



Strengthening data governance with SDMX

OCTOBER 30, 2023

Allen Boddie

The Vision, Benefits, and Opportunities of iData

The iData project was established to mitigate the risk that IMF's legacy system (EcOS) fails. It is also an opportunity to modernize data management and dissemination at the Fund.

- Key Benefits to **Data Producers** include:
 - ▶ Tools supporting enhanced collaboration
 - ▶ Integration with analytic tools
 - ▶ Off-the-shelf solutions providing access to a large ecosystem of resources

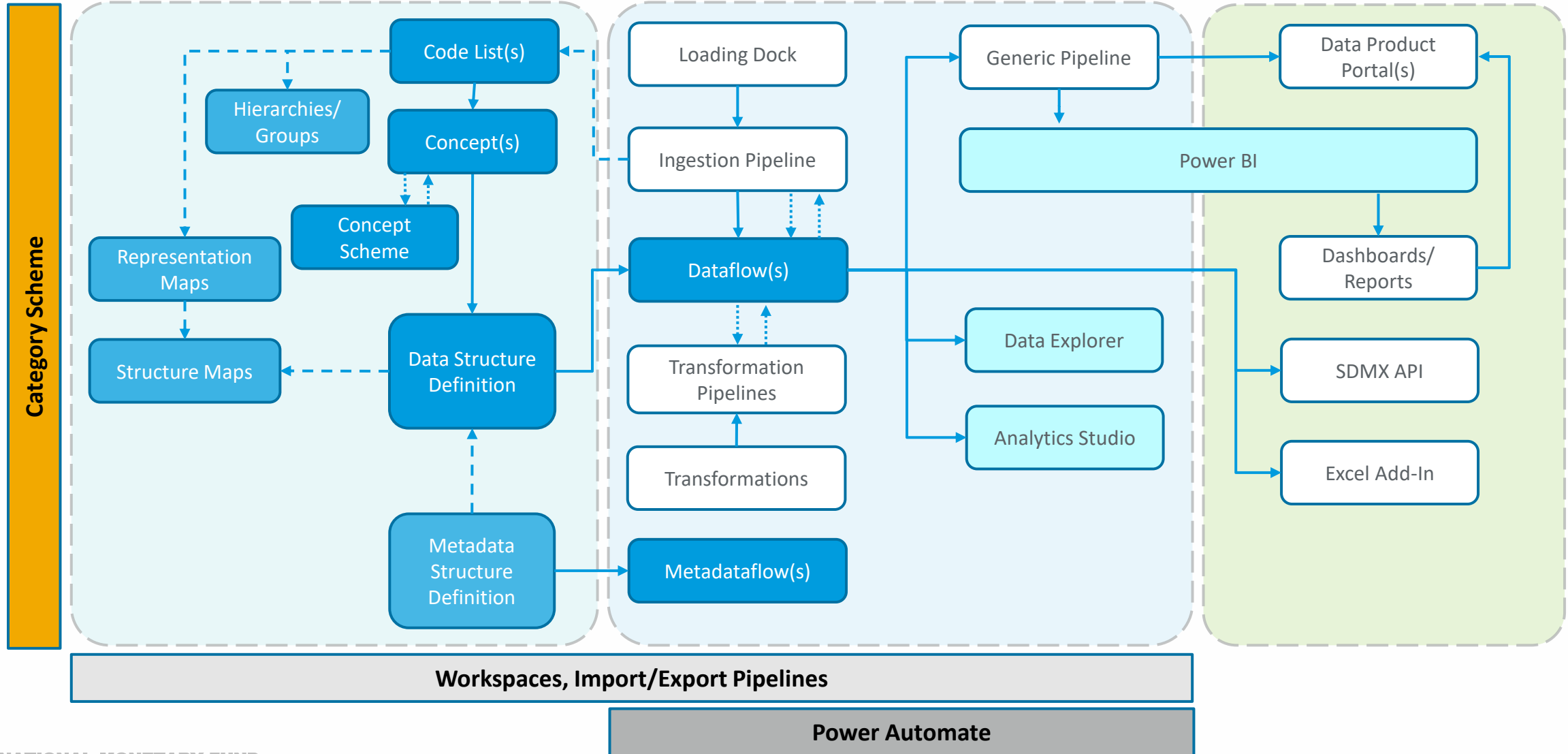
- Key Benefits to **Users** include:
 - ▶ Enhanced data access and discovery
 - ▶ Industry-leading visualization
 - ▶ Multilingual access

iData Big Picture

Design Data

Collect, Manage, Analyze Data

Disseminate Data



Problem: Inconstant Names Across Datasets

▲ Countries X ...

Type to filter items 🔍

| | |
|---|--------|
| <input checked="" type="checkbox"/> Bahrain, Kingdom of | 1,198 |
| <input checked="" type="checkbox"/> Bahrain | 345 |
| <input type="checkbox"/> Hungary | +4,390 |
| <input type="checkbox"/> Solomon Islands | +4,331 |
| <input type="checkbox"/> Cabo Verde | +4,256 |

▼

Dataset owners maintain their own list of countries relevant to their dataset. This includes country groupings specific to that dataset.

