

Towards common metadata using GSIM and DDI 3.2

EDDI conference, Copenhagen
2-3 December 2015

Mogens Grosen Nielsen

mgn@dst.dk

Flemming Dannnevang

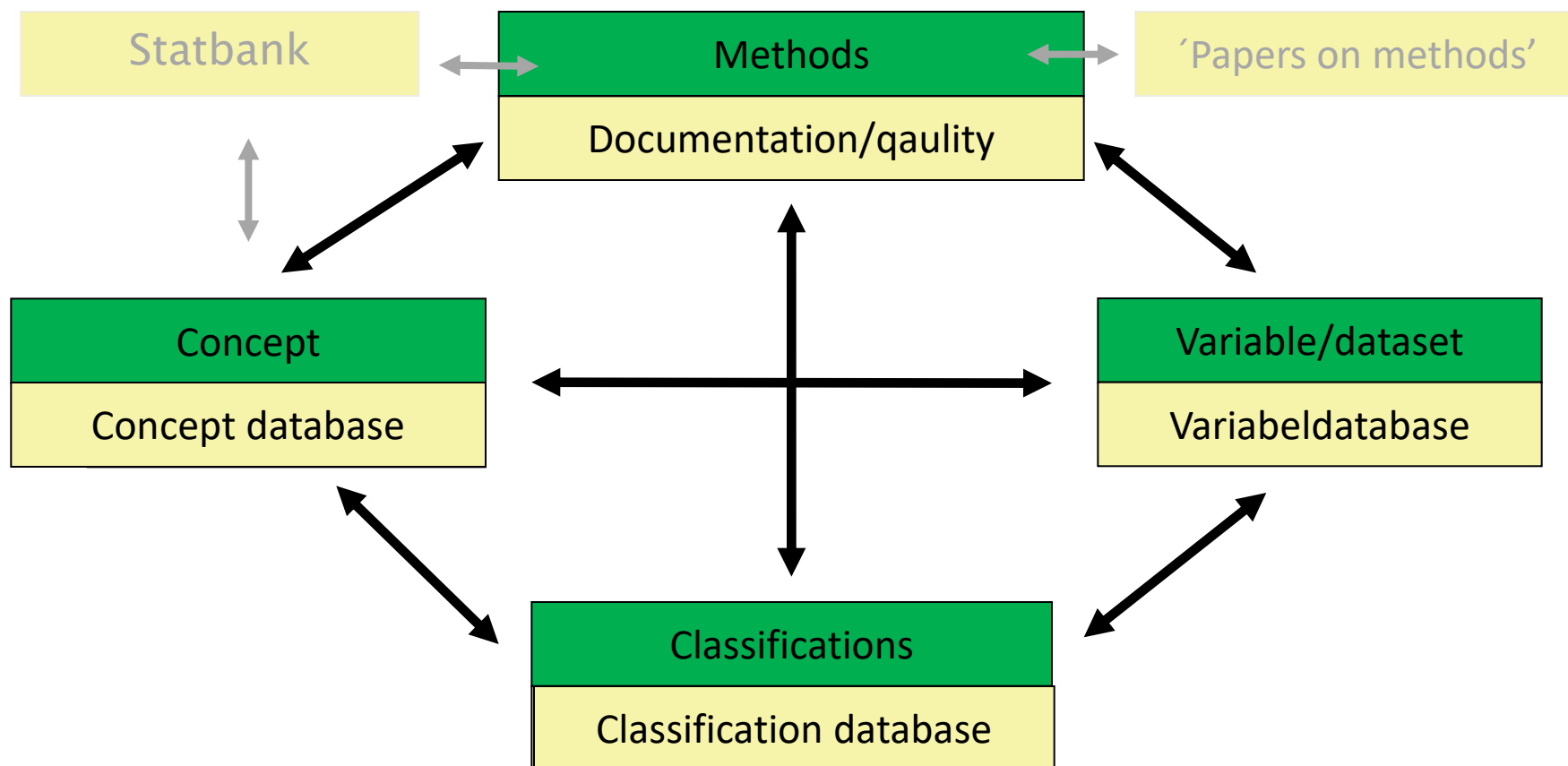
fda@dst.dk

Statistics Denmark

Agenda

1. Vision, strategy and challenges
2. Claims in the paper
3. Metadata users and business processes
4. Metadata terminology and models
5. Metadata portal
6. Conclusion

Vision: Integrated and reusable metadata



Strategy on quality and metadata

Fulfill user-needs, comply with quality requirements and increased efficiency

Standards: GSBPM, GSIM, DDI, SDMX

Principles a) Metadata integrated into GSBPM), b) reuse of metadata c) metadata used actively

Challenges

- Documentation attached to statistical product after it has been published
- User needs introduced too late
- Lack of awareness on common models and standards
- GSIM seems too complicated

Claims

Reusable metadata require improved understanding of

1. the role of metadata in relation to users
2. metadata in relation to production processes
3. metadata-terminology

Metadata, users and business processes

General Environment: Political/legal context, Technology/standards



Ressources:
staff, IT-
systems etc.

