Session 3

Documenting Data – Document, Microdata & DDI

LNADA Capacity Building Training

23 February 2026 | 13:30–14:45

Session Agenda (75 minutes)

NEW structure: Document FIRST, then Microdata

Time	Duration	Activity
13:30–13:45	15 min	**Exercise 3a** — Document a publication
13:45–13:53	8 min	**Publish document** to NADA
13:53–14:03	10 min	**DDI Codebook 2.5 structure** presentation
14:03–14:08	5 min	**Introduce sample microdata dataset**
14:08–14:40	32 min	**Exercise 3b** — Document microdata (6 steps)
14:40–14:45	5 min	**Save, review, and preview S4**

Why this order? Start simple (Document) to build confidence before tackling the complex DDI workflow.

Exercise 3a: Document a Publication

(~15 minutes)

Goal: Document a research publication using the Document metadata schema.

This builds confidence before the complex DDI exercise.

Schema: The Document schema is a mix of **Dublin Core, MARC21, and BibTex** elements.
See: Schema Guide — Document

Reference: Follow the official quick start guide:
World Bank Document Quick Start Guide

Data: Files provided in `quick_start_files/document/`

Steps: Open the quick start guide and follow along step-by-step.

Publishing Your Document to NADA

From Metadata Editor → LNADA Catalog

Now that your document is documented, let's publish it!

1. In the Metadata Editor, open your Document project

2. Click **Publish to NADA** (or **Publish** button)

3. Enter your **NADA catalog URL:** `https://nada.lsb.lao-stat.de`

4. Enter your **API key** (created in Exercise 1c)

5. Select **data access type:** Open Access (for this document)

6. Select publishing status: **Published**

7. Click **Confirm**

Data access type controls who can download your data — we'll cover all 7 access types in detail in S4.

Verify Your Published Document

Check it on NADA

1. Open browser → nada.lsb.lao-stat.de

2. Search for your document by title

3. Verify: title, author, abstract, thumbnail all appear correctly

Congratulations! Your first published entry on LNADA.

Next: We'll document microdata (more complex), then publish that too.

What is Microdata?

Individual-level records from surveys, censuses, or administrative data

- **Household surveys** — one record per household, variables for each household attribute

- **Individual surveys** — one record per person (nested within households, communities, or regions)

- **Census data** — population enumeration with personal/household characteristics

- **Administrative records** — health registries, education enrollments, tax records, etc.

- **Hierarchical structure** — household + individual records linked by identifier

Key feature: Lowest level of data detail, protecting privacy through de-identification

DDI Codebook 2.5: Overview

Five core sections organize all metadata

1. **Document Description** (doc_desc)

Metadata about the metadata itself — who documented it, when

2. **Study Description** (study_desc)

Everything about the study — title, producers, geography, methodology

Introducing the Sample Dataset

Generic synthetic survey from World Bank quick start files

Two Stata data files:

- `WLD_2023_SYNTH-SVY-HLD-EN_v01_M.dta` — **Household** level (primary units)
- `WLD_2023_SYNTH-SVY-IND-EN_v01_M.dta` — **Individual** level (nested within households)

Supporting documents:

- `synthetic_survey_questionnaire.xlsx` — Survey instrument with questions
- `synthetic_survey_info.xlsx` — Study information summary
- `synthetic_data_logo.jpg` — Thumbnail image

Why synthetic? Safe for training — no real respondent privacy concerns

Exercise 3b: Microdata Quick Start Reference

Follow the official guide alongside this exercise

Reference: World Bank Microdata Quick Start Guide

Open this guide now — it walks through the same steps we'll cover below.

Schema: DDI Codebook 2.5 — the international standard for microdata documentation.

Exercise 3b: Review metadata

Review your metadata:

1. Click **Export** or **Preview**

2. Choose **JSON preview** to see the structured metadata output

3. Examine the structure:

   - All sections present? (doc*desc*, *study*desc, files, variables)

   - Variables populated with labels and statistics?

   - External resources linked?

What's next (Session S4):
Publishing to LNADA via API