This results in:

- Deviation from IMF's official country names
- A poor experience for data users
- High level of effort for:
 - Name changes
 - Country groupings change

Solution: Inherited Code Lists

Official country code list with official names:

- Maintained by a central team
- Used as basis for cross-dataset portal facets
- Datasets which need additional terms can inherit from the official code list
- Datasets directly using official code list can reuse hierarchies built by central team
- **Remaining Issue:** Can't reuse hierarchies for inherited code lists

Structured Artifacts

- Code Lists
- Hierarchies
- Concept Schemes
- Data Structure Definitions
- Dataflows
- Metadata Structure Definitions
- Metadataflows
- Representation Maps
- Structure Maps
- Category Schemes
- Data Factory**
- Analytics
- Data Management Tools
- Settings










CONTENT ATTRIBUTES PARENT EXTENSION PROPERTIES ANNOTATIONS DEPENDENTS PRECEDENTS AUDIT Delete Clone Export

Sub-Saharan Africa Regional Economic Outlook (AFRREO) Country: 374 Add

| ID | Name ↑ | Description |
|----------------|--|---|
| 184 | Spain | |
| Inherited Term | Sri Lanka | |
| 769 | SSA Frontier and Emerging Market Economies | Countries in Sub-Saharan Africa who have access to international capi |
| 651 | SSA low-income countries excluding countries in fragile situations | |
| 703 | SSA middle-income countries excluding South Africa and Nigeria | |
| 044 | SSA Non-Oil Resource Intensive Countries | Countries in sub-Saharan Africa where nonrenewable natural resourc |
| 045 | SSA Non-Resource Intensive Countries | Countries in sub-Saharan Africa that are not classified as either oil exp |
| 699 | SSA oil-exporting countries | |
| 700 | SSA oil-exporting countries excluding Nigeria | |
| 697 | SSA oil-importing countries | |
| 650 | SSA oil-importing countries excluding South Africa | |
| 043 | SSA Resource Intensive Countries | Includes countries in Sub-Saharan Africa that are classified as oil expo |
| 856 | St. Helena | |
| 361 | St. Kitts and Nevis | |
| 362 | St. Lucia | |
| 363 | St. Pierre and Miquelon | |
| 364 | St. Vincent and the Grenadines | |
| 603 | Sub-Saharan Africa (SSA) | EMDEs located in the Sub-Sahara Africa region. |
| 604 | Sub-Saharan Africa (SSA) excluding Nigeria and South Africa | EMDEs located in the Sub-Sahara Africa region, excluding Nigeria and |
| 602 | Sub-Saharan Africa (SSA) excluding South Sudan | EMDEs located in the Sub-Sahara Africa region, excluding South Sudan |

Inherited Code List for AFR Regional Economic Outlook

Problem: Limited Global Usage

| Country | Users ↓ | New Users | Sessions |
|---|---|---|---|
| | 6,898,004 % of Total: 100.00% (6,898,004) | 6,922,374 % of Total: 100.00% (6,922,374) | 10,612,588 % of Total: 100.00% (10,612,588) |
| 1.  United States | 3,411,242 (49.29%) | 3,402,701 (49.16%) | 3,938,798 (37.11%) |
| 2.  India | 278,729 (4.03%) | 280,132 (4.05%) | 458,542 (4.32%) |
| 3.  United Kingdom | 258,097 (3.73%) | 257,927 (3.73%) | 451,113 (4.25%) |
| 4.  China | 252,831 (3.65%) | 253,575 (3.66%) | 530,834 (5.00%) |
| 5.  Japan | 191,840 (2.77%) | 193,596 (2.80%) | 344,525 (3.25%) |
| 6.  Germany | 121,990 (1.76%) | 122,044 (1.76%) | 224,239 (2.11%) |
| 7.  South Korea | 110,216 (1.59%) | 111,190 (1.61%) | 199,534 (1.88%) |
| 8.  Russia | 95,072 (1.37%) | 95,743 (1.38%) | 193,310 (1.82%) |
| 9.  Spain | 90,696 (1.31%) | 91,791 (1.33%) | 161,653 (1.52%) |

Despite the IMF severing a global membership, the usage of our data portal is limited:

- Around 50% of our users are US based
- Current portal content only available in English
- Current portal very slow away from Eastern United States

- Structured Artifacts
- Code Lists
- Hierarchies
- Concept Schemes
- Data Structure Definitions
- Dataflows
- Metadata Structure Definitions
- Metadataflows
- Representation Maps
- Structure Maps
- Category Schemes
- Data Factory
- Analytics
- Data Management Tools
- Settings