Respondents/
registers etc.

Management processes

Specify Needs

Design

Build

Collect

Process

Analyse

Disseminate

Evaluate

Support processes: Quality, **metadata**, methods
& IT

Users

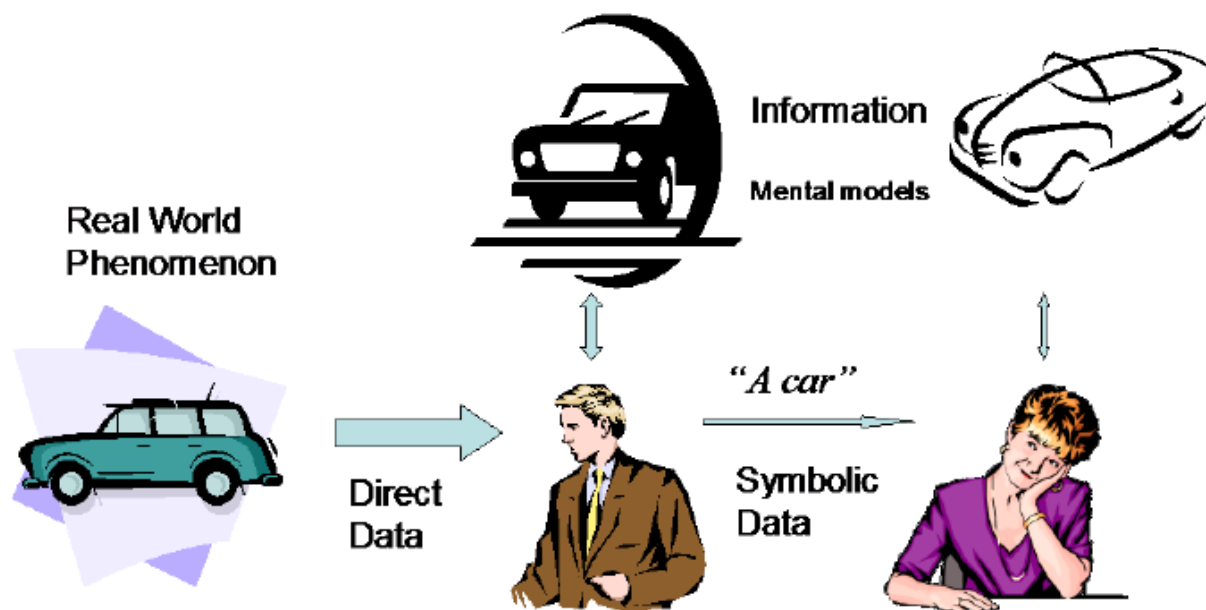
User needs / orders

Metadata and business processes

Phase		2. Design					
		2.1 Design outputs	2.2 Design variable descriptions	2.3 Design data collection methodology	2.4 Design frame and sample methodology	2.5 Design statistical processing methodology	2.6 Design production systems and workflow
Metadata object							
Statistical outputs	1	Define					
Concepts	2		Create/Reuse				
Data Source							
Variables (Observation)			Define				
Variables (Derived)	2		Define				
Statistical unit			Identify				
Classifications/Code lists			Create/Reuse				
Data collection mode				Define			
Questionnaire				Define			
Questions				Create/Reuse			
Flow logic				Create			
Response templates				Create/Reuse			
Interviewer instructions				Create			
Target population					Define		
Register(Universe)					Define		
Frame design					Define		
Sampling method					Define		
Processing methods						Define	
Operational processes							Define

Metadata terminology: Reality, data and information

Compatible frames of reference needed!



Sharing and reusing data and metadata

- *Sharing of data requires compatible frames of reference*
- *Compatible frames of references can be ensured via metadata*
- *Need for both metadata-terminology and domain-specific metadata*

Frames of references related metadata terminology and statistical metadata.

