

# Fisheries business Metadata & The CWP role for global harmonization

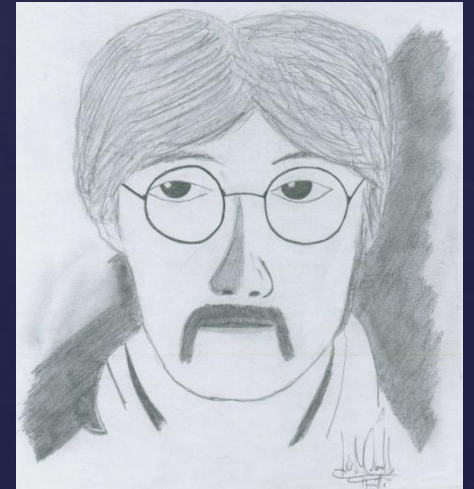
FAO, Erik van Ingen  
Swakopmund, Namibia, 25-27 February 2015

all the people could

find – use - load – display -  
interpret - interchange –  
analyze

all fisheries data

Imagine



Harmonization  
Standardization  
*Interoperability*

Better data at lower cost  
Fisheries Management

Business Case



**FLUX** - Fisheries Language for Universal eXchange - A single language for all data exchanges

**SDMX** - Statistical Data and Metadata eXchange - Foster standards for the exchange of statistical information.

**DDI** - Data Documentation Initiative - A metadata specification for the social and behavioral sciences

FiMES, DarwinCore, OGC and ISO Observations and Measurements , etc.

# Information Standards

standard	body
FLUX	UN/CEFACT
SDMX	BIS - ECB - EUROSTAT - IMF - OECD - UN - World Bank
DDI	Widespread community
FiMES	FIRMS & FAO
DarwinCore	TDWG (Biodiversity Information Standards)
OGC WFS, WMS, WCS	Number of Active Members: 509 WFS accepted as ISO Standard
Observations and Measurements	ISO

# Standard setting bodies

**Observation data** (also referred to as raw data or operational data ) is a measure taken at a certain moment in time (FLUX)

**Statistical data (timeseries)** is often processed and aggregated observation data (SDMX)

# Observation data & Statistical data

Sample #	pH
1	6.39
2	6.42
3	6.35
4	6.85

# Generic Statistical Business Process Model

from United Nations Economic Commission for Europe (UNECE)

GSBPM

Quality Management / Metadata Management							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	2.3 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Produce dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	2.5 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

The GSBPM is a reference model for the 'statistical data process'

The GSBPM can help to put information standards within a context

# GSBPM & CWP

Quality Management / Metadata Management							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & inputs	6.4 Apply disclosure control	7.4 Produce dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			



**Quality Management / Metadata Management**

Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

Quality Management / Metadata Management

1	2	3	4	5	6	7	8	9
Specify Needs	Design	Build	Collect	Process	Analyze	Disseminate	Archive	Evaluate
1.1 Determine needs for information	2.1 Design outputs	3.1 Build data collection instruments	4.1 Select complete	5.1 Integrate data	6.1 Review draft outputs	7.1 Update output products	8.1 Define archive rules	9.1 Gather evaluation inputs
1.2 Consider constraints	2.2 Design interface definitions	3.2 Build or enhance process components	4.2 Setup collection	5.2 Classify records	6.2 Update outputs	7.2 Produce dissemination products	8.2 Manage for products	9.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design data collection methodology	3.3 Integrate instruments	4.3 Run collection	5.3 Perform quality control	6.3 Update outputs	7.3 Review release of dissemination products	8.3 Archive data and associated metadata	9.3 Develop action plan
1.4 Identify concepts	2.4 Design data transfer controls methodology	3.4 Test production system	4.4 Manage collection	5.4 Estimate inventory & validation needs	6.4 Apply database control	7.4 Review dissemination products	8.4 Dispose of obsolete metadata	
1.5 Check data availability	2.5 Design automated processing methodology	3.5 Test automated systems	5.5 Conduct quality control	6.5 Analyze outputs	7.5 Produce dissemination products			
1.6 Prepare budget & cost	2.6 Design production cycle & workflow	3.6 Finalize production system	6.6 Calculate aggregate	7.6 Manage user support				
			6.7	6.8				
				Finalize data files				

