SDMX GC 2023 – Advanced Capacity Building

SDMX, micro data in action

Stratos Nikoloutsos, Olivier Sirello (Bank for International Settlements)

A standard for micro data: trade-offs and challenges

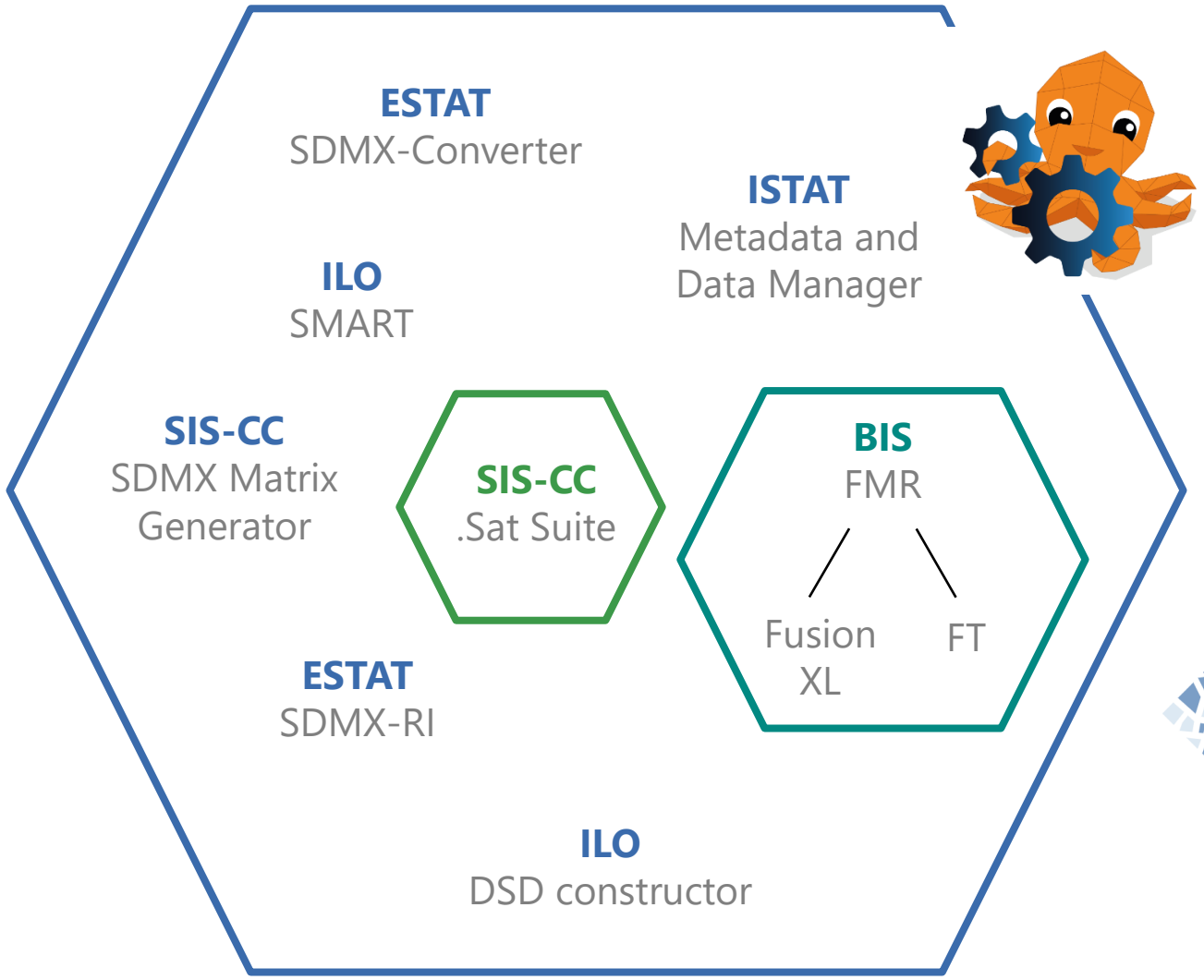
1. **Standardisation but also customisation**
   of data and metadata is key to facilitate interpretability, comparability and data lineage

2. **Reconciling micro and macro data: zooming in and out**
   with the help of SDMX 3.0

3. **Proper modelling: from a top-down to a bottom-up approach**
   to ensure consistency and standardisation across different data sets

4. **Data sharing with more performant and tailored queries**
   made simpler and more efficient via SDMX

SDMX comes with a variety of tools



The **sdmx.io** ecosystem includes a collection of open source SDMX software tools cooperating to solve official statistics use cases.

https://www.sdmx.io/tools/ecosystem/

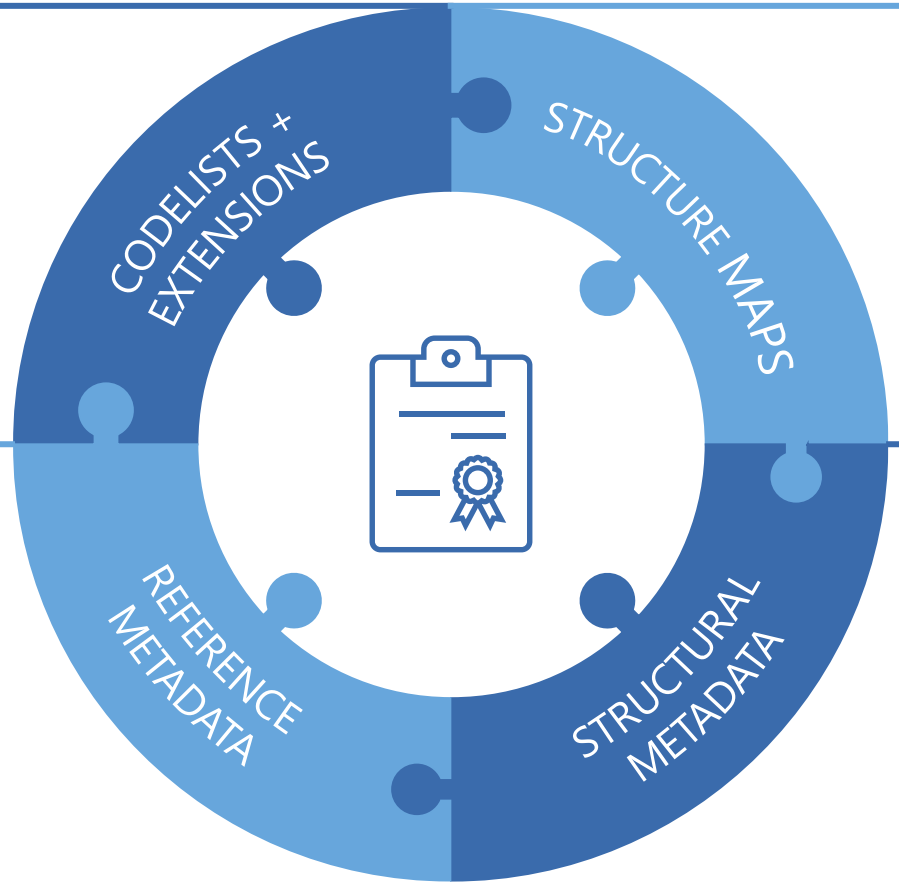
SDMX/FMR is managed by the BIS and will be used as an example in this presentation although other tools may also offer the same features

