WP 9
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

UNECE Workshop on the Common Metadata Framework
(Vienna, Austria, 4-6 July 2007)

SUMMARY RESULTS OF A QUESTIONNAIRE ON PART C OF THE COMMON METADATA FRAMEWORK (CMF)

UNECE Secretariat[1]

1. *An electronic questionnaire was circulated to the 94 METIS mailing list members in June 2007. A copy of the survey is provided at Appendix A. Out of 57 organisations represented on this email list, 19 have answered the questionnaire (33%).*

2. *The results indicate similarities in the preferred process model for a statistical survey life cycle, and suggest that survey planning and design is the most problematic phase in terms of capturing metadata. There is not a single metadata standard that the majority of the respondents are conforming to. Resistance from subject matter statisticians, employing and retaining people with the right IT skills, and implementing new metadata standards appear to be the greatest challenges for effective implementation of metadata systems and frameworks.*

3. In a survey conducted by UNECE in early June 2007 among members of the METIS e-mail list, the following 19 organizations provided a response:

- Australian Bureau of Statistics
- Statistics Austria
- Statistics Canada
- Croatian Bureau of Statistics
- Statistical Service of Cyprus (CYSTAT)
- Czech Statistical Office
- Statistics Denmark
- Statistics Estonia
- State Statistical Office – The former Yugoslav Republic of Macedonia
- Statistics New Zealand
- Statistics Norway
- Central Statistical Office of Poland
- Statistical Office of the Slovak Republic
- Statistical Office of the Republic of Slovenia
- Statistics South Africa
- Instituto Nacional de Estadistica (INE)(Spain)
- Statistics Sweden
- Swiss Federal Statistical Office
- US Bureau of Labor Statistics

[1] Prepared by Jessica Gardner, jessica.gardner@unece.org and Vanda Czifra, vanda.czifra@unece.org

Views on Proposed Structure of Part C

4. 95% of respondents have both consulted the CMF Part C website and agree with the proposal paper for Part C (http://www.unece.org/stats/cmf/PartC.html). No respondent has suggested any major structural changes to Part C.

Preferred process model for the statistical survey life-cycle

5. The majority of respondents described their preferred process model for the statistical survey life cycle, or plan to do so at the METIS workshop in July 2007. The models were similar to that proposed in Part C (C1 to C7), however, evaluation (C8), was only explicitly mentioned in about a quarter of them.