CONTENT ATTRIBUTES PARENT EXTENSION PROPERTIES ANNOTATIONS DEPENDENTS PRECEDENTS AUDIT

Delete Clone Export

Country: 368

| ID | Name EN | Name ru ↑ | Name FR | Name es | Name ar | Name zh | Name JA | Descrip |
|-----|---------------------|--------------------|--------------------|-------------------|-----------------|---------|--------------|---------|
| 193 | Australia | Австралия | Australie | Australia | أستراليا | 澳大利亚 | オーストラリア | |
| 122 | Austria | Австрия | Autriche | Austria | النمسا | 奥地利 | オーストリア | |
| 912 | Azerbaijan | Азербайджан | Azerbaïdjan | Azerbaiyán | أذربيجان | 阿塞拜疆 | アゼルバイジャン | |
| 919 | Åland Islands | Аландские острова | Åland | Islas Åland | جزر أولاند | 阿兰群岛 | オーランド諸島 | |
| 914 | Albania | Албания | Albanie | Albania | ألبانيا | 阿尔巴尼亚 | アルバニア | |
| 612 | Algeria | Алжир | Algérie | Argelia | الجزائر | 阿尔及利亚 | アルジェリア | |
| 859 | American Samoa | Американское Самоа | Samoa américaines | Samoa Americana | ساموا الأمريكية | 美属萨摩亚 | アメリカ領サモア | |
| 312 | Anguilla | Ангилья | Anguilla | Anguila | أنغيلا | 安圭拉岛 | アンギラ | |
| 614 | Angola | Ангола | Angola | Angola | أنغولا | 安哥拉 | アンゴラ | |
| 171 | Andorra | Андорра | Andorre | Andorra | أندورا | 安道尔 | アンドラ | |
| 311 | Antigua and Barbuda | Антигуа и Барбуда | Antigua-et-Barbuda | Antigua y Barbuda | أنٹیگوا وبربودا | 安提瓜和巴布达 | アンティグア・バーブーダ | |
| 213 | Argentina | Аргентина | Argentine | Argentina | الأرجنتين | 阿根廷 | アルゼンチン | |
| 911 | Armenia | Армения | Arménie | Armenia | أرمينيا | 亚美尼亚 | アルメニア | |
| 314 | Aruba | Аруба | Aruba | Aruba | أروبا | 阿鲁巴 | アルバ | |
| 512 | Afghanistan | Афганистан | Afghanistan | Afganistán | أفغانستان | 阿富汗 | アフガニスタン | |
| 313 | Bahamas, The | Багамские острова | Bahamas | Las Bahamas | جزر البهاما | 巴哈马 | バハマ | |
| 513 | Bangladesh | Бангладеш | Bangladesh | Bangladesh | بنغلاديش | 孟加拉国 | バングラデシュ | |
| 316 | Barbados | Барбадос | Barbade | Barbados | بربادوس | 巴巴多斯 | バルバドス | |
| 419 | Bahrain | Бахрейн | Bahreïn | Bahrein | البحرين | 巴林 | バーレーン | |
| 913 | Belarus | Беларусь | Belarus | Belarús | بيلاروس | 白俄罗斯 | ベラルーシ | |

Solution: Multilingual Artifacts

Solution: Multilingual Artifacts

▲ Country

Tapez pour filtrer les éléments

SelectAll

Nigéria

Ouganda

République centrafricaine

République démocratique du Congo

République du Congo

Rwanda

Remaining Issue: The SDMX standard seems to only allow for downloading all localizations of an artifacts.

- This means the API will take longer to return information needed on the portal, where a user is looking at one language, than needed
- We implemented the ability to return a single localization to improve performance, but this is not in the standard

Using Akamai CDN to improve performance for non-US based users

Problem: No Consistent Metadata for Datasets

To enable portal users to search through datasets and improve data governance around datasets we want to enforce a common set of required metadata for all published datasets.

Solution: MSD Attached to DSD

Data Structure Definition

| Concept Name |
|----------------------------|
| Status |
| ▼ Metadata Attributes (23) |
| DOI |
| Full Description |
| Dataset Methodology Notes |
| Full Source Citation |
| Short Source Citation |
| Suggested Citation |
| Dataset Comment |
| Dataset Topic |
| Dataset Keywords |
| Department |
| Publisher |
| Contact Point |

Centrally managed MSD with all required metadata fields.

- Require population of these fields for published datasets
- Information can be displayed on portal in a uniform way for all datasets

Remaining issues: Harder to maintain for datasets which need their own metadata attributes

Enforcing Use of Central Artifacts

The solutions given require different teams managing artifacts uses to support the same dataset.

To get this all to work together we use semantic versioning and wild carding:

- All artifacts in iData, even non-SDMX artifacts used in data production, follow semantic versioning.
- A non-backward compatible change is a major version change
- For most artifacts, we do not allow wildcarding of major versions

Multi-dataset Queries

We are exploring using multi-dataset queries to eliminate repetitive datasets while still allowing that selection of data to be accessible.

Multi-dataset queries are also used in the watch list on our portals

Remaining issues: The format for multidata queries does not allow for keys to be passed for individual datasets resulting in extra data being displayed.

- We use a POST request where an array of dataset objects with their own filters are passed
- [Fussion Registry](#) modifies the GET request to prefix filters with reference to dataflow

Questions?