the intermediate SDRF file

Made with Streamlit

This case demonstrates how the column was automatically completed as there was only a single term selected.

Home
1. Mapping local metadata
2. Labeling
3. Required columns
4. Additional columns
5. Experiment types

The following columns are empty and required
 start
 characteristics[age]
 characteristics[disease]
 characteristics[individual]
 characteristics[biological replicate]
 technology type
 assay name
 comment[technical replicate]
 comment[fraction identifier]
 comment[instrument]
 comment[cleavage agent details]
 comment[modification parameters]
 comment[fragment mass tolerance]
 comment[precursor mass tolerance]
 undo column

Press to download SDRF file

	Sample 5	Homo sapiens	heart left ventricle	cardiocyte	Not available	None	F
5	Sample 6	Homo sapiens	heart left ventricle	cardiocyte	Not available	None	F
6	Sample 7	Homo sapiens	heart left ventricle	cardiocyte	Not available	None	F
7	Sample 8	Homo sapiens	heart left ventricle	cardiocyte	Not available	None	M
8	Sample 9	Homo sapiens	heart left ventricle	cardiocyte	Not available	None	M
9	Sample 10	Homo sapiens	heart left ventricle	cardiocyte	Not available	None	M

If you have healthy and control samples, indicate healthy samples using *normal*. Input the disease for the other samples using the ontology

Are there multiple characteristics[disease] in your data? How many different characteristics[disease] are in your data?
 No
 Yes 4 - +

Select your matching ontology term using this autocomplete function
normal x hypertrophic car... x dilated cardiomy... x ischemic cardio... x

Or follow the ontology based drop down menu below

> disease
 normal
 Not available
 Not applicable

Submit selection

If all cells are correctly filled in click twice on the update button

Update

source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry cat...]	characteristics
Sample 1	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 2	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 3	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 4	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 5	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 6	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 7	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 8	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 9	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 10	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 11	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 12	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 13	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 14	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 15	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 16	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 17	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty
Sample 18	Homo sapiens	heart left ventricle	cardiocyte	Not available	empty

This example showcases a scenario where multiple terms are relevant. The user's data contains four distinct disease types. After selecting these four diseases, they are then made available as options within the dropdown menu embedded in the 'characteristics[disease]' column.

The screenshot shows the lesSDRF application interface. On the left, a sidebar lists steps: Home, 1. Mapping local metadata, 2. Labeling, 3. Required columns (selected), 4. Additional columns, and 5. Experiment types. Below this, a list of empty required columns is shown, with 'characteristics[disease]' highlighted and enclosed in a red box. At the bottom of the sidebar is a button 'Press to download SDRF file'.

The main area has a title 'disease for the other samples using the ontology'. It asks if there are multiple characteristics[disease] in the data (Yes is selected) and how many (4). A dropdown menu shows 'Choose an option' and a list of diseases: disease, normal, Not available, and Not applicable. A 'Submit selection' button is present.

A note says 'If all cells are correctly filled in click twice on the update button'. Below is a table with 16 rows (Sample 1 to Sample 16) and 6 columns: source name, characteristics[sex], characteristics[disease], characteristics[individual], characteristics[biological replicate], and technology type. The 'characteristics[disease]' column is highlighted with a red box. In Sample 1, 'normal' is selected. In Sample 3, 'hypertrophic cardiomyopathy' and 'dilated cardiomyopathy' are listed. In Sample 4, 'normal' is selected. In Sample 16, 'normal' is selected. All other cells in the table are empty.

This example showcases a scenario where multiple terms are relevant. The user's data contains four distinct disease types. After selecting these four diseases, they are then made available as options within the dropdown menu embedded in the 'characteristics[disease]' column.

Users can input directly within the designated lesSDRF column. However, to prevent accidental edits, only the highlighted column is editable, ensuring other columns remain unchanged.

The hierarchical structure of the ontology tree facilitates easy navigation and selection of terms, especially when users are uncertain about the correct terms to use. This setup is derived from the Ontology Lookup Service, ensuring a user-friendly and intuitive approach to term selection.

The screenshot shows a software application window with a sidebar and a main content area.