Need → Design → Build → Collect → Process → Analyse → Disseminate

Figure 1: Example process model for the statistical survey life-cycle

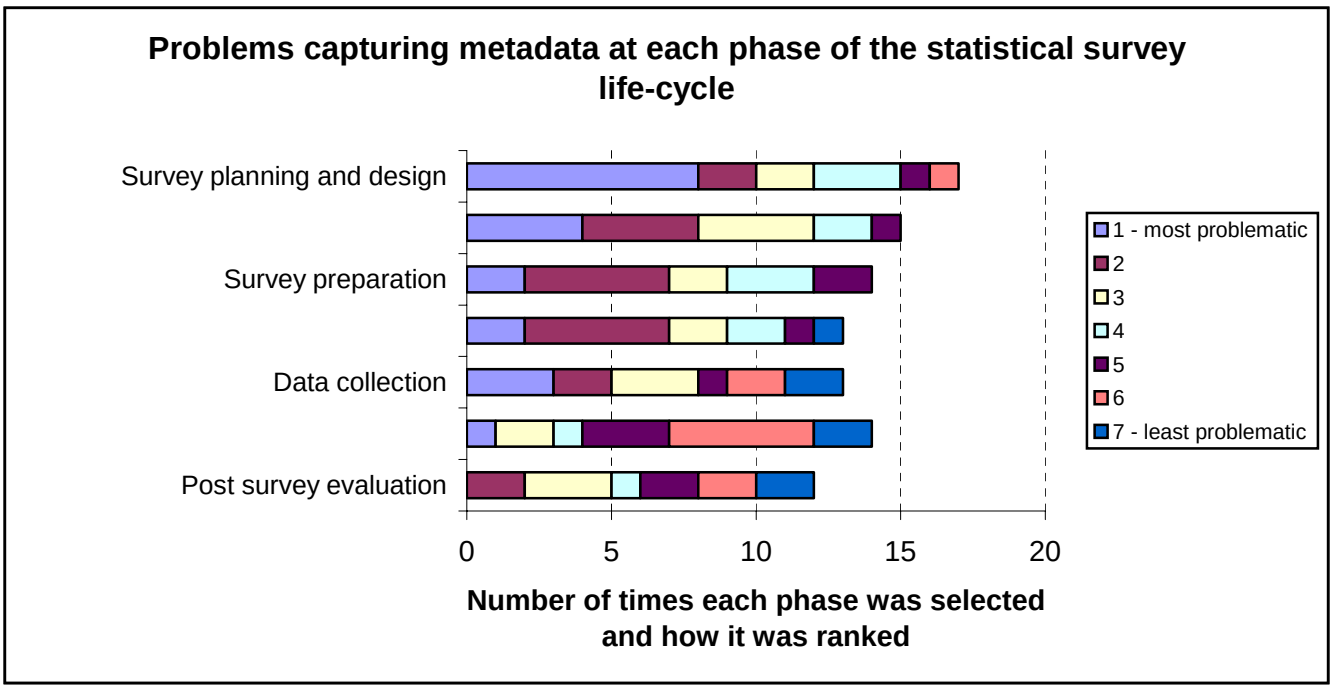
6. Respondents emphasised the importance of an end-to-end (E2E) approach to metadata management throughout the statistical life-cycle. There was also a suggestion to map 'Inputs, Outputs, Guides and Enablers' between sub-processes, in order to help to understand the reusability of metadata throughout the cycle.

7. The results indicate that it is useful to have a generic model, as proposed for Part C, but organizations will adapt it to suit their needs, modifying terminology, and merging, splitting, or adding phases. The models described by survey respondents are provided at Appendix B.

Problems capturing metadata

8. *Survey planning and design* was revealed as the most problematic point of the statistical cycle in terms of capturing metadata (see Table 1 below). The second most problematic point is *Derivation, estimation, aggregation, analysis,* with 74% of respondents selecting it and ranking it in their top five problem areas.

9. *Survey preparation* and *Input processing* phases were next, with around 70% of respondents selected them, and half of those ranking them in their top two problem areas.



10. *Post survey evaluation* was identified as the least problematic phase, since only about half of the respondents selected it and it is ranked fairly low. Approximately 70% of the respondents selected the remaining two phases and the ranking indicates that the degree of problems with *Data collection* varies to some extent, and *Dissemination* is not seen as highly problematic.

Table 1: Results from question 7: "Which points in the statistical cycle present the most problems in terms of capturing metadata? Please select those that apply, and then rank them in order from most to least problematic."

	Selected by % of respondents	**Ranked from most problematic (1) to least problematic (7)** (by number of respondents)						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Survey Planning and design	**95**	**9**	2	2	3	1	1	-
Derivation, estimation, aggregation, analysis	74	3	**4**	**4**	2	1	-	-
Survey preparation	74	2	**5**	2	3	2	-	-
Input processing	68	2	**5**	2	2	1	-	1
Data collection	68	**3**	2	**3**	-	1	2	2
Dissemination	74	1	-	2	1	3	**5**	2
Post survey evaluation	63	-	2	**3**	1	2	2	2

*Highest values in **bold***

Use of standards

11. When it comes to standards, there is not a single one that a clear majority of the respondents are conforming to (see Table 2 below). However, the four most implemented standards, either partially or fully conformed to, are *Dublin Core*, *DDI*, *SDMX*, and *ISO 11179*.

12. The least used standards are *RDF or OWL*, and the *Common Warehouse Model (CWM)*. *XBRL* and *ISO 19115 (GIS Metadata Standard)* are not widely used either, but many are considering them, as they are *ISO 11179, DDI and SDMX*. Among the other standards that the respondents mentioned using, the *Neuchatel terminology* and the *Metanet model* are dominant.

