

Session 5: Day 1 Recap & LSB Microdata

Day 2 Morning: 09:00–10:20 (80 min)

Agenda for Today

- Day 1 Recap (15 min)
- Showcase: How Others Use NADA (15 min)
- Introducing LSB-Specific Datasets (10 min)
- Exercise 5: Document an LSB Microdata Dataset (30 min)
- Additional Data Types – If Time Allows (10 min)

Part 1: Day 1 Recap

15 minutes | Group discussion | Reinforce key concepts

What We Learned Yesterday

The NADA ecosystem:

- Metadata Editor = authoring tool (document metadata)
- LNADA = publication platform (discover & access data)
- PyDataTools = automation helper (extract variable metadata)

The workflow:

- Prepare data → Document in Editor → Publish to LNADA

Key Concepts – Data Preparation

Remember from Session 2:

- The Metadata Editor is NOT a data editing tool
- Clean data BEFORE documenting: quality checks, file organisation, privacy
- Remove direct identifiers (names, addresses, phone numbers)
- Variable labels and value labels are essential

Key Concepts – Metadata Standards

Three standards we practised:

Data Type	Standard	Exercise
Document	Dublin Core / MARC21 / BibTex	Exercise 3a
Microdata	DDI Codebook 2.5	Exercise 3b
Indicator	WB Indicator schema	Exercise 4a

DDI has 5 sections: Doc Description, Study Description, File Description, Variables, Variable Groups

Key Concepts – Publishing

From Session 4:

- Publishing sends metadata from Editor → LNADA via API
- Access levels control who can download data
- Published entries are searchable and discoverable

Questions from yesterday?

- Anything unclear about the Editor workflow?
- Any steps you'd like to revisit today?

Part 2: Showcase – How Others Use NADA

15 minutes | Real-world examples | Discussion prompts

World Bank Microdata Library

microdata.worldbank.org

- Thousands of datasets from 100+ countries
- Surveys, censuses, administrative data
- DDI Codebook standard
- Gold standard for NADA implementation

Philippines PSADA

psada.psa.gov.ph

- National Statistical Office catalog
- Similar context to Laos: single national institution publishing data
- Well-organized by topic
- Good example of domestic NADA deployment

India Microdata Library

microdata.gov.in/NADA

- Large-scale national catalog
- Many surveys, ministries, multiple languages
- Demonstrates scalability
- Shows how multiple data producers can feed one NADA instance

NADA Demo Catalog

nada-demo.ihsn.org

- Sandbox environment (IHSN Secretariat)
- Test drive NADA features
- Public data for exploration
- Low-risk place to learn

Discussion: What Can LSB Learn?

- **Organization:** How are datasets organized by topic, year, geography?
- **Metadata completeness:** What makes some entries more discoverable?
- **Data access:** How do they handle public vs. restricted access?
- **User experience:** What features help researchers find data?

Part 3: Introducing LSB-Specific Datasets

10 minutes | Real data context | Bridge from training to work

From Generic Examples to Real LSB Data

So far, we've documented:

- Generic household survey (training example in S3/S4)
- World Bank poverty indicators (training example)

Now we move to:

- Real LSB surveys in the rebuilt LNADA
- These surveys have real-world complexity
- Incomplete metadata from previous systems
- Lao-specific geography and identifiers
- Different languages (Lao and English)

LSB-Specific Dataset Characteristics

What's different from training examples?

Feature	Training Data	Real LSB Data
Metadata	Clean, complete	Gaps to fill
Geography	Generic regions	Lao provinces/districts
Language	English only	Lao + English mixed
History	Fresh start	Legacy from old LNADA
IDs	Generic	LSB conventions (LAO-LSB-...)

Available LSB Datasets (NESSTAR Files)

Two legacy datasets from previous LNADA (DDI 1.2.2 format):

Dataset	Description	Coverage
EC 2006.xml	Establishment Census 2006	All provinces
LSIS2011.xml	Lao Social Indicator Survey 2011	Selected provinces

Both will be converted and cataloged in the rebuilt LNADA.