Sidebar (Left):

- Home
- 1. Mapping local metadata
- 2. Labeling
- 3. Required columns** (highlighted)
- 4. Additional columns
- 5. Experiment types

The "Required columns" section lists several fields as empty and required:

- start
- comment[instrument]
- comment[cleavage agent details]
- comment[modification parameters]
- comment[precursor mass tolerance]
- comment[fragment mass tolerance]
- undo column

Buttons at the bottom of this section include "Press to download SDRF file" and "Submit selection".

Main Content Area (Right):

A large list of instrument models is shown in a hierarchical tree structure. The "Q Exactive" model is selected (indicated by a checked checkbox). Other models listed include:

- LTQ Velos
- LTQ Velos ETD
- TSQ Vantage
- LTQ Orbitrap Velos
- ISQ
- Velos Plus
- LTO Orbitrap Elite
- Q Exactive** (selected)
- Orbitrap Fusion
- Orbitrap Fusion ETD
- TSQ Quantiva
- TSQ Endura
- Q Exactive HF
- TSQ 8000 Evo
- Exactive Plus
- Q Exactive Plus
- Orbitrap Fusion Lumos
- LTO Orbitrap Classic
- TSQ Altis
- TSQ Quantis
- TSQ 9000
- Q Exactive HF-X
- Orbitrap Exploris 480
- Orbitrap Eclipse
- Orbitrap Exploris 240
- Orbitrap Exploris 120
- LTO Orbitrap Velos Pro
- Orbitrap ID-X
- Q Exactive UHMR
- TSQ Altis Plus
- Hitachi instrument model
- Varian instrument model
- Agilent instrument model
- Dionex instrument model
- Applied Biosystems instrument model
- LECO instrument model

If all cells are correctly filled in click twice on the update button

source name	characteristics[organism]	characteristics[organism part]	characteristics[age]	characteristics[development]	characteristics[...]
Sample 10	Homo sapiens	heart left ventricle	44Y	adult	M

A red "Update" button is visible at the bottom right of the table.

The hierarchical structure of the ontology tree facilitates easy navigation and selection of terms, especially when users are uncertain about the correct terms to use. This setup is derived from the Ontology Lookup Service, ensuring a user-friendly and intuitive approach to term selection.

The screenshot shows a software interface for labeling a dataset. On the left, a sidebar lists steps: Home, 1. Mapping local metadata, 2. Labeling, 3. Required columns (selected), 4. Additional columns, and 5. Experiment types. Below this is a list of required columns:

- start
- characteristics[cell type]
- characteristics[ancestry category]
- characteristics[age]
- characteristics[sex]
- characteristics[disease]
- characteristics[individual]
- characteristics[biological replicate]
- technology type
- assay name
- comment[technical replicate]
- comment[fraction identifier]
- comment[instrument]
- comment[cleavage agent details]
- comment[modification parameters]
- comment[fragment mass tolerance]
- comment[precursor mass tolerance]
- undo column

A button at the bottom says "Press to download SDRF file".

The main area displays a table of 13 samples, all labeled "Homo sapiens" and "heart left ventricle".

	source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry cat...]	characteristics
3	Sample 4	Homo sapiens	heart left ventricle	None	None	None
4	Sample 5	Homo sapiens	heart left ventricle	None	None	None
5	Sample 6	Homo sapiens	heart left ventricle	None	None	None
6	Sample 7	Homo sapiens	heart left ventricle	None	None	None
7	Sample 8	Homo sapiens	heart left ventricle	None	None	None
8	Sample 9	Homo sapiens	heart left ventricle	None	None	None
9	Sample 10	Homo sapiens	heart left ventricle	None	None	None
10	Sample 11	Homo sapiens	heart left ventricle	None	None	None
11	Sample 12	Homo sapiens	heart left ventricle	None	None	None
12	Sample 13	Homo sapiens	heart left ventricle	None	None	None

Input the cell type of your sample

Are there multiple characteristics[cell type] in your data?

No
 Yes

Select your matching ontology term using this autocomplete function

Choose an option

Or follow the ontology based drop down menu below

You need to select a total of 1.