	Frames of reference inside NSI's	Frames of reference of external users
Terminology for statistical metadata	Complex and simplified metadata terminology (e.g. selected terms from GSIM)	Simplified metadata terminology (e.g. classification, variable, code-list)
Statistical metadata	Domain specific metadata (complex and simplified).	Domain specific metadata simplified

Terminology: simplified definition of statistical metadata (from SDMX)

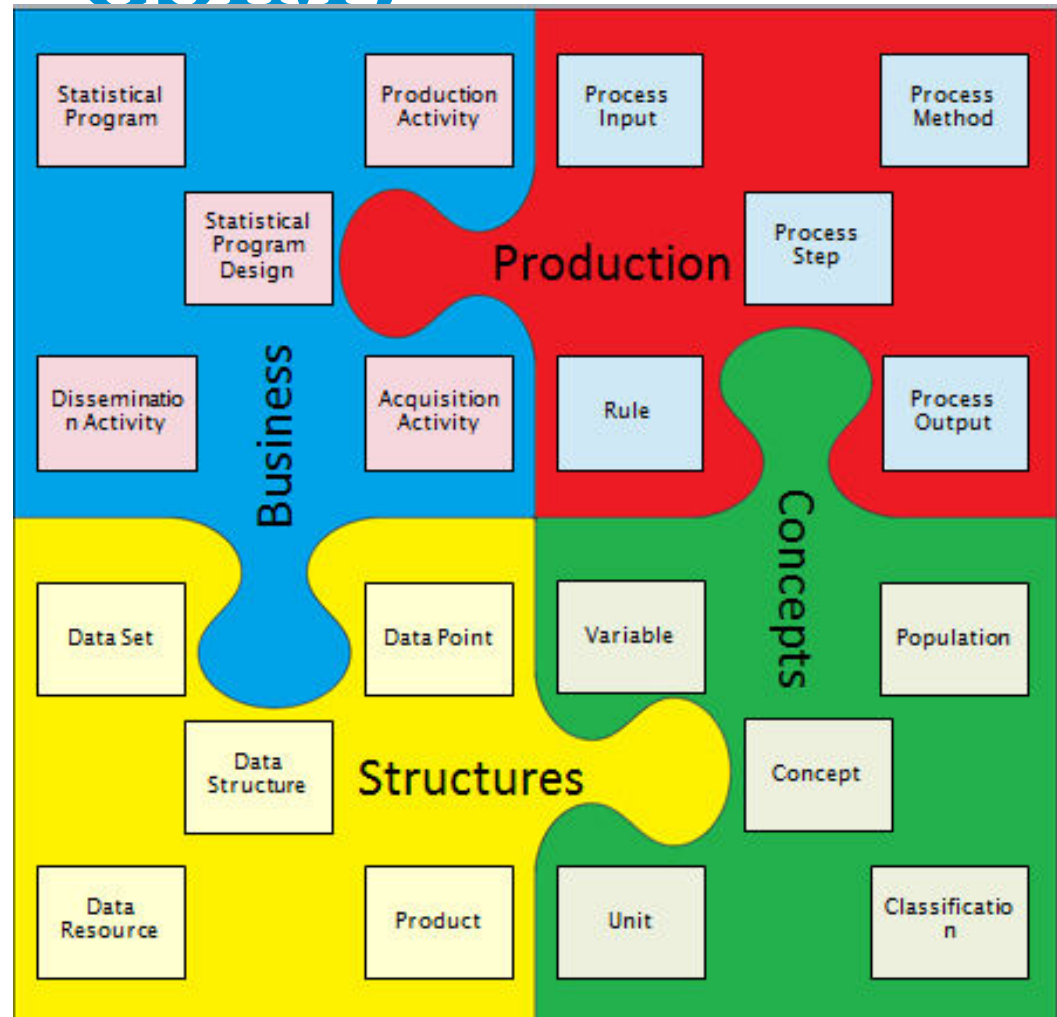
Reference metadata:

- Conceptual metadata (e.g. definition of income)
- Methodological and processing metadata (e.g. description of data processing)
- Quality metadata (e.g. Availability)

Structural metadata:

- Metadata act as identifiers and descriptors of the data (e.g. variables, code-lists, dataset)

Terminology: complex definition of statistical metadata (from GSIM)