Standardisation

of data and metadata is key to facilitate interpretability, comparability and data lineage

Harmonize values across data sets thanks to standard, but customizable codes

Link components between data sets and understand their relationships



Get useful information about data, such as methodology

Define information at the level of each observation, such as its status, confidentiality or break

The right balance between standardisation and customisation

Harmonized values across data sets, but also customisable code lists

Customisation and extension of code lists

- Micro data often come with the **need for customised** and/or **extended code lists**
- This feature is key for **flexible maintenance of codes**, for instance during the collection and compilation phases

In a security-by-security data set, new ISIN codes can be appended after each data collection round

Structure and representations maps

- Data sets with micro data typically are split into **multiple tables**
- Structure maps are key to **describe the relationships** between each of them

Structure maps can be used to describe the relationships between the columns of multiple tables
(also allow to map custom internal codes to standard codes leveraging representation maps)

Structural metadata

- Structural metadata are key to describe statistical data, for each at data set, series, observation and measure level
- With **SDMX 3.0**, a **list of values for attributes is allowed**, increasing the flexibility of the data modelling notably required for micro data

In a security-by-security data set, SDMX 3.0 allows to set attributes for multiple measures,
such as face, nominal and market value per each security per period

Extension of Codelists

ISO_3166-A2
(249 Codes)

How can I…
- add 10 legacy Country Codes
- add continents and regions (eg 29 Codes)
- change a few labels (eg 20 Codes)

Before SDMX 3.0:
- Create a new Country Codelist with 259 Codes
- Create a new Area Codelist with 278 Codes
- Create a new Country Codelist with 249 Codes

CL_COUNTRY
(259 Codes)

CL_AREA
(278 Codes)

CL_COUNTRY
(249 Codes)

Extension of Codelists

ISO_3166-A2
(249 Codes)

How can I…
- add 10 legacy Country Codes
- add continents and regions (eg 29 Codes)
- change a few labels (eg 20 Codes)

With SDMX 3.0:
- Extend ISO_3166-A2 with 10 Codes
- Extend ISO_3166-A2 with 29 Codes
- Extend ISO_3166-A2 with 20 Codes

CL_COUNTRY
(10 Codes)

CL_AREA
(29 Codes)

CL_COUNTRY
(20 Codes)

Arrays

Observation Status

Viewing: Observation Status [2.2]

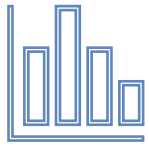
Position	Id	Name	Description
1	A	Normal value	To be used as default value if no value is provided or when no special coded qualificati...
2	B	Time series break	Observations are characterised as such when different content exists or a different met...
3	D	Definition differs	Used to indicate slight deviations from the established methodology (footnote-type info...
4	E	Estimated value	Observation obtained through an estimation methodology (e.g. to produce back-casts) ...
5	F	Forecast value	Value deemed to assess the magnitude which a quantity will assume at some future p...
6	G	Experimental value	Data collected on the basis of definitions or (alternative) collection methods under dev...
7	I	Value imputed by a receiving agency	Observation imputed by a receiving agency to replace or fill gaps in reported data serie...
8	K	Data included in another category	This code is used when data for a given category are missing and are included in anot...
9	W	Includes data from another category	This code is used when data include another category, or go beyond the scope of the ...
10	O	Missing value	This code is to be used when no breakdown is made between the reasons why data ar...
11	M	Missing value; data cannot exist	Used to denote empty cells resulting from the impossibility to collect a statistical value ...
12	P	Provisional value	An observation is characterised as "provisional" when the source agency - while it bas...
13	S	Strike and other special events	Special circumstances (e.g. strike) affecting the observation or causing a missing value.
14	L	Missing value; data exist but were not collected	Used, for example, when some data are not reported/disseminated because they are b...
15	H	Missing value; holiday or weekend	Used in some daily data flows.
16	Q	Missing value; suppressed	Used, for example, when data are suppressed due to statistical confidentiality consider...
17	J	Derogation	Clause in an agreement (e.g. legal act, gentlemen's agreement) stating that some prov...
18	N	Not significant	Used to indicate a value which is not a "real" zero (e.g. a result of 0.0004 rounded to z...
19	U	Low reliability	This indicates existing observations, but for which the user should also be aware of the...

Showing 1 to 19 of 20 entries

Search

Arrays as value for Attribute/Measure

Observation Status



How can I…
- provide 2 statuses for an observation?
- provide more than one statuses without knowing the exact number on design time?

Before SDMX 3.0:
- Add two OBS_STATUS Attributes
- ?

```
OBS_STATUS_1: F (Forecast)
OBS_STATUS_2: U (Low reliability)
```

```
OBS_STATUS_1: F (Forecast)
OBS_STATUS_2: U (Low reliability)
...
OBS_STATUS_N: D (Definition differs)
```

Arrays as value for Attribute/Measure

Observation Status

How can I…
- provide 2 statuses for an observation?
- provide more than one statuses without knowing the exact number on design time?

With SDMX 3.0:
- Add one OBS_STATUS Attribute with max = 2
- Add one OBS_STATUS Attribute with unbounded upper limit (or high enough)

```
OBS_STATUS: [
  F (Forecast),
  U (Low reliability)
]
```

```
OBS_STATUS: [
  F (Forecast),
  U (Low reliability),
  ...
  D (Definition differs)
]
```

Easy reconciliation of micro and macro data

with the help of SDMX

Hierarchies to quickly drill down from aggregates or conversely

Mappings to link representations and understand relationships between concepts

Track the validity of concepts and codes, their groupings, as well as their mappings, across time

Attributes at multiple levels with fine grained attachment

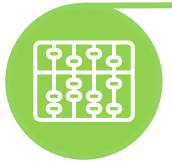
From micro to macro and from macro to micro

Quickly drill down from aggregates and conversely

Hierarchies to zoom in and zoom out

- Hierarchies are key to **drill down on the most granular level** from aggregates
- Groups and hierarchies share the same standardized codes, ensuring consistency

Share the codes across different groups, such as a country belonging to multiple economic groupings
Derive from the hierarchy the underlying entities that have been aggregated



Mappings to better understand relationships between concepts

- Map representations, also **leveraging regular expressions**, to other representations and concepts
- Mapping also include **free text** and can be **one-to-many or many-to-many**

Map the initial two letters of the ISIN code to the country dimension

Attributes and multiple measures

- Measure-specific attributes: **an attribute** can be explicitly related **to one or more measures**

As an example, it might be possible define attribute A “1” for Gender and attribute A “2” for Occupation and filter according to their values

Hierarchies in SDMX 3.0

Viewing: Country groupings in the Data Dictionary [1.0]