Submit selection

Selection contains: ['primary cultured cell', 'spheroplast']

If all cells are correctly filled in click twice on the update button

Update

source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry cat...]	characteristics
Sample 1	Homo sapiens	heart left ventricle	empty	empty	empty
Sample 2	Homo sapiens	heart left ventricle	empty	empty	empty
Sample 3	Homo sapiens	heart left ventricle	empty	empty	empty
Sample 4	Homo sapiens	heart left ventricle	empty	empty	empty
Sample 5	Homo sapiens	heart left ventricle	empty	empty	empty

The structural questions also act as an additional safeguard to prevent users from making incorrect selections. For instance, if a user specifies that there's only one cell type present in the dataset, but then selects two terms from the ontology tree, lesSDRF will flag this discrepancy. The system will then prompt the user to make the necessary corrections.

3. Required columns

This is your current SDRF file.

	source name	characteristics[organism]	characteristics[organism part]	characteristics[cell type]	characteristics[ancestry category]	characteristics[age]	characteristic
0	Sample 1	Homo sapiens	heart left ventricle	cardiocyte		None	None
1	Sample 2	Homo sapiens	heart left ventricle	cardiocyte		None	None
2	Sample 3	Homo sapiens	heart left ventricle	cardiocyte		None	None
3	Sample 4	Homo sapiens	heart left ventricle	cardiocyte		None	None
4	Sample 5	Homo sapiens	heart left ventricle	cardiocyte		None	None
5	Sample 6	Homo sapiens	heart left ventricle	cardiocyte		None	None
6	Sample 7	Homo sapiens	heart left ventricle	cardiocyte		None	None
7	Sample 8	Homo sapiens	heart left ventricle	cardiocyte		None	None
8	Sample 9	Homo sapiens	heart left ventricle	cardiocyte		None	None
9	Sample 10	Homo sapiens	heart left ventricle	cardiocyte		None	None

Input the ancestry of your samples

Are there multiple characteristics[ancestry category] in your data?

No
 Yes

Select your matching ontology term using this autocomplete function
Or follow the ontology based drop down menu below

Choose an option

- > European
- > Asian
- > African
- > Native American
- > Hispanic or Latin American
- > Greater Middle Eastern (Middle Eastern, North African or Persian)
- > African American or Afro-Caribbean
- > Oceanian
- > uncategorised population
- > genetically isolated population
- Aboriginal Australian
- > undefined ancestry population

Not available

Submit selection

In some instances, the SDRF template includes columns that demand information which may not be accessible or relevant to a specific experimental scenario. To address this, terms like "Not available" have been incorporated into the ontology tree.

The screenshot shows a software interface for managing biological data. On the left, a sidebar lists navigation options: Home, 1. Mapping local metadata, 2. Labeling, 3. Required columns (which is selected and highlighted in grey), 4. Additional columns, and 5. Experiment types. Below this, a message states: "The following columns are empty and required: start, assay name, technology type, comment[technical replicate], comment[fraction identifier], comment[instrument], comment[cleavage agent details], comment[modification parameters], comment[precursor mass tolerance], comment[fragment mass tolerance], undo column". A button "Press to download SDRF file" is also present.

The main area displays a table titled "Update" with the following columns: s[disease], characteristics[biological re..., characteristics[individual], assay name, technology type, comment[data file]. The "assay name" column is highlighted with a red box. The table contains numerous rows, each with a unique ID (e.g., 30, 31, 32, 33) and corresponding values for the other columns. The "assay name" column values include "run 1", "run 10", "r", and several entries labeled "empty".

In lesSDRF, only a handful of columns permit manual input. The "assay name" column is one such example, where the value is unique and determined by the user's discretion. Users can input directly within the designated lesSDRF column. However, to prevent accidental edits, only the highlighted column is editable, ensuring other columns remain unchanged.

The screenshot shows a user interface for managing sample metadata. On the left, a sidebar lists steps: Home, 1. Mapping local metadata, 2. Labeling, 3. Required columns (selected), 4. Additional columns, and 5. Experiment types. Below this, a note says 'The following columns are empty and required' followed by a list of optional fields, with 'characteristics[age]' checked and highlighted with a red box. A button 'Press to download SDRF file' is also present.

The main area contains a table titled 'Update' with the following columns: source name, organism, characteristics[organism part], characteristics[cell type], characteristics[ancestry cat...], and characteristics[age]. The table lists 33 samples, all of which have 'heart left ventricle' in the first three columns and 'cardiocyte' in the second column. The 'characteristics[age]' column contains numerical values ranging from 46Y to 66Y. The last row, Sample 33, has '66' in the 'characteristics[age]' column, which is highlighted with a red box. A red box also highlights the error message 'The age column is not in the correct format, please check and try again' at the bottom of the table area.