Table 2: Results from question 8: Please indicate your use of the following standards (respondents could select as many options as applied)

	No current use	Considering / following ideas	Implementing, but not conforming	Conforming	Able to output data in this format
	Number of respondents who selected each combination				
Data Documentation Initiative (DDI)	4	8	-	4	4
SDMX (ISO 17369)	5	7	3	2	3
ISO/IEC 11179 Metadata Registry Standard	5	8	2	2	3
Common Warehouse Model (CWM)	12	3	-	1	1
Dublin Core	10	1	2	3	1
ISO 19115 (GIS Metadata Standard)	9	6	2	-	-
XBRL	11	7	1	1	-
EDI or UN/EDIFACT	9	1	1	1	4
RDF or OWL	13	2	1	-	-
Metanet*	-	-	2	-	-
Neutchatel*	-	1	-	2	2

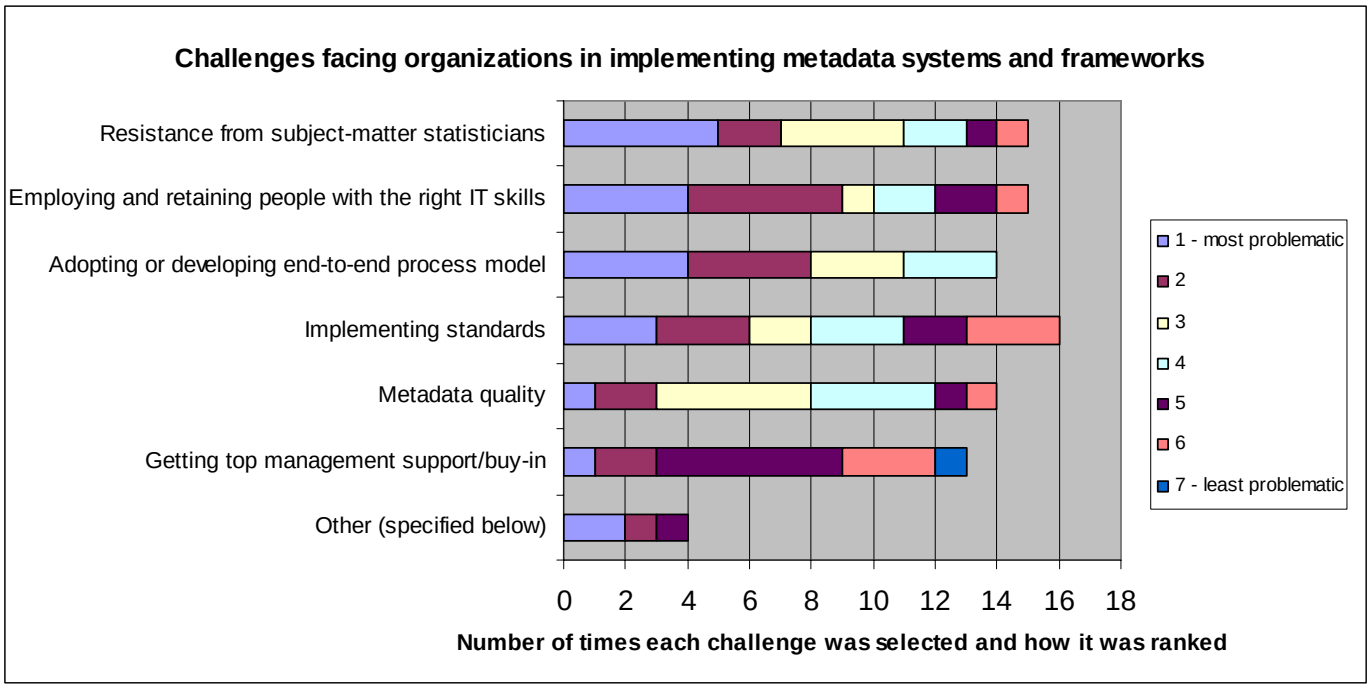
FGDC*	-	-	-	1	-
METS*	-	-	1	-	-

**'Other' standards specified by respondents*

Challenges faced in implementing metadata systems and frameworks

13. Respondents were given a list six challenges to consider. They were asked to select those relevant to their organization, and then rank them in order from most to least challenging. All six challenges were selected by at least 68% of respondents, and a further four challenges were specified under 'other'.