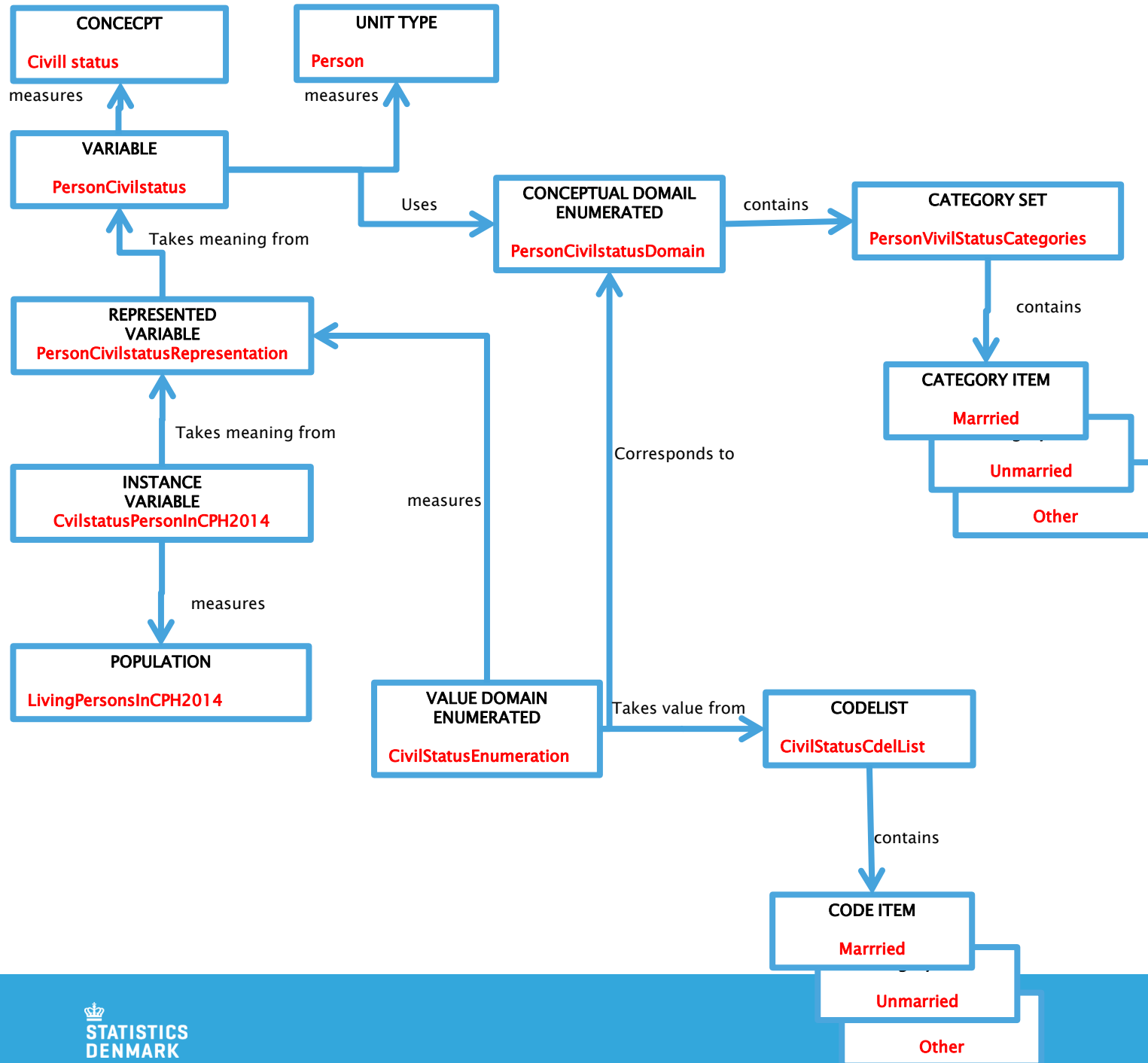
Frame of reference for complex metadata terminology

Model / level	Use in Stat DK
Conceptual	Selection of variable, concept etc from GSIM
Conceptual	GSIM compliant DDI model (3.2)
Logical	DDI model (3.2)
Physical	GSIM compliant DDI model implemented in Colectica

Metadata for dimensional data using GSIM terminology – an example

LivingPersonsInCPH2014_NoOf		Gender	
Civil Status		f	m
Married		14500	15000
Unmarried		20000	22100
Other		400	350