- G000 - World
  - G100 - Africa
    - G110 - Northern Africa
      - DZ - Algeria
      - EG - Egypt
      - LY - Libya
      - MA - Morocco
      - SD - Sudan
      - TN - Tunisia
      - EH - Western Sahara
    - G120 - Sub-Saharan Africa
      - G121 - Eastern Africa
        - IO - British Indian Ocean Territory (the)
        - BI - Burundi
        - KM - Comoros
        - DJ - Djibouti
        - ER - Eritrea

United Nations
Statistics Division

Time period:

Viewing: Country groupings in the Data Dictionary [1.0]

- PT - Portugal
- EU - European Union (the)
  - BE - Belgium
  - DE - Germany
  - DK - Denmark
  - ES - Spain
  - FR - France
  - GB - United Kingdom (the)
  - GR - Greece
  - IE - Ireland
  - IT - Italy
  - LU - Luxembourg
  - NL - Netherlands (the)
  - PT - Portugal
- EU- - European Union (the) excl. Luxembourg
- XM - Euro Area
- EURO- - Euro Area excl. Luxembourg

Time period: 1990

Invalid codes:

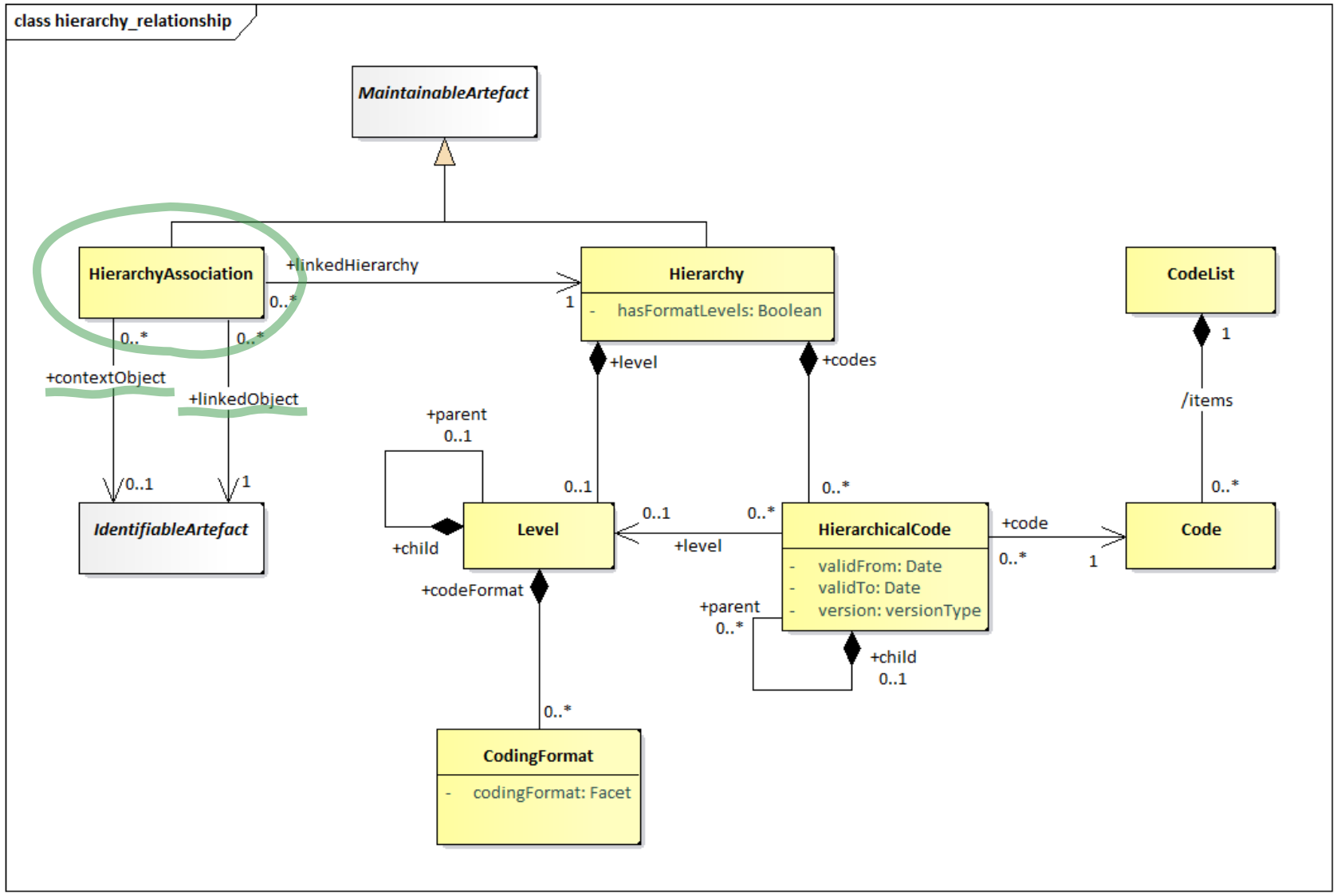
Hide

Hierarchies in SDMX 3.0

May be related to a context
(Hierarchy Association)

Linked to an object
(eg a Dimension)

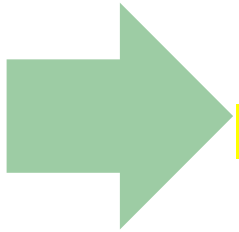
For a given context
(eg a Dataflow)