For columns like 'characteristics[age]' that need manual input—albeit in a specific format—lesSDRF monitors format compatibility. In the given example, when a user inputs the age for all samples, an error occurs in the last row: it displays "66" instead of the correct "66Y". Such discrepancies in format are promptly identified and flagged by lesSDRF.

The screenshot shows the lesSDRF application interface. On the left, a sidebar lists steps: Home, 1. Mapping local metadata, 2. Labeling, 3. Required columns (selected), 4. Additional columns, and 5. Experiment types. Below this, a message says "The following columns are empty and required" with a link. A list of required columns includes "start" and "characteristics[sex]" (which is highlighted with a red border). A button "Press to download SDRF file" is at the bottom. The main area shows a table with 12 rows of sample data. Above the table, a question "Are there multiple sexes in your data?" has a dropdown menu set to "Yes". Below the table, an "Update" button is visible.

	source name	characteristics[sex]	characteristics[disease]	characteristics[individual]	characteristics[biological replicate]	technol...
5	Sample 6	Homo sapiens	heart left ventricle	cardiocyte	Not available	55Y
6	Sample 7	Homo sapiens	heart left ventricle	cardiocyte	Not available	64Y
7	Sample 8	Homo sapiens	heart left ventricle	cardiocyte	Not available	55Y
8	Sample 9	Homo sapiens	heart left ventricle	cardiocyte	Not available	48Y
9	Sample 10	Homo sapiens	heart left ventricle	cardiocyte	Not available	44Y
10	Sample 11	Homo sapiens	heart left ventricle	cardiocyte	Not available	65Y
11	Sample 12	Homo sapiens	heart left ventricle	cardiocyte	Not available	35Y
12	Sample 13	Homo sapiens	heart left ventricle	cardiocyte	Not available	40Y

For columns like 'characteristics[sex]', which offer a limited set of possible entries, lesSDRF provides a dropdown menu embedded directly within the SDRF column. This approach minimizes the risk of users making errors due to typos.

When addressing the biological replicates column, lesSDRF aims to streamline the process for the user. For instance, if there are no biological replicates, the entire column is automatically populated with the value "1". This approach eliminates the manual task for the user and adheres to SDRF guidelines. However, if there are replicates present, the user will be prompted to specify this directly within the column.

The screenshot shows the lesSDRF interface. On the left, a sidebar lists steps: Home, 1. Mapping local metadata, 2. Labeling, 3. Required columns (selected), 4. Additional columns, and 5. Experiment types. Below this, a list of empty required columns is shown, with 'characteristics[individual]' highlighted by a red box. A button 'Press to download SDRF file' is at the bottom. The main area displays a table of 35 samples. Above the table, a message asks 'Do you have multiple individuals in your data?' with a dropdown set to 'Yes' and a count of '35'. A checkbox 'Ready for input?' is checked. The table has columns: source name, assay cat..., characteristics[cell type], characteristics[disease], characteristics[biological re...], and characteristics[individual]. The 'characteristics[individual]' column for rows 1 through 23 is highlighted with a red box. A small red box also highlights the '35' in the count dropdown.

A notable benefit of lesSDRF is its ability to mimic functionalities commonly found in spreadsheet software like Excel, despite being a browser-based tool. For instance, when a user needs to input individual numbers, after specifying the total count, they can pick these numbers from a dropdown menu. But rather than manually selecting each entry, lesSDRF offers a drag-and-drop feature. This functionality, similar to what's found in spreadsheet software, allows users to automatically fill consecutive numbers across successive rows, streamlining the process.

The screenshot shows the lesSDRF application interface. On the left, a sidebar lists steps: Home, 1. Mapping local metadata, 2. Labeling, 3. Required columns (selected), 4. Additional columns, and 5. Experiment types. Below this, it says "The following columns are empty and required" and lists options like start, technology type (which is selected and highlighted with a red box), and various comment and undo options. At the bottom of the sidebar is a "Press to download SDRF file" button. The main area is titled "3. Required columns" and contains a table of SDRF data. A red box highlights the "technology type" section in the table header. Below the table is a modal dialog titled "Choose the technology type from these options:" with three radio buttons: "proteomic profiling by mass spectrometry" (selected and highlighted with a red box), "metabolomics profiling by mass spectrometry", and "other". A "Submit" button is at the bottom of the dialog. The footer of the page says "Made with Streamlit".