Exercise 5: Document an LSB Microdata Dataset

[LSB dataset to be confirmed by trainer]

Context:

- Real survey, real metadata gaps
- Real geography (provinces, districts)
- Some fields in Lao, some in English
- LSB ID convention: LAO-LSB-[SURVEY]-[YEAR] or similar

Part 4: Exercise 5 – Document an LSB Microdata Dataset

30 minutes | Hands-on | Bridge to real work

Exercise 5: Dataset to be Documented

[LSB dataset to be confirmed by trainer]

What you'll do:

1. Open the Metadata Editor
2. Create a NEW project for the LSB survey
3. Fill in all 5 DDI sections (just like Exercises 3b & 4b on Day 1)
4. Handle real-world metadata gaps and Lao-specific fields
5. Document geographic coverage (provinces/districts)
6. Apply LSB naming conventions

Exercise 5: Expected Challenges (1/2)

Challenge 1: Incomplete metadata from legacy system

- **Solution:** Consult with data manager or survey documentation.
Note what's missing.

Challenge 2: Fields in Lao language

- **Solution:** Translate key fields where needed. Document language mix in metadata.

Exercise 5: Expected Challenges (2/2)

Challenge 3: Geographic coverage unclear

- **Solution:** Cross-check with survey report or sampling docs. List provinces/districts explicitly.

Challenge 4: Old metadata format (legacy NESSTAR/DDI 1.2.2)

- **Solution:** Focus on translating concepts to DDI 2.5. Structure may differ; content remains.

Exercise 5: Steps (30 min guided)

1. **Login & setup** (2 min) – Open Metadata Editor
2. **Doc Description** (3 min) – Title, version, date, documentation contact
3. **Study Description** (5 min) – Survey design, universe, sampling, data collection period
4. **File Description** (3 min) – Data files, formats, case quantity
5. **Variable-level metadata** (12 min) – Variable names, labels, type, categories, notes
6. **Review & discuss** (5 min) – What worked? What was harder than training?

Exercise 5: Key Takeaways

After completing Exercise 5, you will:

- ✓ Have documented a REAL LSB survey (not training data)
- ✓ Understand real-world metadata gaps and how to handle them
- ✓ Apply LSB naming conventions and geography
- ✓ Be ready to catalog LSB surveys for the rebuilt LNADA

This exercise = your first real cataloging task!

Discussion: Real vs. Training Data

Reflect as a group:

- What was harder about the LSB dataset?
- What metadata was missing or unclear?
- How did you solve those problems?
- What would help you catalog more LSB surveys?

Part 5: Additional Data Types – If Time Allows

10 minutes (optional) | If schedule allows | Quick overview

Other Data Types Available in NADA

Beyond microdata (covered in S3/S4 and Exercise 5):

1. **Indicators** — Time series (poverty rates, GDP, inflation)
2. **Documents** — Reports, papers, working documents
3. **Geographic** — Shapefiles, spatial data
4. **Images** — Photos, infographics, data visualizations
5. **Videos** — Training, data literacy, public awareness
6. **Scripts** — R, Python, data cleaning code
7. **Statistical tables** — Aggregate census/survey results

Quick Reference: Data Type & Standard

Data Type	Metadata Standard	Example	LSB Use Case
Microdata	DDI Codebook 2.5	Household survey	✓ Main focus
Indicator	Indicator schema	Poverty rate over time	Future
Document	Dublin Core / MARC21 / BibTex	Annual report, protocol	Future
Geographic	ISO 19139	Admin boundaries	Future
Image	Simple metadata	Survey photos, infographics	Optional
Video	Simple metadata	Training video	Optional
Script	Simple metadata	Data processing code	Optional
Table	Custom schema	Census aggregate tables	Optional

Self-Study: Quick Starts for Other Data Types

If you want to explore on your own:

- **Indicator quick start:** `quick_start_files/indicator/` – World Bank poverty data
- **Document quick start:** `quick_start_files/document/` – Book cover example
- **Geographic quick start:** `quick_start_files/geographic/` – Refugee camp boundaries
- **Image quick start:** `quick_start_files/image/` – Food market photo
- **Video quick start:** `quick_start_files/video/` – Food security training video
- **Script quick start:** `quick_start_files/script/` – (Coming soon)

Summary: Day 1 Recap & LSB Microdata

Learning outcomes:

- ✓ Recalled Day 1 concepts (recap discussion)
- ✓ Saw real-world NADA catalogs in action
- ✓ Understood how LSB's real data differs from training examples
- ✓ Documented a REAL LSB microdata survey (Exercise 5)
- ✓ (Optional) Explored other data types for future use

Next: Session S6 — Cataloging Workflows & LSB Metadata Strategy