Structure and Representation Mappings

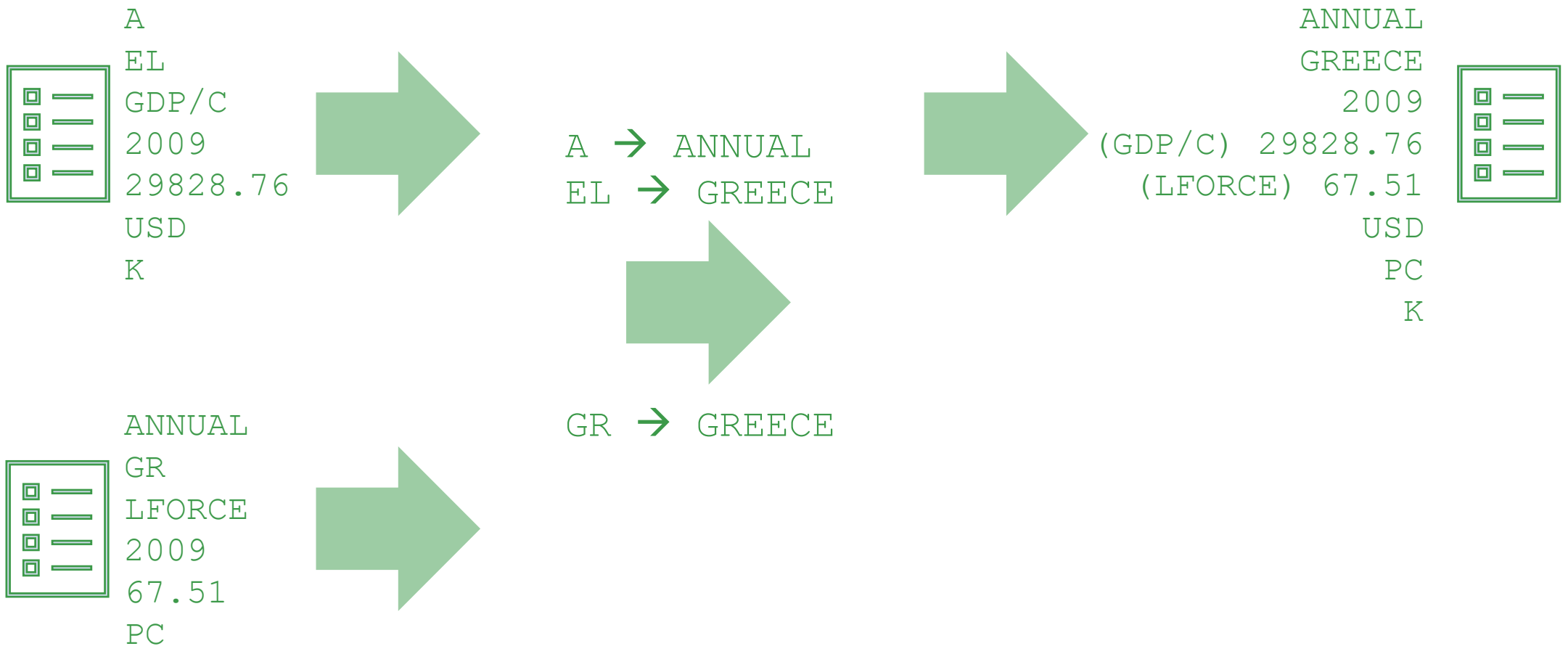
ANNUAL
GREECE
GDP/C
2009
29828.76
USD
K

ANNUAL
GREECE
LFORCE
2009
67.51
PC



ANNUAL
GREECE
2009
(GDP/C) 29828.76
(LFORCE) 67.51
USD
K
PC

Structure and Representation Mappings



Structure and Representation Mappings

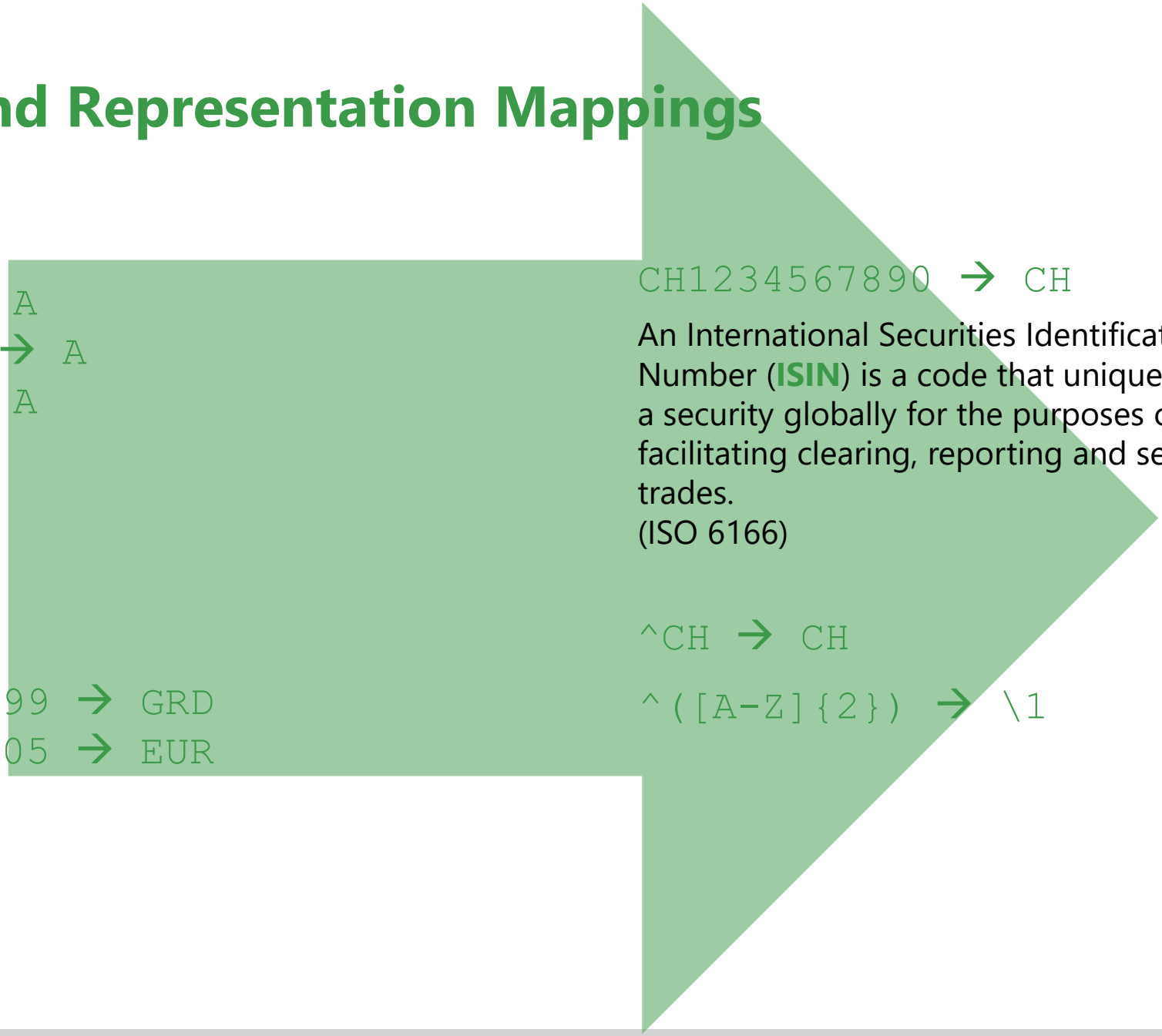
A
EL
GDP/C
2009
29828.76
US dollars
K

ANNUAL
GR
LFORCE
2009
67.51
PC

US Dollars → USD

ANNUAL
GREECE
2009
(GDP/C) 29828.76
(LFORCE) 67.51
USD
PC
K

Structure and Representation Mappings

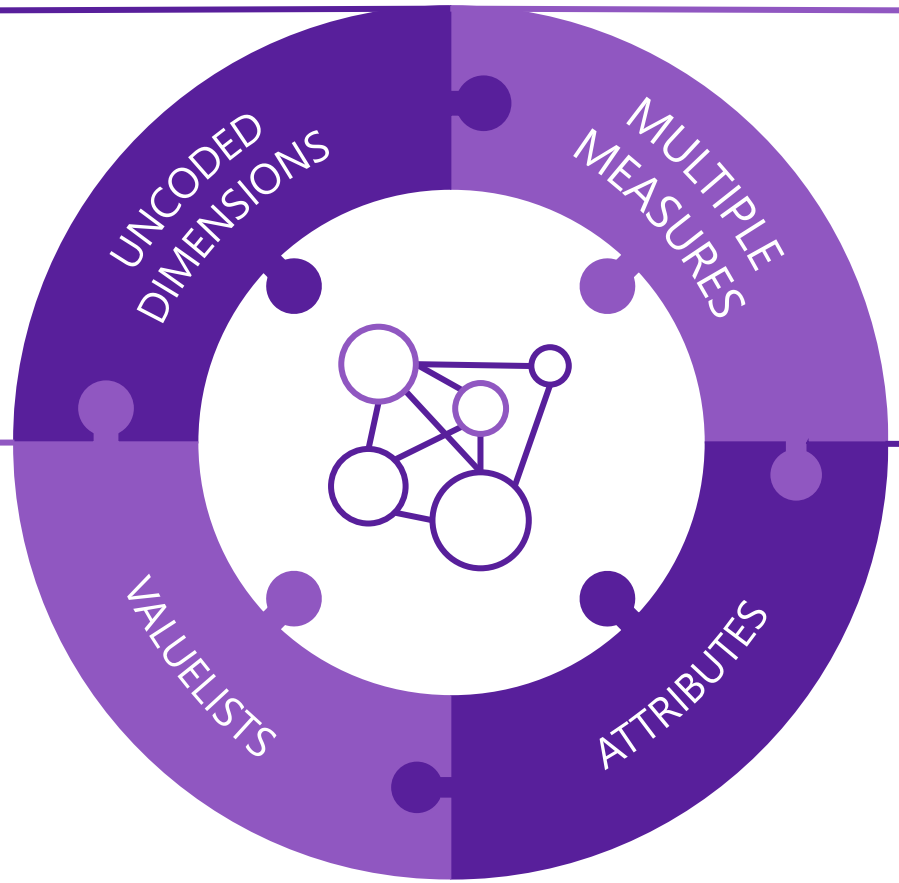


Proper modelling

of micro data ensures consistency and standardisation across different data sets

Leveraging patterns to define the uncoded dimensions of the data

Multiple measures to model micro data as one observation



ValueLists as flexible sets of characters

Attributes can take unbounded lists of values to give more flexibility in the modelling

Achieve a flexible modeling, a bottom-up perspective

Leveraging patterns to define the uncoded dimensions of the data

A component based on a **value domain that follows a pattern**, without requiring the creation of a list of code

It is possible to derive the country ISO2 code from the first two letters of the column "ISIN code" from a security-by-security database
ISIN code –> Reference area, thus CH0000000000 -> CH

Multiple measures to model micro data as one observation

More than one measurement per record, allowing also to provide fine grained metadata per measure – rationalizing/simplifying data modelling

A **security-by-security table may contain three measures** for the amount outstanding, **face**, **nominal** and **market** value.
SDMX allows to **define attributes at the measure level**, for example to flag confidential only some specific values.
It also allows to define **several statuses for a given value**, eg *provisional* and *unvalidated* value for market value on 2023-20

Time period	ISIN code	Face value	Nominal value	Market value
2023-10	CH0123456789	12	11.5	14[P, V]
2023-09	CH0123456789	12	11.6[CONF]	13
2023-08	CH0123456789	12	12	15

Achieve a flexible modeling, a bottom-up perspective

Leveraging patterns to define the uncoded dimensions of the data

A component based on a **value domain that follows a pattern**, without requiring the creation of a list of code

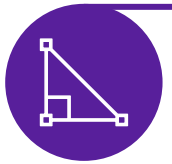
It is possible to derive the country ISO2 code from the first two letters of the column “ISIN code” from a security-by-security db
ISIN code –> Reference area, thus CH0000000000 -> CH

```
<str:Dimension>