Columns such as "technology type" accept only specific inputs as dictated by SDRF guidelines. To streamline the process and alleviate the potential for user confusion or the need to navigate the comprehensive SDRF documentation, lesSDRF clearly displays and integrates these options. Users can simply make their selection with a click, ensuring both accuracy and convenience.

The screenshot shows the lesSDRF software interface. On the left, a sidebar lists steps: Home, 1. Mapping local metadata, 2. Labeling, 3. Required columns (selected), 4. Additional columns, and 5. Experiment types. Below this, a note says 'The following columns are empty and required' with a 'start' button. A dropdown menu is open, showing 'comment[cleavage agent details]' with a red border around it, and other options like 'comment(modification parameters)', 'comment(precursor mass tolerance)', 'comment(fragment mass tolerance)', and 'undo column'. At the bottom of the sidebar is a 'Press to download SDRF file' button.

3. Required columns

This is your current SDRF file.

	source name	characteristics[organism]	characteristics[organism part]	characteristics[age]	characteristics[developmental stage]	characteristics[sex]	characteristics[
0	Sample 10	Homo sapiens	heart left ventricle	44Y	adult	M	not available
1	Sample 11	Homo sapiens	heart left ventricle	65Y	adult	F	not available
2	Sample 12	Homo sapiens	heart left ventricle	35Y	adult	M	not available
3	Sample 14	Homo sapiens	heart left ventricle	53Y	adult	F	not available
4	Sample 18	Homo sapiens	heart left ventricle	26Y	adult	M	not available
5	Sample 24	Homo sapiens	heart left ventricle	42Y	adult	M	not available
6	Sample 26	Homo sapiens	heart left ventricle	40Y	adult	M	not available
7	Sample 28	Homo sapiens	heart left ventricle	51Y	adult	M	not available
8	Sample 31	Homo sapiens	heart left ventricle	58Y	adult	M	not available
9	Sample 33	Homo sapiens	heart left ventricle	46Y	adult	M	not available

Input the cleavage agent details of your sample

Select the cleavage agents used in your sample If no cleavage agent was used e.g. in top down proteomics, choose NoEnzyme

Choose an option

- NoEnzyme
- NT=2-iodobenzoate; CS='(?:<=W)'
- NT=Arg-C; CS='(?:<=R)(?:P)'
- NT=Asp-N; CS='(?:[BD])'
- NT=Asp-N ambic; CS='(?:[DE])'
- NT=CNBr; CS='(?:<=M)'
- NT=Chymotrypsin; CS='(?:<=[FYWL])(?:P)'
- NT=Formic acid; CS='(?:<=[DNL])(?:D)'

Similarly, when inputting the cleavage agent, lesSDRF simplifies the process. The specific naming conventions, formatting of the cleavage pattern, and other details are directly incorporated into the tool. Users can effortlessly type the name of their cleavage agent into the search bar and quickly find a matching option, ensuring accuracy and adherence to SDRF standards.

3. Required columns

This is your current SDRF file.

	source name	characteristics[organism]	characteristics[organism part]	characteristics[age]	characteristics[developmental stage]	characteristics[sex]	characteristics[
0	Sample 10	Homo sapiens	heart left ventricle	44Y	adult	M	not available
1	Sample 11	Homo sapiens	heart left ventricle	65Y	adult	F	not available
2	Sample 12	Homo sapiens	heart left ventricle	35Y	adult	M	not available
3	Sample 14	Homo sapiens	heart left ventricle	53Y	adult	F	not available
4	Sample 18	Homo sapiens	heart left ventricle	26Y	adult	M	not available
5	Sample 24	Homo sapiens	heart left ventricle	42Y	adult	M	not available
6	Sample 26	Homo sapiens	heart left ventricle	40Y	adult	M	not available
7	Sample 28	Homo sapiens	heart left ventricle	51Y	adult	M	not available
8	Sample 31	Homo sapiens	heart left ventricle	58Y	adult	M	not available
9	Sample 33	Homo sapiens	heart left ventricle	46Y	adult	M	not available

