

# Exercise 1: Handling Raw Data

**Time:** 30 minutes

**Goal:** Transform unstructured CSV/TSV files into a valid BIDS/PRISM dataset

The screenshot shows the MRI-LAB GRAZ PRISM Studio software interface in Light Mode. At the top, there is a navigation bar with icons for Home, Projects, Validator, Converter, Tools, Specs, Docs, and Quit. The main title is "Prism Converter" with a subtitle "Convert Survey, Biometrics, Physio, and Eyetracking inputs to PRISM/BIDS-style outputs." Below this, a "Current Project" section shows "Port\_Vario" with a library path of "/Volumes/Evo/data/prism\_hub/Port\_Vario". A "Survey" tab is selected, showing survey questionnaires. Other tabs include Biometrics (Physical measures), Physio (ECG, Respiration), Eyetracking (Gaze data), and Participants (Demographics). The main workspace is titled "Survey Data Conversion" and "Conversion Mode". It offers two options: "Data Conversion" (selected) which converts responses to BIDS format, and "Template Generation" which extracts structure only. A "Survey File (.xlsx, .csv, .tsv, .lss, .iss)" input field is present, along with a "Choose File" button and a note about LimeSurvey (.lss/.iss) or Data Dictionary (.xlsx/.csv/.tsv). Below this, "Select specific survey" and "Participant ID Column" fields are shown, with "e.g. phq9" and "Auto-detect (PRISM surveys only)". An "ID Mapping File (optional)" field is also present. On the right, a "Session ID \*" field is set to "Select session..." with a note about creating sessions like "ses-01", "ses-02", etc. A "Language" dropdown is set to "Auto (template default)". At the bottom, there are "Preview (Dry-Run)" and "Convert" buttons, and a note about participant data extraction using the Participants tab. The footer includes the MRI-Lab Graz logo, contact information (karl.koschutnig@uni-graz.at, GitHub: MRI-Lab-Graz, Report an Issue), and credits (Maintained by Karl Koschutnig, Built with ❤️ for the research community).

Figure 1: Exercise 1 UI (Light Mode)

## What You'll Learn

By the end of this exercise, you will:

- Understand BIDS folder hierarchy (Dataset → Subject → Session → Modality)
- Know BIDS file naming conventions
- Use the GUI converter to create structured datasets
- Recognize the importance of sidecar JSON files

---

## Starting Materials

Look in the `raw_data/` folder:

- **wellbeing.tsv** - A survey about general well-being and life satisfaction.
- **fitness\_data.tsv** - Biometric measurements (heart rate, strength, etc.) from a physical fitness assessment.

These are typical “raw” data files - tab-delimited exports from your data collection tools.

---

## Your Task

Convert both the Wellbeing and Fitness data into a proper BIDS/PRISM dataset with the correct folder structure and file naming.

---

## Step-by-Step Instructions

### Step 1: Launch PRISM Studio

1. Open your web browser
2. Go to: <http://localhost:5001>
3. You should see the PRISM Studio home page

### Step 2: Open the Converter Tool

1. Click on “Converter” in the navigation menu (top or sidebar)
2. Select “Survey Data Converter”

### Step 3: Load Your Data (Wellbeing Survey)

1. Click “Browse” or “Choose File”
2. Navigate to: `demo/workshop/exercise_1_raw_data/raw_data/wellbeing.tsv`
3. Click “Upload” or “Load File”
4. Preview your data - you should see columns like `participant_id`, `session`, `age`, `WB01`, etc.

#### **Step 4: Map Columns**

The converter needs to know which column represents what:

**Participant ID:** - In the dropdown, select: “**This column represents → participant\_id**”

**Session:** - Select: “**This column represents → session**”

**Survey Name:** - Enter: **wellbeing** - This will appear in your filenames as **task-wellbeing**

**Modality:** - Select: **survey**

**Data Columns:** - The columns WB01 through WB05 are your survey items. The demographic columns (**age**, **sex**, etc.) will be automatically handled.

#### **Step 5: Configure Output**

##### **1. Output Directory:**

- Click “**Set Output Folder**”
- Navigate to: **demo/workshop/exercise\_1\_raw\_data/**
- Create a new folder called: **my\_dataset**
- Select this folder

##### **2. Preview Filename:**

- Check the preview: **sub-{id}\_ses-{session}\_task-wellbeing\_survey.tsv**
- This should look correct!