    <str:ConceptIdentity>urn:...Concept=SDMX:CONCEPTS(1.0).ISIN</str:ConceptIdentity>
    <str:LocalRepresentation>
        <str:TextFormat pattern="^[A-Z]{2}[A-Z0-9]{9}\d$" />
    </str:LocalRepresentation>
</str:Dimension>
```

Two-letter country code: exactly two uppercase letters from 'A' to 'Z'

The 9-character alphanumeric code that follows the country code: can be uppercase letters or digits from '0' to '9'.

Achieve a flexible modeling, a bottom-up perspective



Leveraging patterns to define the uncoded dimensions of the data

A component based on a **value domain that follows a pattern**, without requiring the creation of a list of code

It is possible to derive the country ISO2 code from the first two letters of the column “ISIN code” from a security-by-security db
ISIN code –> Reference area, thus CH0000000000 -> CH

```
<str:Dimension>
    <str:ConceptIdentity>urn:...Concept=SDMX:CONCEPTS(1.0).ISIN</str:ConceptIdentity>
    <str:LocalRepresentation>
        <str:TextFormat
pattern="^(?:AD|AE|AF|AG|AI|AL|AM|AO|AQ|AR|AS|AT|AU|AW|AX|AZ|BA|BB|BD|BE|BF|BG|BH|BI|BJ|
BL|BM|BN|BO|BQ|BR|BS|BT|BV|BW|BY|BZ|CA|CC|CD|CF|CG|CH|CI|CK|CL|CM|CN|CO|CR|CU|CV|CW|CX|C
Y|CZ|DE|DJ|DK|DM|DO|DZ|EC|EE|EG|EH|ER|ES|ET|FI|FJ|FK|FM|FO|FR|GA|GB|GD|GE|GF|GG|GH|GI|GL
|GM|GN|GP|GQ|GR|GS|GT|GU|GW|GY|HK|HM|HN|HR|HT|HU|ID|IE|IL|IM|IN|IO|IQ|IR|IS|IT|JE|JM|JO|
JP|KE|KG|KH|KI|KM|KN|KP|KR|KW|KY|KZ|LA|LB|LC|LI|LK|LR|LS|LT|LU|LV|LY|MA|MC|MD|ME|MF|MG|M
H|MK|ML|MM|MN|MO|MP|MQ|MR|MS|MT|MU|MV|MW|MX|MY|MZ|NA|NC|NE|NF|NG|NI|NL|NO|NP|NR|NU|NZ|OM
|PA|PE|PF|PG|PH|PK|PL|PM|PN|PR|PS|PT|PW|PY|QA|RE|RO|RS|RU|RW|SA|SB|SC|SD|SE|SG|SH|SI|SJ|
SK|SL|SM|SN|SO|SR|SS|ST|SV|SX|SY|SZ|TC|TD|TF|TG|TH|TJ|TK|TL|TM|TN|TO|TR|TT|TV|TW|TZ|UA|U
G|UM|US|UY|UZ|VA|VC|VE|VG|VI|VN|VU|WF|WS|YE|YT|ZA|ZM|ZW)[A-Z0-9]{9}\d$" />
```

Achieve a flexible modeling, a bottom-up perspective

Multiple measures to model micro data as one observation

More than one measurement per record, allowing also to provide fine grained metadata per measure – rationalizing/simplifying data modelling

A **security-by-security table may contain three measures** for the amount outstanding, **face**, **nominal** and **market** value.
SDMX allows to **define attributes at the measure level**, for example to flag confidential only some specific values.
It also allows to define **several statuses for a given value**, eg *provisional* and *unvalidated* value for market value on 2023-20

Time period	ISIN code	Face value	Nominal value	Market value
2023-10	CH0123456789	12	11.5	14 P, V
2023-09	CH0123456789	12	11.6 CONF	13
2023-08	CH0123456789	12	12	15