Input the cleavage agent details of your sample

Select the cleavage agents used in your sample If no cleavage agent was used e.g. in top down proteomics, choose *NoEnzyme*

tryps

- NT=Trypsin/P; CS="(?<=[KR])"
- NT=Trypsin; CS="(?<=[KR])(?!P)"
- NT=Chymotrypsin; CS="(?<=[FYWL])(?!P)"
- NT=TrypChymo; CS="(?<=[FYWLKR])(?!P)"

Similarly, when inputting the cleavage agent, lesSDRF simplifies the process. The specific naming conventions, formatting of the cleavage pattern, and other details are directly incorporated into the tool. Users can effortlessly type the name of their cleavage agent into the search bar and quickly find a matching option, ensuring accuracy and adherence to SDRF standards.

The screenshot shows the 'Required columns' step in the lesSDRF software. A red box highlights the 'comment[modification parameters]' option under 'Additional columns'. Below this, a search bar shows 'Oxidation' and 'Carbamidomethyl'. A dropdown menu for 'Oxidation' shows 'Fixed' selected, 'Anywhere' for position, and 'M' for target amino acid. A list of modifications is at the top, and a table of sample metadata is at the bottom.

The intricate formatting required for modifications is streamlined in lesSDRF to enhance user experience. Initially, users can broadly search for the modification name using an autocomplete search function. Once the appropriate modifications are selected, lesSDRF prompts users to provide specific details for each modification. This includes the type of modification (either fixed or variable), its position (be it anywhere, C-terminus, or N-terminus), and the target amino acid. These inputs are then structured into the SDRF format. After finalizing their entries and selecting the 'Submit modification' checkbox, lesSDRF generates a new column in the SDRF file for each specified modification.

Step 4: Input additional columns

Beyond the required columns needed for a valid SDRF, there are other optional columns available. To prevent overwhelming users with choices, these are incorporated into step 4 and are distinctly labeled as "additional columns". This design choice ensures clarity while providing users the flexibility to include any supplementary data they deem relevant.

Examples of such columns are compound, enrichment, collision energy, etc.

Of these columns, the 'factor value' stands out as particularly significant. This column represents the study variable, essentially highlighting the main focus or subject of investigation in the study.

The screenshot shows a software application window with a sidebar on the left containing navigation links: Home, 1. Mapping local metadata, 2. Labeling, 3. Required columns, 4. Additional columns (which is selected and highlighted in blue), and 5. Experiment types. Below these are some instructions and a list of options for adding columns, with 'factor value' being the selected option (indicated by a red box).

4. Additional columns

This is your current SDRF file.

	source name	characteristics[organism]	characteristics[organism part]	characteristics[age]	characteristics[developmental stage]	characteristics[sex]	characteristics[
0	Sample 10	Homo sapiens	heart left ventricle	44Y	adult	M	not available
1	Sample 11	Homo sapiens	heart left ventricle	65Y	adult	F	not available
2	Sample 12	Homo sapiens	heart left ventricle	35Y	adult	M	not available
3	Sample 14	Homo sapiens	heart left ventricle	53Y	adult	F	not available
4	Sample 18	Homo sapiens	heart left ventricle	26Y	adult	M	not available
5	Sample 24	Homo sapiens	heart left ventricle	42Y	adult	M	not available
6	Sample 26	Homo sapiens	heart left ventricle	40Y	adult	M	not available
7	Sample 28	Homo sapiens	heart left ventricle	51Y	adult	M	not available
8	Sample 31	Homo sapiens	heart left ventricle	58Y	adult	M	not available
9	Sample 33	Homo sapiens	heart left ventricle	46Y	adult	M	not available

You can choose one or multiple columns that define the **factor value** in your experiment.
This column indicates which experimental factor/variable is used as the hypothesis to perform the data analysis.

If there are multiple factor values, we ask the user to take care in assigning biological and technical replicates.

Select a factor value: Save factor value?