##### **3. Options to Enable:**

- Generate sidecars (JSON files)
- Create **participants.tsv**
- Create **dataset\_description.json**

#### **Step 6: Convert!**

1. Click “**Convert to BIDS**”
2. Wait for the progress bar
3. Success message should appear.

#### **Step 7: Convert Biometrics (Bonus)**

Repeat the process for **fitness\_data.tsv**: 1. Load **fitness\_data.tsv** 2. Map **participant\_id** and **session** 3. Enter Survey/Task Name: **fitness** 4. Change Modality to: **biometrics** 5. Select the same **my\_dataset** output folder 6. Click “**Convert to BIDS**”

---

#### **Step 8: Explore Your Dataset**

Navigate to **my\_dataset/** and explore the structure:

```
my_dataset/
    dataset_description.json
    participants.tsv
    sub-DEM0001/
        ses-baseline/
            survey/
                sub-DEM0001_ses-baseline_task-wellbeing_survey.tsv
                sub-DEM0001_ses-baseline_task-wellbeing_survey.json
            biometrics/
                sub-DEM0001_ses-baseline_task-fitness_biometrics.tsv
                sub-DEM0001_ses-baseline_task-fitness_biometrics.json
```

Open some files and look inside!

---

### Checkpoint: Did It Work?

**You should have:** - [ ] A `my_dataset/` folder with proper structure - [ ] `dataset_description.json` at the root - [ ] `participants.tsv` at the root - [ ] Folders named `sub-DEM0001/`, `sub-DEM0002/`, etc. - [ ] Inside each: `ses-baseline/survey/` (and `biometrics/` if you did the bonus) - [ ] `.tsv` data files with proper BIDS naming - [ ] `.json` sidecar files (one for each `.tsv`)

**File naming should follow this pattern:** - `sub-DEM0001` (with hyphen, not `subDEM0001`) - `ses-baseline` (with hyphen, not `sesbaseline`) - `task-wellbeing` (with hyphen, not `taskwellbeing`) - Underscores `_` separate the entities - Example: `sub-DEM0001_ses-baseline_task-wellbeing_survey.tsv`

---

### Quick Validation Test

Let's check if your dataset is valid:

1. Go to “Home” or “Validator” in PRISM Studio
2. Click “Select Dataset”
3. Choose your `my_dataset/` folder
4. Click “Validate Dataset”

**Expected Result:** - Warnings about missing metadata (this is OK! We'll fix this in Exercise 2) - No critical errors about file structure or naming - All files detected correctly

**If you see errors about file naming or structure:** - Double-check the filename pattern - Make sure there are hyphens after `sub-`, `ses-`, `task-` - Ask your instructor for help!

---

## What Just Happened?

You converted unstructured data into a standardized format!

**Before:** Just a CSV file sitting somewhere on your computer **After:** A properly structured dataset that:

- Follows international standards (BIDS)
- Can be understood by automated tools
- Has a clear hierarchy (subject → session → modality)
- Includes metadata files (JSON sidecars)
- Is ready for sharing and archiving

---

## Key Concepts

### BIDS Hierarchy

Dataset (study level)

    Subject (participant level) - sub-DEM0001, sub-DEM0002, ...  
    Session (visit level) - ses-baseline, ses-followup, ...  
    Modality (data type) - survey, biometrics, ...  
    Files (actual data)

### File Naming Rules

- **Entities** are key-value pairs: sub-DEM0001, ses-baseline, task-wellbeing
- **Separator** between entities: underscore \_
- **Separator** within entities: hyphen -
- **Suffix** describes the modality: survey, biometrics
- **Extension** is the file type: .tsv, .json

### Sidecar Files

- Every data file (.tsv, .json, etc.) should have a .json sidecar
  - The sidecar contains metadata about the data file
  - Same filename, just different extension
  - Example:
    - Data: sub-DEM0001\_ses-baseline\_task-wellbeing\_survey.tsv
    - Sidecar: sub-DEM0001\_ses-baseline\_task-wellbeing\_survey.json
- 

## Troubleshooting

**Problem:** “Invalid column mapping”

**Solution:** Make sure you selected a column for participant\_id

**Problem:** “Invalid characters in filename”

**Solution:** Check that task name doesn’t have spaces or special characters

**Problem: “Output folder not found”**

**Solution:** Make sure you created the `my_dataset/` folder first

**Problem: “No data rows found”**

**Solution:** Check that your CSV has data (not just headers)

**Problem: “No survey item columns matched the selected templates”**

**Solution:** - Make sure you loaded `wellbeing.tsv` from this exercise (`exercise_1_raw_data/raw_data/wellbeing.tsv`) - Ensure the file is tab-delimited TSV (not semicolon-separated) - Select the matching wellbeing template (WB01-WB05) or clear template filters and retry

---

## Next Steps

**Congratulations!** Your data is now structured.

But wait - the JSON sidecars are mostly empty! They only have basic information.

In **Exercise 2**, you'll learn how to fill in the metadata to make your dataset truly self-documenting.

---

## Bonus Challenge (If You Have Extra Time)

1. Try with participants data:
    - Load `participants_raw.tsv`
    - See if you can update the main `participants.tsv` file
  2. Add a second survey:
    - If there's a `gad7_anxiety.csv` file, convert it too
    - It should go into the same dataset structure
    - Files will be named: `sub-01_ses-01_task-gad7_survey.tsv`
  3. Explore the converter settings:
    - Can you change the file suffix from `survey` to `beh`?
    - What happens if you choose a different modality?
- 

**Ready for Exercise 2?** → Go to `../exercise_2_hunting_errors/`