Achieve a flexible modeling, a bottom-up perspective

Multiple measures to model micro data as one observation

More than one measurement per record, allowing also to provide fine grained metadata per measure – rationalizing/simplifying data modelling

A **security-by-security table may contain three measures** for the amount outstanding, **face**, **nominal** and **market** value.
SDMX allows to **define attributes at the measure level**, for example to flag confidential only some specific values.
It also allows to define **several statuses for a given value**, eg *provisional* and *unvalidated* value for market value on 2023-20

Time period	ISIN code	Face value	Nominal value	Market value
2023-10	CH0123456789	12	11.5	14 P, V
2023-09	CH0123456789	12	11.6 CONF	13
2023-08	CH0123456789	12	12	15

Achieve a flexible modeling, a bottom-up perspective

Multiple measures to model micro data as one observation

More than one measurement per record, allowing also to provide fine grained metadata per measure – rationalizing/simplifying data modelling

A **security-by-security table may contain three measures** for the amount outstanding, **face**, **nominal** and **market** value.

SDMX allows to **define attributes at the measure level**, for example to flag confidential only some specific values.

It also allows to define **several statuses for a given value**, eg *provisional* and *unvalidated* value for market value on 2023-20

```
<str:Measure>
    <str:ConceptIdentity>urn:...Concept=SDMX:CONCEPTS(1.0).FACE</str:ConceptIdentity>
</str:Measure>
<str:Measure>
    <str:ConceptIdentity>urn:...Concept=SDMX:CONCEPTS(1.0).NOMINAL</str:ConceptIdentity>
</str:Measure>
<str:Measure>
    <str:ConceptIdentity>urn:...Concept=SDMX:CONCEPTS(1.0).MARKET</str:ConceptIdentity>
</str:Measure>
```

Achieve a flexible modeling, a bottom-up perspective

Multiple measures to model micro data as one observation

More than one measurement per record, allowing also to provide fine grained metadata per measure – rationalizing/simplifying data modelling

A **security-by-security table may contain three measures** for the amount outstanding, **face**, **nominal** and **market** value.
SDMX allows to **define attributes at the measure level**, for example to flag confidential only some specific values.
It also allows to define **several statuses for a given value**, eg *provisional* and *unvalidated* value for market value on 2023-20

```
<str:Attribute>
    <str:ConceptIdentity>urn:...Concept=SDMX:CONCEPTS(1.0).CONF</str:ConceptIdentity>
    <str:AttributeRelationship>
        <str:Observation/>
    </str:AttributeRelationship>
    <str:MeasureRelationship>
        <str:Measure>FACE</str:Measure>
    </str:MeasureRelationship>
</str:Attribute>
```

Achieve a flexible modeling, a bottom-up perspective

Multiple measures to model micro data as one observation

More than one measurement per record, allowing also to provide fine grained metadata per measure – rationalizing/simplifying data modelling

A **security-by-security table may contain three measures** for the amount outstanding, **face**, **nominal** and **market** value.
SDMX allows to **define attributes at the measure level**, for example to flag confidential only some specific values.
It also allows to define **several statuses for a given value**, eg *provisional* and *unvalidated* value for market value on 2023-20

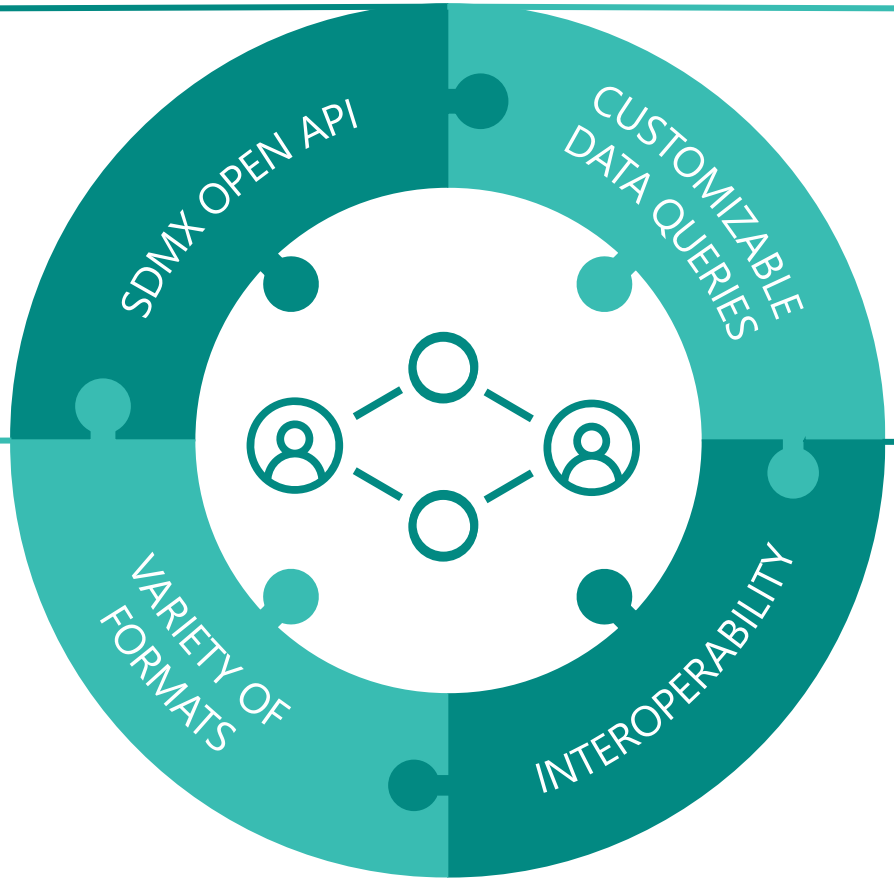
```
<str:Attribute>
    <str:ConceptIdentity>urn:...Concept=SDMX:CONCEPTS(1.0).CONF</str:ConceptIdentity>
    <str:LocalRepresentation minOccurs="1" maxOccurs="3">
        <str:Enumeration>urn:...Codelist=SDMX:CL_OBS_STATUS(1.0)</str:Enumeration>
    </str:LocalRepresentation>
    ...
</str:Attribute>
```

Data sharing

made simpler and more efficient via SDMX

SDMX Open API to programmatically access data, metadata and structures

A variety of formats to retrieve and store data and metadata



Customizable data queries, custom filters or specific values retrieval

Interoperability with other formats such as xBRL

Open API, more formats and interoperability

A powerful new Open API with increased flexibility

- Accessing data, metadata, structures within a **client application** also to ease their maintenance
- New parameters and operators for more flexible data and metadata querying

Querying for data for a range of values of a measure or attribute, including string matching

A variety of formats to retrieve and store data and metadata

- More data and metadata formats
- **Combining data with** reference **metadata**

New powerful XML messages, covering all standards
JSON targeting data visualization with combined data and structural metadata
CSV for human readable/processible datasets