From GSIM to DDI 3.2 to Colectica – main activities and issues

To DDI 3.2

- High/Low level mapping
- Data structure conceptually very different

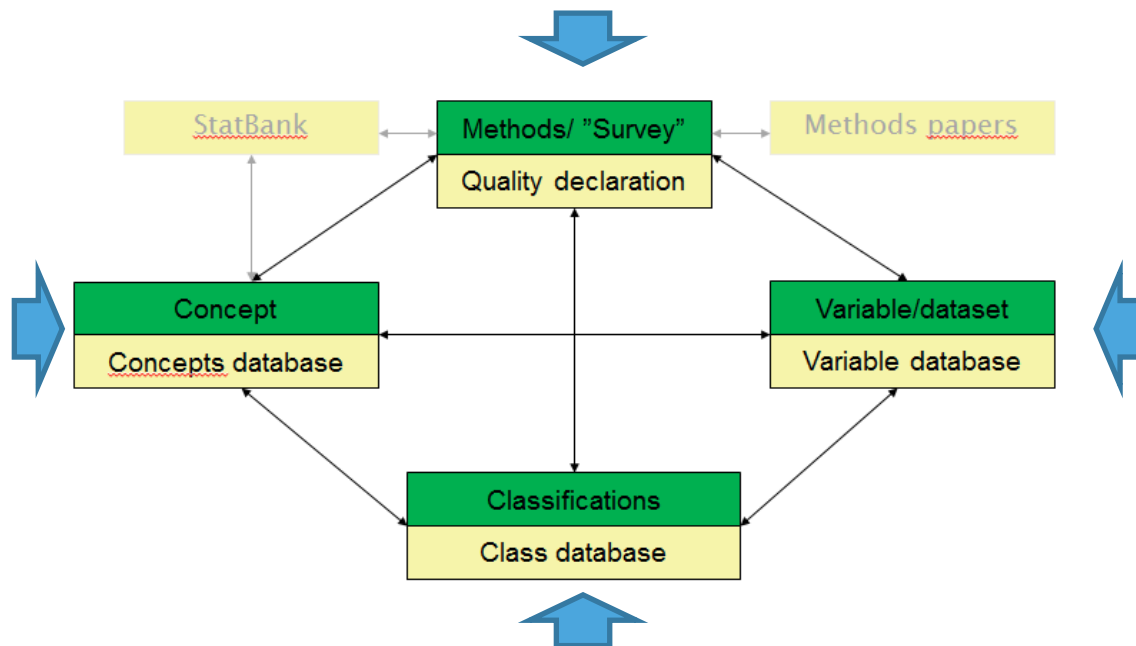
To Colectica

- Import, organising and linking elements
- Organising data so common metadata can be easily administrated and reused

Models implemented and used in metadata portal- simplified model

Lists of statistics

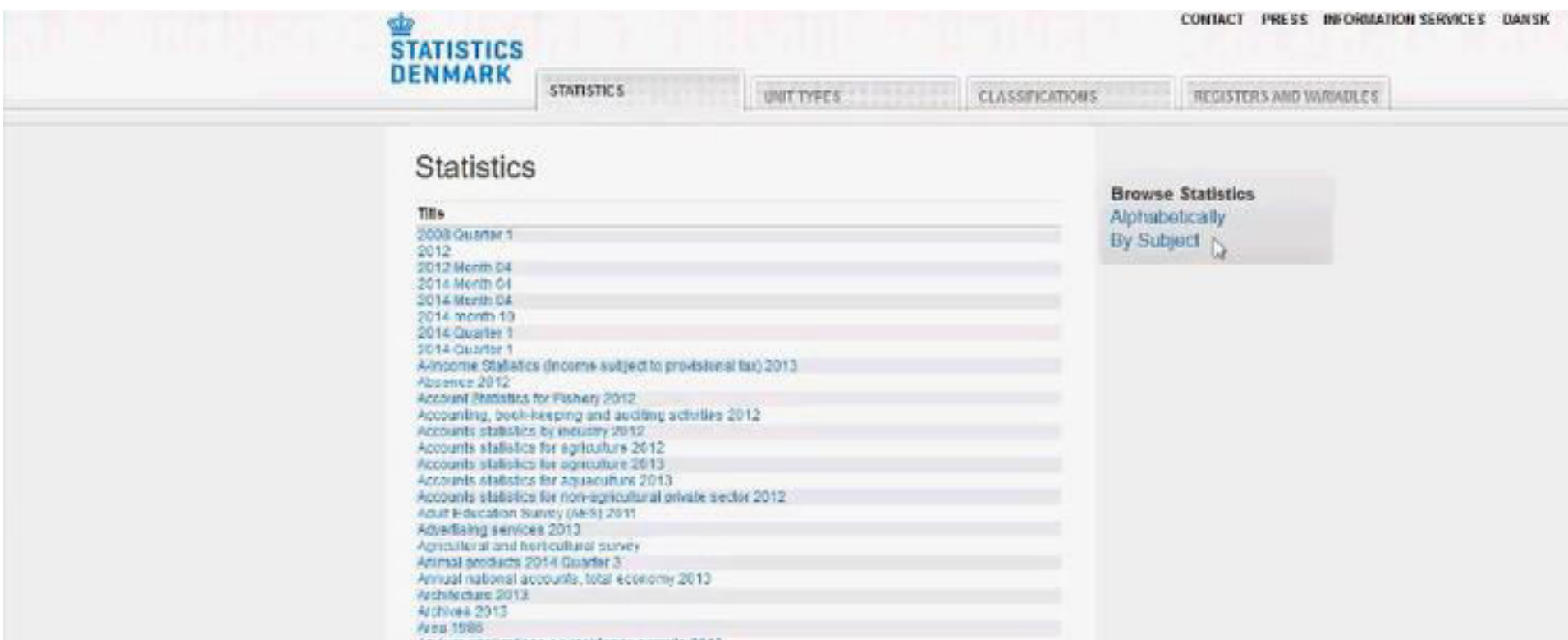
Lists of concepts
(unittypes and
pro-perties)