14. *Implementing standards* was the most selected challenge (84% of respondents), but its ranking varies considerably (see graph and Table 3 below). It seems that *Resistance from subject matter statisticians* is the most challenging issues facing organizations, followed by *Employing and retaining people with the right IT skills* and *Adopting or developing an E2E process model.*



15. *Getting top-management support/buy-in* was considered the least challenging issue and *Metadata quality* seems also to represent moderate challenges.

16. Other challenges specified by respondents include:

- identifying standard metadata requirements in environments which have previously worked in isolation;
- difficulties embedding the E2E process within actual workflows;
- insufficient integration of standard metadata system with data production systems; and
- lack of resources.

Table 3: Results from question 9: "What challenges does your organization face in implementing metadata systems and frameworks? Please select as many options as apply, and then rank them in order from *most* to *least* challenging."

	Selected by % of respondents	**Ranked from most challenging (1) to least challenging (7)** (by number of respondents)						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Implementing standards	**84**	**3**	**3**	2	**3**	2	**3**	-
Resistance from subject-matter statisticians	79	**5**	2	4	2	1	1	-
Employing and retaining people with the right IT skills	79	4	**5**	1	2	2	1	-
Adopting or developing end-to-end process model	74	**4**	**4**	3	3	-	-	-
Metadata quality	74	1	2	**5**	4	1	1	-
Getting top management support/buy-in	68	1	2	-	-	**6**	3	1
Other (specified below)	21	2	1			1		
Embedding E2E process into workflows*		1	-	-	-	-	-	-
Lack of resources*		1	-	-	-	-	-	-
Identifying common metadata standards*		-	-	-	-	1	-	-
Insufficient integration of metadata system with data production systems*		-	1	-	-	-	-	-

**'Other' challenges specified by respondents*

Other issues

17. Finally, respondents highlighted the following other issues:

- To determine which standards will be most beneficial often requires technical solutions to be implemented to fully understand the scope of what we currently have. We are finding it will need to be a cyclical process of analysing, implementing and re-assessing to continually improve upon the gains from each step.
- It would be interesting to see a metadata system that could output equally well in all the standards.
- In Part C, it is important to get a many case studies from different countries as possible. Once the documentation is put together and reviewed, it would useful to determine the degree of convergence among the metadata models / approaches and see if the METIS Steering Committee could recommend a common approach or best practice for metadata to support the survey life cycle. This could be reported at the METIS 2008 meeting.
- The use of the word survey in the statistical 'survey' life-cycle process is misleading. It leads to the exclusion of processes based on administrative data; where no survey has been carried out. The term "statistical task" could be more appropriate for the statistical key process. This issue needs some discussion.
- Appreciation of the importance of the work of METIS.
- There is a lot work left to do in the field of describing the statistical production process and understanding how metadata (including process metadata) is used and produced at the different stages.
- The business case of investing in metadata systems in the process of data collection, input processing, and derivation and estimation is not defined, because it is internal process. The business case of metadata systems is much clearer for dissemination process.

Acknowledgements

18. The UNECE Secretariat would like to thank the respondents and to acknowledge the following people for their contribution to the development of the questionnaire: Daniel Gillman (US Bureau of Labor Statistics), Gary Dunnet and Matjaz Jug (Statistics New Zealand), Alice Born (Statistics Canada), Alistair Hamilton (Australian Bureau of Statistics), and Joza Klep (Statistics Slovenia).

Appendix A

METIS Questionnaire - Common Metadata Framework, Part C

Editors of the Common Metadata Framework (CMF) have developed this brief questionnaire to gather information relating to Part C of the framework.

Please take five minutes to complete the following 10 questions.

Please email the completed questionnaire to jessica.gardner@unece.org or fax to +41 22 917 0040.

1) Name (optional):

2) Organization (mandatory):

3) Have you consulted the Common Metadata Framework (CMF) Part C website (www.unece.org/stats/cmf/PartC.html) ?

- ☐ **No** ---> question 6
- ☐ **Yes** ---> question 4

4) Do you agree with the proposal paper for Part C (which appears on the CMF website)?