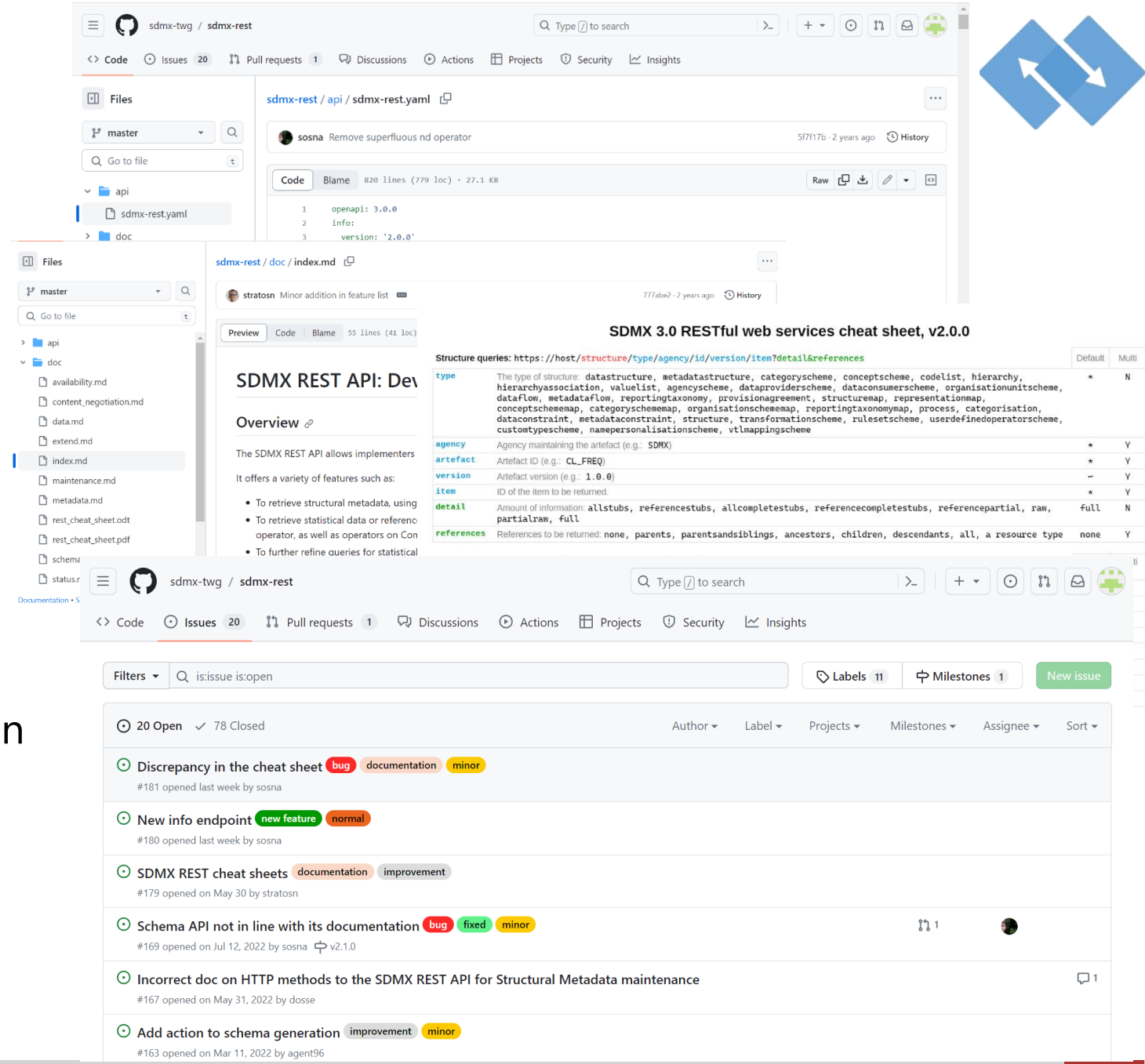
Interoperability with other formats (xBRL)

- Ongoing work to **link the two standards**

xBRL-SDMX converter to facilitate the interoperability

SDMX 3.0 RESTful API v2.0.0

- The normative part of the specification, i.e. the Open API definition
- The Developers' documentation, including a cheat sheet
- Request features and report issues on GitHub



The structure queries

- Re-organized and enriched
- Supports multiple instances of search terms, wildcarding

The unique identifier:
- Agency ID
- Artefact ID
- Version
- Item ID (for Item Schemes)

https://host/structure/type/agency/id/version/item?detail&references

The type of structure:
datastructure, metadatastructure, categoryscheme, conceptscheme, codelist, hierarchy, hierarchyassociation, valuelist, agencyscheme, dataproviderscheme, dataconsumerscheme, organisationunitscheme, dataflow, metadataflow, reportingtaxonomy, provisionagreement, structuremap, representationmap, conceptschememap, categoryschememap, organisationschememap, reportingtaxonomymap, process, categorisation, dataconstraint, metadataconstraint, structure, transformationscheme, rulesetscheme, userdefinedoperatorscheme, customtypescheme, namepersonalisationscheme, vtlmappingscheme

Amount of information:
allstubs, referencestubs, allcompletestubs, referencecompletestubs, referencepartial, raw, partialraw, full

References to be returned:
none, parents, parentsandsiblings, ancestors, children, descendants, all, a resource type

The data queries

The unique identifier of the context:
- Agency ID
- Artefact ID
- Version

https://host/data/context/agency/id/version/key?c
&updatedAfter
&firstNObservations
&lastNObservations
&attributes
&measures
&dimensionAtObservation
&includeHistory