Choose an option:

source name	characteristics[organism]	characteristics[organism part]	characteristics[age]	characteristics[developmental stage]	characteristics[sex]	characteristics[
			adult		M	not available	
			adult		F	not available	
			adult		M	not available	
			adult		F	not available	
			adult		M	not available	
			adult		M	not available	
			adult		M	not available	
			adult		M	not available	
			adult		M	not available	
7	Sample 28	Homo sapiens	heart left ventricle	51Y	adult	M	not available
8	Sample 31	Homo sapiens	heart left ventricle	58Y	adult	M	not available
9	Sample 33	Homo sapiens	heart left ventricle	46Y	adult	M	not available

The 'factor value' pinpoints the study variable and is derived from a column that has been previously populated. Users can designate which column will serve as the 'factor value' by selecting from the provided dropdown menu.

4. Additional columns

This is your current SDRF file.

id[modification parameters]_1	comment[precursor mass tolerance]	comment[fragment mass tolerance]
1 iamidomethylAC=4;CF=H(3) C(2) N O;MM=57.021464;MT=Fixed;PP=Anywhere	20 ppm	20 ppm
2 iamidomethylAC=4;CF=H(3) C(2) N O;MM=57.021464;MT=Fixed;PP=Anywhere	20 ppm	20 ppm
3 iamidomethylAC=4;CF=H(3) C(2) N O;MM=57.021464;MT=Fixed;PP=Anywhere	20 ppm	20 ppm
4 iamidomethylAC=4;CF=H(3) C(2) N O;MM=57.021464;MT=Fixed;PP=Anywhere	20 ppm	20 ppm
5 iamidomethylAC=4;CF=H(3) C(2) N O;MM=57.021464;MT=Fixed;PP=Anywhere	20 ppm	20 ppm
6 iamidomethylAC=4;CF=H(3) C(2) N O;MM=57.021464;MT=Fixed;PP=Anywhere	20 ppm	20 ppm
7 iamidomethylAC=4;CF=H(3) C(2) N O;MM=57.021464;MT=Fixed;PP=Anywhere	20 ppm	20 ppm
8 iamidomethylAC=4;CF=H(3) C(2) N O;MM=57.021464;MT=Fixed;PP=Anywhere	20 ppm	20 ppm
9 iamidomethylAC=4;CF=H(3) C(2) N O;MM=57.021464;MT=Fixed;PP=Anywhere	20 ppm	20 ppm
10 iamidomethylAC=4;CF=H(3) C(2) N O;MM=57.021464;MT=Fixed;PP=Anywhere	20 ppm	20 ppm

You can choose one or multiple columns that define the **factor value** in your experiment.
This column indicates which experimental factor/variable is used as the hypothesis to perform the data analysis.

If there are multiple factor values, we ask the user to take care in assigning biological and technical replicates.

Select a factor value: Save factor value?

Choose an option

	source name	characteristics[organism]	characteristics[organism part]	characteristics[age]	characteristics[developmental stage]	characteristics[sex]	characteristics[
0	Sample 10	Homo sapiens	heart left ventricle	44Y	adult	M	not available
1	Sample 11	Homo sapiens	heart left ventricle	65Y	adult	F	not available
2	Sample 12	Homo sapiens	heart left ventricle	35Y	adult	M	not available
3	Sample 14	Homo sapiens	heart left ventricle	53Y	adult	F	not available
4	Sample 18	Homo sapiens	heart left ventricle	26Y	adult	M	not available
5	Sample 24	Homo sapiens	heart left ventricle	42Y	adult	M	not available
6	Sample 26	Homo sapiens	heart left ventricle	40Y	adult	M	not available
7	Sample 28	Homo sapiens	heart left ventricle	51Y	adult	M	not available
8	Sample 31	Homo sapiens	heart left ventricle	58Y	adult	M	not available
9	Sample 33	Homo sapiens	heart left ventricle	46Y	adult	M	not available

Selecting a column for 'factor value' leads to its duplication under that designation. For instance, if the "characteristics[disease]" column is chosen, it results in the creation of a new column named "factor value[disease]" appended to the end of the SDRF file, as illustrated above.