Lists of registers, data-
set and variables

Lists of classifications and
code lists.

Statistics - list



The screenshot shows the Statistics Denmark website. At the top, there is a navigation bar with the logo on the left and links for CONTACT, PRESS, INFORMATION SERVICES, and DANSK on the right. Below the navigation bar, there are four tabs: STATISTICS (selected), UNIT TYPES, CLASSIFICATIONS, and REGISTERS AND VARIABLES. The main content area is titled "Statistics" and contains a list of statistical items. On the right side of the main content area, there is a sidebar with the heading "Browse Statistics" and two links: "Alphabetically" and "By Subject".

STATISTICS DENMARK CONTACT PRESS INFORMATION SERVICES DANSK

STATISTICS UNIT TYPES CLASSIFICATIONS REGISTERS AND VARIABLES

Statistics

Titles

- 2008 Quarter 1
- 2012
- 2012 Month 04
- 2014 Month 01
- 2014 Month 04
- 2014 month 10
- 2014 Quarter 1
- 2014 Quarter 1
- Admission Statistics (income subject to provisional tax) 2013
- Absence 2012
- Account Statistics for Fishery 2012
- Accounting, book-keeping and auditing activities 2012
- Accounts statistics by industry 2012
- Accounts statistics for agriculture 2012
- Accounts statistics for agriculture 2013
- Accounts statistics for aquaculture 2013
- Accounts statistics for non-agricultural private sector 2012
- Adult Education Survey (AES) 2011
- Advertising services 2013
- Agricultural and horticultural survey
- Animal products 2014 Quarter 3
- Annual national accounts, total economy 2013
- Architecture 2013
- Archives 2013
- Area 1986
- Arrest statistics, as residence permits 2013

Browse Statistics

- Alphabetically
- By Subject

Statistics – “public expenditure”

The screenshot shows the Statistics Denmark website. At the top, there is a navigation bar with links for CONTACT, PRESS, INFORMATION SERVICES, and DANSK. Below this, there are tabs for STATISTICS, UNIT TYPES, CLASSIFICATIONS, and REGISTERS AND WRIDLES. The main content area is titled "Public expenditure and revenue on the environment 2012". On the left side, there is a sidebar with a note "Note: Navigation below is not yet functional" and a menu with links for CONTACT and INTRODUCTION. The main content area contains the following information:

Contact
Head of Section, Government Finances Flintull Annica Eriksson +45 3917 3573
aer@dst.dk