DDI

SDMX DDI

Schema based has 2 layers:

1. schema (domain specific)
2. data

Artefact based has 3 layers:

1. schema
2. artefact (domain specific)
3. data

Schema or  
artefact based



*For the techies in the audience:*

- ⌘ A **schema** is often an XML schema
- ⌘ An **artefact** is often an XML document
- ⌘ The **data** is often an XML document, also called a **message**

Note: XML has a successor called JSON, which is more lightweight, less verbose and web friendly

# Schema or artefact based



An information standard can be *schema* or *artefact* based.

*Schema* or *artefact* refers to the type of information standard. **Or:** it refers to the style or approach taken when the information standard was developed.

Schema or  
artefact based



```
graph LR; Data([Data  
(message)]) -- depends directly on the --> Schema[Schema  
(metadata, domain specific*1)];
```

Data

(message)

depends directly on the

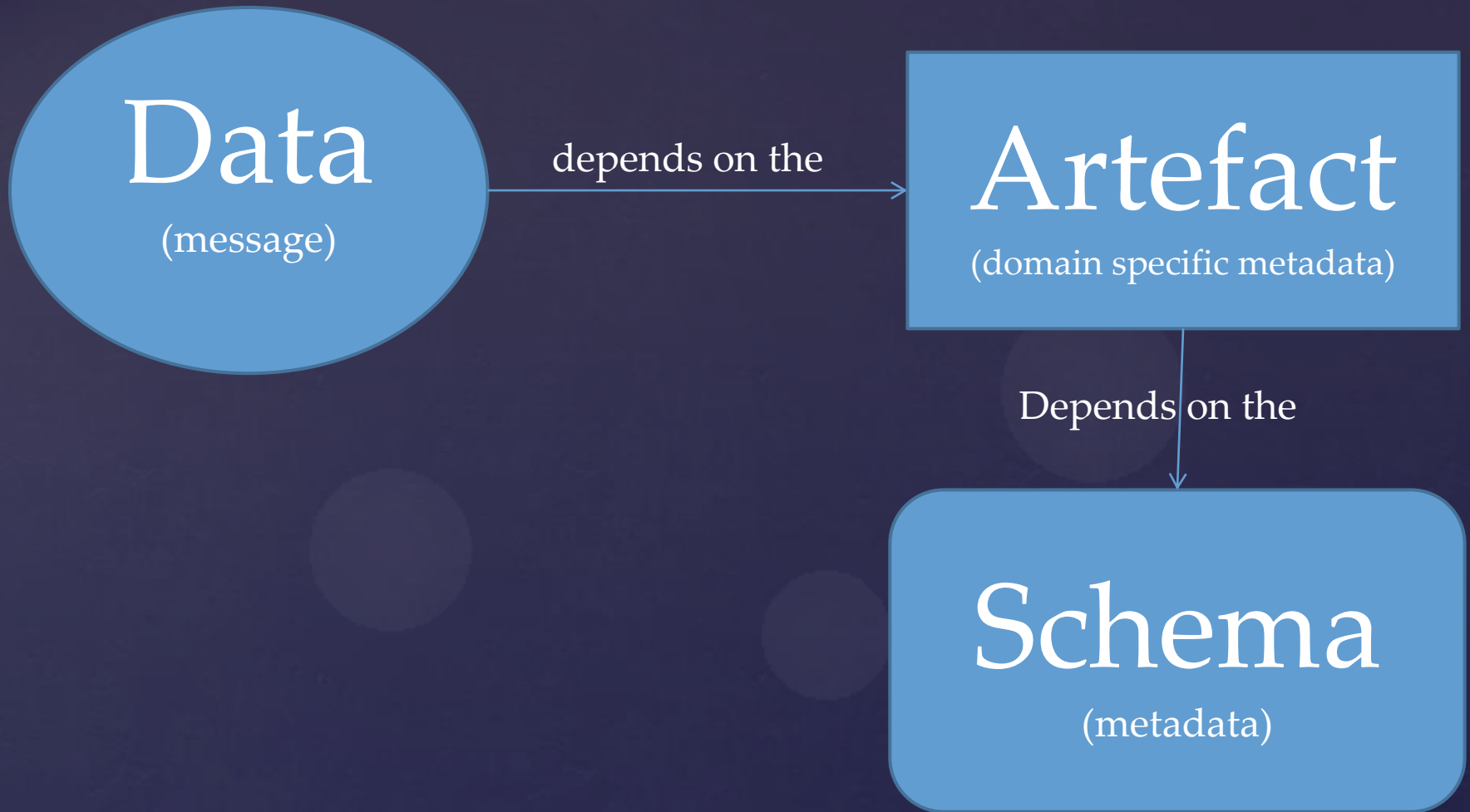
Schema

(metadata, domain specific\*1)

Schema based

\*1) In this presentation we boldly consider all schema based standards as domain specific.

In reality this is not always the case, because a domain generic schema based standard can have domain specific extensions.



