Standards-first approaches to building open data systems:

*The UNICEF experience*
Where we were a couple of years ago…

Education  Nutrition  WASH  Protection  Health  Poverty  Demography

Dataset  Dataset  Dataset  Dataset  Dataset  Dataset  Dataset
A “Wild Wild West” of approaches: R, STATA, SPSS, Python, Excel

Lots of manual processes

Dashboards (Tableau, Shiny, PowerBI)

Websites (data.unicef.org, childmortality.org, countdown2030)

Infographics (Infogram, Canva, Excel)

Data sources lying around all over the place.
A single shared place to store all of our raw, semi-/un-structured source data and metadata.
Where we wanted to be…

Data and structures in a single, PUBLIC shared place.

Data warehouse

- Nutrition
- WASH
- Demography
- Poverty
- Protection
- Health

Dashboards
Websites
Infographics
Where we wanted to be…

Data warehouse

Reference Data Manager

Ref data and Ref metadata in a single shared place.

DATA USE UNIT DAPM / DATA & ANALYTICS
JAN 26, 2021

YVES JAQUES, ICT SPECIALIST
And of course, we wanted a robust set of standards to help maximize the “ilities”

- Usability
- Comparability
- Exchangeability
- Maintainability
- Versionability
- Reusability
- and so on…
So we made some **choices**

- **Standards**
  - SDMX for our data and reference data
  - DDI for our raw data sources

- **Processes**
  - Standardize reference data in code lists
  - Reuse existing community code lists
  - Reuse code lists across data structures
  - Publish once, push to many channels

- **Tools**
  - Joined the SIS-CC community to share tools
  - Deployed SDMX-compliant data warehouses (.Stat & Fusion)
  - Implemented a Reference data manager (bespoke system)
  - Deployed a DDI-compliant Data Catalog (World Bank NADA)
What’s working today?

- **Source Data Catalog**
  - **Midstream** on our reboot using DDI-compliant [NADA](#). Great product!
  - 1000’s of raw data sources.

- **Indicator Data Warehouse (IDW)**
  - Virtually all of our indicator data on the state of women and children are now hosted in an [SDMX-compliant registry](#).
  - Millions of indicator data points for hundreds of indicators across 65 dataflows and 13 technical sectors.
  - Fully implements the SDMX standard.

- **Reference data and metadata manager (RDM)**
  - Bespoke system in operation and steadily gaining metadata.
  - [Public API](#) makes it easy for anyone to access.
What’s working today?

• Data Consultations
  • Bespoke tool CONSULT, for conducting UNICEF and SDG consultations with global partners. Pulls from IDW and the RDM.

• Publication channels
  • We feed a number of sites and dashboards dynamically from single source IDW using REST APIs and Web Data Connectors. More are on the way!
    • https://data.unicef.org/dv_index/
    • https://childmortality.org/
    • https://profiles.countdown2030.org/#/ds
    • https://data.humdata.org/organization/unicef-data
What did existing standards and tools NOT help us solve?

- **SDMX**
  - Has an overly complex model for Reference Metadata, and tool support for reference metadata is poor.
    - We had to build our own solution, the RDM.
  - Lack of Code List inheritance means we often create new code lists that extend existing code lists, but must do so informally.

- Time-series centric cube model leads (in our case) to:
  - Sparse and partially intersecting cubes that
  - Create terrible data discovery and comparability challenges,
  - The fact that our survey-centric data are mostly NOT time-series (i.e. w/regular periodicity) compounds this problem.
  - We had to [deploy a search engine](#) and build indexes to overcome these issues.
    - The public facing tool for this is a work-in-progress.
What did existing standards and tools NOT help us solve?

• **Working System**
  
  • Neither SDMX nor DDI nor any of their associated tools do much to overcome a central business process challenge, the need for a statistical working system.
  
  • We have begun cracking this area by first attacking microdata harmonization.
  
  • It’s probably the hardest area, and so of course, we saved it for last…😊
Thanks for listening!