Contact organisation
Statistics Denmark

Contact organisation unit
Government Finances, Economic Division

Contact name
Flintull Annica Eriksson

Contact mail address
Sejregade 11, 2100 Copenhagen

Contact email address
aer@dst.dk

Contact phone number
+45 3917 3573

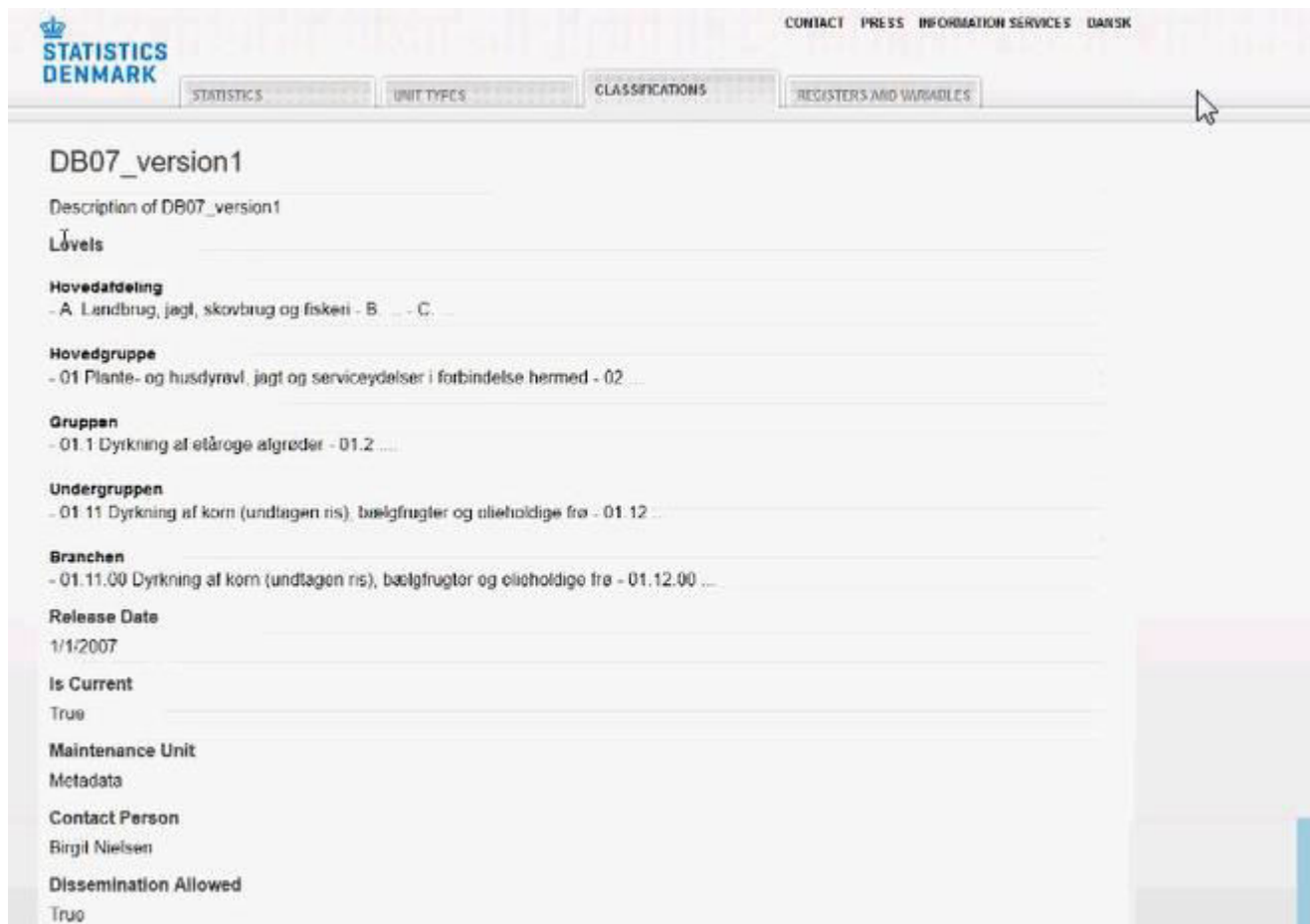
Contact fax number
+45 3917 3999

Quality management
Statistics Denmark is using the principle of Code of Practice (CoP) and the Quality Assurance Framework (QAF).

Classifications - list

STATISTICS DENMARK	
CONTACT PRESS INFORMATION SERVICES DANSK	
STATISTICS UNIT TYPES CLASSIFICATIONS REGISTERS AND VARIABLES	
Classifications	
Name	Description
Gender Codes	
OSDPM Quality Filter	
REGISTRERESKlassifikation	
AJO_BOPADL_KILDE_KODE	
AJO_BRANCHER07_KILDE_KODE	
AJO_BRANCHER07_OK	
AJO_ESR_KILDE_KODE	
AJO_FOKTIV_ANR_KODE	
AJO_FOEDDATO_KILDE_KODE	
AJO_FUNKTIONSKODE	
AJO_FUNKTIONSKODE_KILDE_KODE	
AJO_IMPUTERING_REGEL_KODE	
AJO_KOEN_KILDE_KODE	
AJO_LOENTIME_IMPUT_NIVEAU	
ANBRINGELSESTED	
ATP BIDRAGS SATS PA JOBBET	
AVS_AROV_STATUS	
BRANCHEKODER - DB07	
DB03_version1	
DB07	
DB07_version1	Description of DB07_version1
DB07_version2	
FUNKTIONSKODE	
Gender codes	
IMPUTERINGSTYPE	
INDKOMST ART KODE	
INDKOMST TYPE KODE	
KODE FOR EN PERSONS KØN	
Kommunekode	
Kommunekode2	
LOVAJODTAGERENS KOMMUNEKODE	
LOENTIMEKODE	