The context of data retrieval:
datastructure, dataflow,
provisionagreement

Key(s) of the series to be returned:
eg M.GR.EUR.SP00

with wildcarding:
eg M.*.EUR.SP00

With support for multiple keys:
eg M.GR.EUR.SP00, M.CY.EUR.SP00

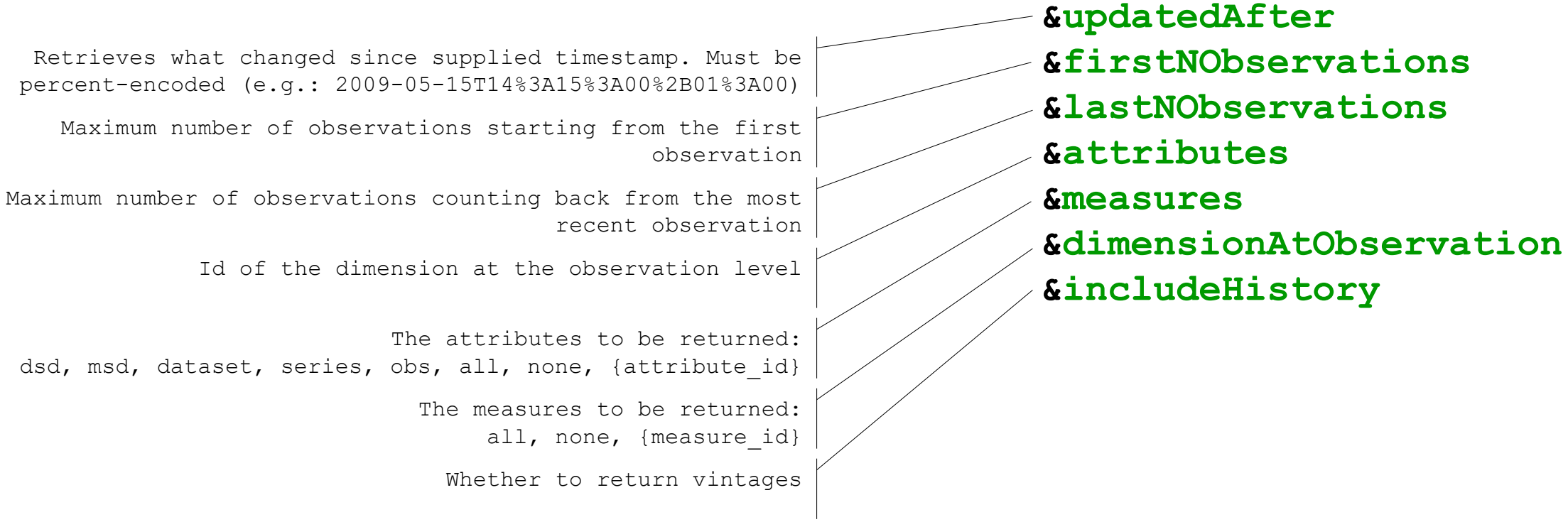
Component-based filters (for any Dimension, Attribute or Measure):
eg **c**[REF_AREA]=CH&**c**[CONF_STATUS]=F

Support for operators:
eg **c**[ICP_ITEM]=**sw**:01&**c**[TIME_PERIOD]=**ge**:2015

eq	Equal
ne	Not equal
lt	Less than
le	Less than or equal to
gt	Greater than
ge	Greater than or equal to
co	Contains
nc	Does not contain
sw	Starts with
ew	Ends with
nd	And
or	Or

The data queries

https://host/data/context/agency/id/version/key?c



Other queries

- Data validity

```
https://host/schema/context/agency/id/version?dimensionAtObservation
                                              &explicitMeasure
```

- Data availability

```
https://host/availability/context/agency/id/version/key/componentId?c
                                                                   &updatedAfter
                                                                   &references
                                                                   &mode
```

- Metadata

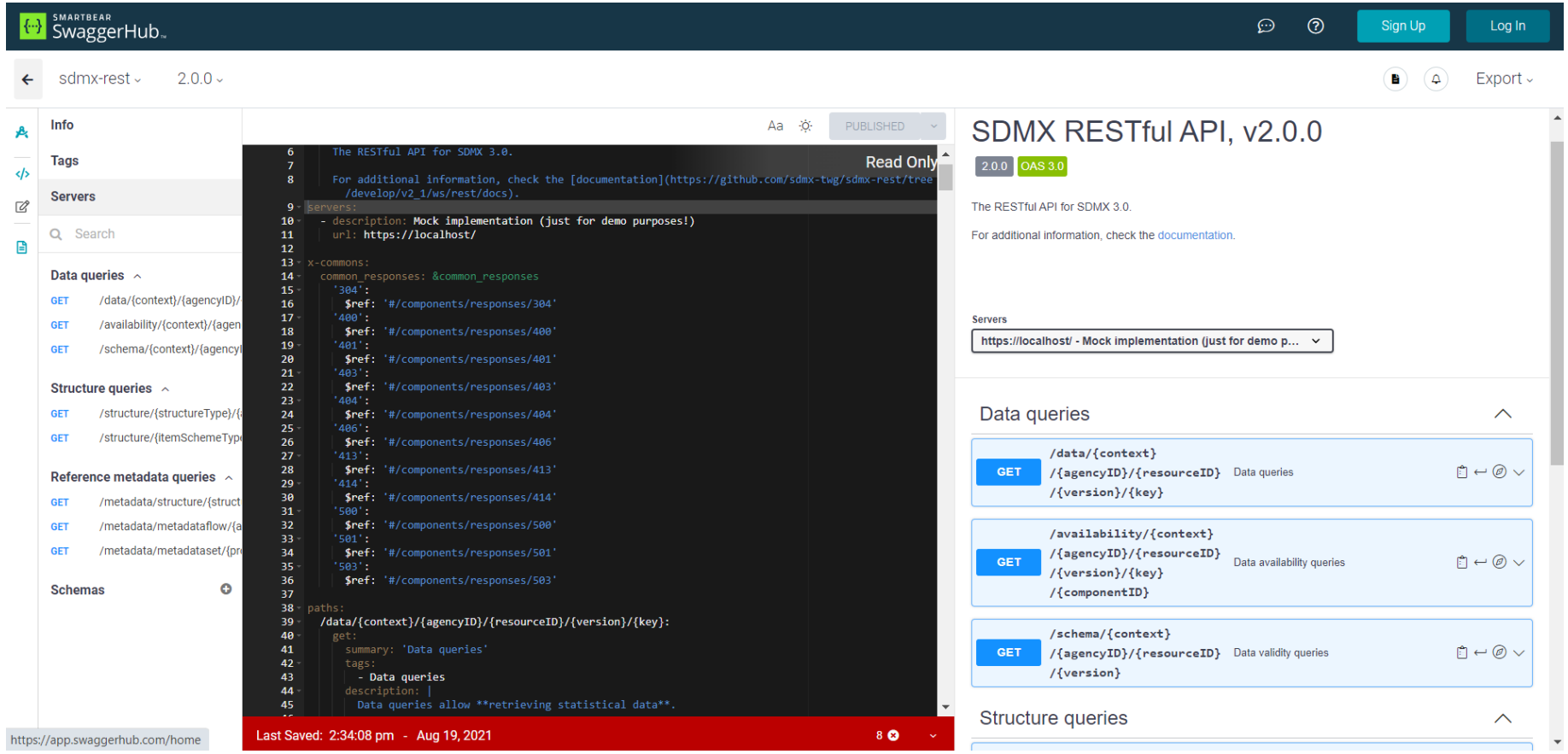
```
https://host/metadata/metadataset/provider/id/version?detail

https://host/metadata/metadataflow/agency/id/version/provider?detail

https://host/metadata/structure/type/agency/id/version/provider?detail
```

The formats

SDMX-ML Data	application/vnd.sdmx.data+xml;version=3.0.0
SDMX-ML Structure	application/vnd.sdmx.structure+xml;version=3.0.0
SDMX-ML Metadata	application/vnd.sdmx.metadata+xml;version=2.0.0
SDMX-JSON Data	application/vnd.sdmx.data+json;version=2.0.0
SDMX-JSON Structure	application/vnd.sdmx.structure+json;version=2.0.0
SDMX-JSON Metadata	application/vnd.sdmx.metadata+json;version=2.0.0
SDMX-CSV Data	application/vnd.sdmx.data+csv;version=1.0.0
SDMX-CSV Metadata	application/vnd.sdmx.metadata+csv;version=2.0.0

The SDMX v2 API in action

- See the API spec on SwaggerHub



Thank you!

Olivier.Sirello@bis.org

Stratos.Nikoloutsos@bis.org