

Session S6: Cataloging Workflows & LSB Metadata Strategy

LNADA Training – Day 2, 10:40–12:00

Session Overview

- End-to-end cataloging workflow (7 steps)
- World Bank dataset list strategy
- NESSTAR to Metadata Editor migration
- LSB metadata conventions
- Maintenance best practices
- Advanced: API & AI metadata

Duration: 80 minutes | **Presenter:** Frank Lieber

End-to-End Cataloging Workflow (Part 1)

The 7-Step Lifecycle

1. **Data Production** – survey/census conducted, data collected
2. **Data Processing** – cleaning, editing, quality assurance
3. **Data Preparation** – file organisation, privacy checks, labelling
4. **Metadata Creation** – use Metadata Editor, follow standards
5. **Quality Review** – peer review of metadata completeness

End-to-End Cataloging Workflow (Part 2)

The 7-Step Lifecycle (continued)

- 1. **Publishing** – push to LNADA, configure access levels (Open, PUF, Licensed, Restricted)
- 2. **Maintenance** – update metadata as corrections/new versions arrive

Roles & Responsibilities

Role	Responsibility
Data Producer	Collect data, conduct survey/census
Metadata Editor	Document data in Metadata Editor
Data Custodian	Manage data files, approve publishing
System Administrator	Maintain LNADA platform

Metadata Management at LSB (Part 1)

ID Conventions

Example standard format: LAO-LSB-LECS-2023-V01

LAO = Country code
LSB = Agency
LECS = Survey acronym
2023 = Year/round
V01 = Version

Benefits

- **Machine-readable** – easy to parse
- **Consistent** – standard format
- **Traceable** – version control clear
- **Unique** – no duplicates

Metadata Management at LSB (ID Convention Details)

ID Format Explained

Full format: LAO-LSB-LECS-2023-V01

- **LAO** = Country code (ISO 3166-1 alpha-3)
- **LSB** = Agency (Lao Statistics Bureau)
- **LECS** = Survey acronym
- **2023** = Year or survey round
- **V01** = Version number

Metadata Management at LSB (Part 2)

Controlled Vocabularies

Why standardize terms? Users search for "poverty," not "pauvreté" or "ความยากจน."

Example: Vocabulary Sources

- **CESSDA** (Common European Research Infrastructure Consortium) – topics, keywords

Metadata Management at LSB (Part 3)

Custom Templates

What's a template? A pre-filled Metadata Editor form with LSB defaults.

Example fields pre-filled:

- Country: Lao PDR
- Producer: Lao Statistics Bureau
- Language: Lao, English
- Theme: [select from LSB themes]

Benefits

- Faster documentation (don't re-enter the same fields every time)
- Consistency (all datasets have the same producer info, language, etc.)
- Reduced errors (less typing)

Best Practices: Regular Review Cycles

Regular Review Cycles

Schedule: Every 6–12 months, review all published metadata.

Questions to ask:

- Is the dataset still relevant?
- Has a new version been released?
- Are links still valid?
- Have users reported issues?

Best Practices: Version Control

Version Control

When to create a new version:

- Data corrections or updates
- Methodological changes
- New variables or indicators added
- Major metadata changes

Example: LECS 2012/13 v1.0 → v1.1 (typo fix) → v2.0 (new variables added)

Best Practices for Catalog Maintenance (Part 2)

Linking Related Datasets

Example relationships:

- LECS 2012/13 v2.0 → supersedes → LECS 2012/13 v1.0
- LECS 2018/19 → continuation of → LECS 2012/13
- LSIS 2011 → companion to → LECS 2012/13

Role-Based Workflow

- **Editor** – creates/edits metadata
- **Reviewer** – checks completeness, suggests changes
- **Custodian** – approves and publishes
- **Admin** – manages users and platform

Quality Assurance Checklist

QA Checklist (before publishing)

- Title and ID are unique
- Abstract is complete (150+ words)
- All variables have labels
- Variable labels are clear and specific
- Value labels are complete (where applicable)
- Geographic coverage is specified
- Data collection dates are specified
- Data access level is appropriate
- External resources (questionnaire, report) are attached

Advanced Preview: API Access

Metadata Editor API

What's an API? Application Programming Interface – a way for computer programs to talk to the Metadata Editor.

Use cases:

- Bulk publish metadata from Excel/CSV
- Automated metadata updates from databases
- Integration with other systems (government portals)
- Analytics on metadata completeness

Getting Started

1. Login to Metadata Editor
2. Go to User Profile → API Keys
3. Generate a new key
4. Use in R, Python, or other scripts

Advanced Preview: AI-Enhanced Metadata

LLM4Data: AI for Metadata Enrichment

What's happening at the World Bank? Researchers are using Large Language Models (LLMs) to help with metadata creation.

Capabilities:

- **Theme generation** – read variable labels → auto-suggest themes
- **Keyword extraction** – from abstract/title → suggest keywords
- **Abstract generation** – read data description → auto-write abstract

Implementation Timeline & Next Steps (Part 1)

Immediate Actions (Next 1 Month)

1. Establish metadata conventions (ID scheme, vocabularies)
2. Create Metadata Editor templates for LSB
3. Prioritize first 10 datasets from World Bank list
4. Assign editors and reviewers

Short-term (Next 2 Months)

1. Document first 5 datasets
2. Publish to LNADA
3. Solicit user feedback

Implementation Timeline & Next Steps (Part 2)

Short-term (continued)

1. Refine processes based on feedback

Long-term (This Year)

1. Complete all ~35 datasets
2. Establish regular review cycle
3. Build team capacity (training others)
4. Explore API automation

Session Wrap-Up & Q&A

Key Takeaways

- Cataloging is a **process**, not a one-off event
- **Quality** matters – good metadata takes time
- **Consistency** across datasets makes the catalog more usable
- **Legacy data** can be valuable if re-documented
- **Planning** and **roles** are essential for success

Questions?

Next: Lunch (12:00–13:30), then Session S7 on LNADA System Administration.