Home
1. Mapping local metadata
2. Labeling
3. Required columns
4. Additional columns
5. Experiment types

These are all possible columns you may want to add:

- start
- factor value
- characteristics[cell line]
- characteristics[compound]
- characteristics[concentration of compound]
- characteristics[enrichment process]
- comment[collision energy]
- comment[depletion]
- comment[dissociation method]
- undo column

[Press to download SDRF file](#)

	source name	characteristics[organism]	characteristics[organism part]	characteristics[age]	characteristics[developmental stage]	characteristics[sex]	characteristics[<i>z</i>]
0	Sample 10	Homo sapiens	heart left ventricle	44Y	adult	M	not available
1	Sample 11	Homo sapiens	heart left ventricle	65Y	adult	F	not available
2	Sample 12	Homo sapiens	heart left ventricle	35Y	adult	M	not available
3	Sample 14	Homo sapiens	heart left ventricle	53Y	adult	F	not available
4	Sample 18	Homo sapiens	heart left ventricle	26Y	adult	M	not available
5	Sample 24	Homo sapiens	heart left ventricle	42Y	adult	M	not available
6	Sample 26	Homo sapiens	heart left ventricle	40Y	adult	M	not available
7	Sample 28	Homo sapiens	heart left ventricle	51Y	adult	M	not available
8	Sample 31	Homo sapiens	heart left ventricle	58Y	adult	M	not available
9	Sample 33	Homo sapiens	heart left ventricle	46Y	adult	M	not available

There are still columns that you can add to your SDRF file.
Similar to the previous section, you can select a column to annotate in the sidebar. When a column is filled, it will disappear from the sidebar so you can keep track on which columns you can still add.
When you want to reannotate a previously filled in column, you can do so by clicking on the **undo column** button in the sidebar.

At the bottom of the page there are now two download buttons. The first one will download your intermediate SDRF file.
The second button will download your final SDRF file after some final checks.

Made with Streamlit

Once the user has finalized their SDRF file, they can simply click the "Download" button. This allows them to download their file in TSV format, ready for submission to a public proteomics repository.

Step 5: Experiment types

The final segment of lesSDRF is the "Experiment types" tab. Currently, this section is in development and heavily relies on feedback from the community. The ultimate aim is to enable users to specify the kind of experiment they've conducted. Subsequently, they can then fill out columns suggested by the community. This feature is particularly beneficial for studies that are relatively new to the field of proteomics, like metaproteomics or single cell proteomics. The community-driven approach ensures that the tool remains up-to-date and relevant to emerging research trends.

This screenshot shows the 'Experiment types' section of a SDRF application. On the left, a sidebar lists steps: Home, 1. Mapping local metadata, 2. Labeling, 3. Required columns, 4. Additional columns, and 5. Experiment types (which is selected). A red box highlights the 'Join the community effort' button. The main area displays a table of sample metadata:

	source name	characteristics[organism]	characteristics[organism part]	characteristics[age]	characteristics[developmental stage]	characteristics[sex]	characteristics[
0	Sample 10	Homo sapiens	heart left ventricle	44Y	adult	M	not available
1	Sample 11	Homo sapiens	heart left ventricle	65Y	adult	F	not available
2	Sample 12	Homo sapiens	heart left ventricle	35Y	adult	M	not available
3	Sample 14	Homo sapiens	heart left ventricle	53Y	adult	F	not available
4	Sample 18	Homo sapiens	heart left ventricle	26Y	adult	M	not available
5	Sample 24	Homo sapiens	heart left ventricle	42Y	adult	M	not available
6	Sample 26	Homo sapiens	heart left ventricle	40Y	adult	M	not available
7	Sample 28	Homo sapiens	heart left ventricle	51Y	adult	M	not available
8	Sample 31	Homo sapiens	heart left ventricle	58Y	adult	M	not available
9	Sample 33	Homo sapiens	heart left ventricle	46Y	adult	M	not available

A yellow banner at the bottom says 'Under development'. Below it is a section titled '5. Experiment types' with a note about the application being in development. It includes a list of suggested experiment types: Metaproteomics, Single cell proteomics, and Immunopeptidomics. A red box highlights the 'Metaproteomics' button.

Some experiment types have an atypical SDRF structure. Here you can find the community-suggested SDRF columns for such experiments.

The suggested columns are:

- Add characteristics[environmental material] to SDRF file
- Add characteristics[diet] to SDRF file
- Add characteristics[biome] to SDRF file
- Add characteristics[environmental condition] to SDRF file

Choosing an experiment type will display the columns suggested by the community. Users can then decide to integrate these columns into their metadata file if they're not already present.

On the left side, there's a dedicated button inviting users to be part of this community-driven initiative. Clicking on this button will redirect them to a GitHub issues page, facilitating discussions and deliberations regarding the columns and their relevance.