Classifications: "Activity codes – NACE"



STATISTICS DENMARK

CONTACT PRESS INFORMATION SERVICES DANSK

STATISTICS UNIT TYPES CLASSIFICATIONS REGISTERS AND VARIABLES

DB07_version1

Description of DB07_version1

Levels

Hovedafdeling
- A Landbrug, jagt, skovbrug og fiskeri - B - C -

Hovedgruppe
- 01 Plante- og husdyravl, jagt og serviceydelser i forbindelse hermed - 02 -

Gruppen
- 01.1 Dyrkning af etårige afgrøder - 01.2 -

Undergruppen
- 01.11 Dyrkning af korn (undtagen ris), bælgrugter og olieholdige frø - 01.12 -

Branchen
- 01.11.00 Dyrkning af korn (undtagen ris), bælgrugter og olieholdige frø - 01.12.00 -

Release Date
1/1/2007

Is Current
True

Maintenance Unit
Metadata

Contact Person
Birgit Nielsen

Dissemination Allowed
True

Registers and variables - list

STATISTICS DENMARK	
CONTACT	PRESS INFORMATION SERVICES DANSK
STATISTICS	UNIT TYPES CLASSIFICATIONS REGISTERS AND VARIABLES
Registers and Variables	
Title	
Templets	
Annual Income Statistics (income from which is drawn provisional tax)	
Absence	
Account Statistics for Fishery	
Accounting, book-keeping and auditing activities	
Accounts statistics by industry	
Accounts statistics for agriculture	
Accounts statistics for aquaculture	
Adult Education Survey (AES)	
Advertising services	
Agricultural and horticultural survey	
Animal products	
Annual national accounts, total economy	
Architecture	
Archives	
Area	
Asylum applications, eg residence permits	
Auction	
Bachelors and long-term higher education	
Balance of payments	
Balance of payments and the international investment position	
Balance of payments and the international investment position	
Bankruptcy	
Benefits during sickness or in connection with childbirth	
Berlin	
Book production	
Budgets of general government	
Business and management consultancy activities	
Business Cycle Tracer	
Business demography	
Business Enterprise Research and Development (BERD)	
Business services in the sector for market research and public opinion polling	
Business services in the sector for technical testing and analysis	
Car register and publications	
Cash Benefits	
Cattle	
Census of buildings, 1st January	
Census of housing	

Registers and variables – ”Person gender”

The screenshot shows the Statistics Denmark website interface. At the top, there is a navigation bar with the logo on the left and links for CONTACT, PRESS, INFORMATION SERVICES, and DANSK on the right. Below the navigation bar is a horizontal menu with four tabs: STATISTICS, UNIT TYPES, CLASSIFICATIONS, and REGISTERS AND VARIABLES. The REGISTERS AND VARIABLES tab is selected. The main content area displays the details for the 'Person gender instance variable'. It includes sections for Label, Representation Type, Code List, Dataset Name, and Universe. At the bottom of the page, there is a footer with contact information for Statistics Denmark and a note about the proof-of-concept metadata portal.

STATISTICS DENMARK

CONTACT PRESS INFORMATION SERVICES DANSK

STATISTICS UNIT TYPES CLASSIFICATIONS REGISTERS AND VARIABLES

Person gender instance variable

Label
Person gender instance variable

Representation Type
Code

Code List
Gender Codes

Dataset Name
Population 2014 Data

Universe
Population population 2014 in Denmark

Statistics Denmark
Søstergade 11
DK-2100 Copenhagen
dat@stat.dk
+45 39 17 29 17

This proof-of-concept metadata portal is
powered by Collection

Conclusions

Common and reusable metadata require

- Coordinated work on handling metadata in relation to user needs and business processes and efforts in developing precise metadata terminology
- Development of metadata terminology and applications require careful modelling going from conceptual level to physical level.
- Compatible frames of reference is needed to ensure common understanding