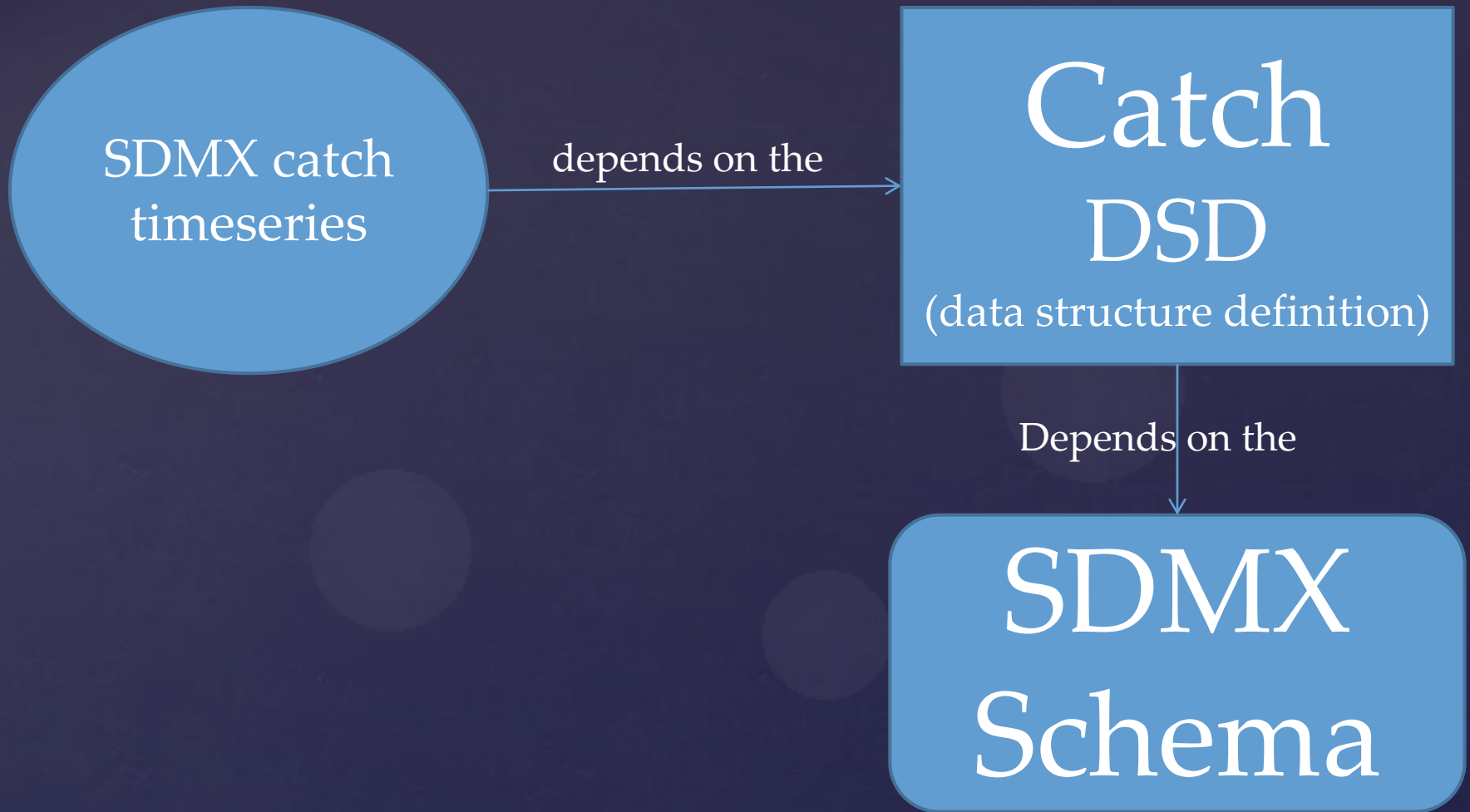
Artefact based

VMS  
data

depends directly on the

UN/CEFACT  
FLUX Schema

Example Schema based



Example Artefact based

	characteristics	pro	con	examples
schema based	<p>often domain specific</p> <p>2 layers</p> <p>Single Maintenance Model</p>	<p>less freedom</p> <p>pragmatic</p> <p>easier to implement</p> <p>central led harmonization</p>	<p>every domain has its own standard</p>	<p>FLUX</p> <p>FishFrame</p> <p>DublinCore</p> <p>FIMES</p>
artefact based	<p>not domain specific</p> <p>3 layers</p> <p>Delegated Maintenance Model</p>	<p>more freedom</p> <p>'organic' harmonization</p>	<p>higher complexity</p>	<p>DDI</p> <p>SDMX</p>

## How is a codelist represented?

schema  
based

In the schema, often as a enumeration.  
This is however not practical for long lists.

When it comes to long lists, often a pointer is used to the list itself, which is the same as the below mentioned technique.

artefact  
based

As a separate independent artefact

MDM Capacity for the management of

& Reference data

& Code lists

& Classifications

& Mappings

& Harmonization and its related workflow

# Master Data Management (MDM)

The **business** layer defines the metadata of the information (schema or artefact based). “What is in the truck?”

The **transport** layer defines how the data and metadata will be transmitted. “What are the truck and the road?”

*Note: terminology copied from FLUX*

# Business & Transport layer



implementation friendly  
developer friendly  
user friendly

# Requirements

(related to information standards)



every standard is welcome  
avoid overlap - objective  
neutral - evaluate - advice  
stimulate adoption  
normative

(I) CWP &  
Information Standards



**Prio1:** CWP investigates, evaluates and advises on models (GSBPM, etc.) and standards (FLUX, SDMX, etc.)

Make sure all the CWP concepts & definitions are consistently applied

**Prio2:** Time & budget permitting, the CWP can go standard specific (e.g. FLUX&SDMX Catch DSD)

## (II) CWP & Information Standards



Recommending participation of RFB agencies  
in FLUX and SDMX projects

Ensure coherence over models and standards  
like FLUX and SDMX.

Promoting the structures, concepts and  
definitions as valuable towards any data  
exchange and harmonization effort.

Recommending a MDM initiative for fisheries  
facilitate datasets harmonization

# CWP consideration and discussion

thank you

FAO, Erik van Ingen  
Swakopmund, Namibia, 25-27 February 2015