- ☐ **No**
- ☐ **Yes**

5) Please list the major structural changes you want to see in CMF Part C.

6) What is your preferred process model for the statistical survey life-cycle? Please describe it briefly and name all the phases.

7) Which points in the statistical cycle present the most problems in terms of capturing metadata?

Please select those that apply by numbering them in order from most (number 1) to least problematic.

Leave blank those that are not applicable.

- **Survey planning and design**
- **Survey preparation**
- **Data collection**
- **Input processing**
- **Derivation, estimation, aggregation, analysis**
- **Dissemination**
- **Post survey evaluation**

8) Please indicate your use of the following standards.

	No current use	Considering / following ideas	Implementing, but not conforming	Conforming	Able to output data in this format
Data Documentation Initiative (DDI)					
Statistical Data and Metadata Exchange (SDMX ISO 17963)					
ISO/IEC 11179 Metadata Registry Standard					
Common Warehouse Metamodel (CWM)					
Dublin Core					
ISO 19115 (GIS Metadata Standard)					
XBRL					
EDI or UN/EDIFACT					
RDF or OWL					
Other (Please specify) :					

9) What challenges does your organization face in implementing metadata systems and frameworks?

Please select those that apply by numbering them in order from most (number 1) to least problematic. Leave blank any that are not applicable.

- ☐ **Getting top management support/buy-in**
- ☐ **Resistance from subject-matter statisticians**
- ☐ **Implementing standards**
- ☐ **Adopting or developing an end-to-end process model**
- ☐ **Employing and retaining people with the right IT skills**
- ☐ **Metadata quality**
- ☐ **Other (please specify:)**

Finally, are there any other issues you would like to mention?

Thank you for taking the time to complete this questionnaire. Your responses will help guide further development of the Common Metadata Framework (CMF).

The collated results of this survey will be sent to METIS mailing list members in July 2007.

Appendix B

Descriptions of preferred process models for the statistical survey life-cycle

Model proposed in Part C (July 2006)	C1. survey planning and design; C2. survey preparation; C3. data collection; C4. input processing; C5. derivation, estimation, aggregation; C6. analysis; C7. dissemination; C8. post survey evaluation.
Survey responses	
1.	1. Target, customer demand 2. Frame and sample 3. Data collection 4. Data preparation 5. Statistical computation 6. Dissemination and communication 7. evaluation/customer satisfaction
2.	1. Study planning and design (term 'study' refers to both surveys and other statistical activities) 2. Study preparation (resources, methodology) 3. Data collection 4. Data processing (data input, data validation, aggregation, tabulation) 5. Dissemination (production of tables and publications)
3.	1. Evaluation of the requirements on the statistical information (the positive decision initiates the key process) 2. Subject-matter specification of a statistical task (under term "statistical task" we understand all activities of the key process; survey, if exists, is the integral part of the task) 3. Technical specification and software development 4. Preparation of a processing 5. Primary processing (data collection, validation and missing data imputations) 6. Aggregation and estimations 7. Dissemination 8. Evaluation of the statistical task
4.	1. Detection of a user' need. New interest areas transmitted by principal agents (for instance Eurostat, government, university, etc.) 2. Planning and design of the project 3. Developing of the necessary tools (programs, questionnaires, etc.) 4. Data collection 5. debugging and analysis of data, including imput processing 6. dissemination 7. Analysis of the user' satisfaction 8. Analysis of the product and process' quality
5.	1. Survey planning and design 2. Survey preparation 3. Data collection 4. Input processing

	5. Derivation,estimation,aggregation,analysis 6. Dissemination
6.	1. Conceptualization Design (for the sample, estimation, questions, questionnaire, data, and databases) 2. Collection 3. Processing 4. Analysis 5. Dissemination (Analysis precedes dissemination because we release tables and time series data, not the microdata, and the analysis is done on the microdata.)
7.	1. Survey planning 2. Development and design 3. Survey preparation 4. Data collection 5. Processing 6. Analysis 7. Presentation and dissemination
8.	1. Initialisation (covers parts C1 and C2) 2. Data collection (same as C3) 3. Transformation (C4 and C5) 4. Analysis (C6) 4. Information enrichment (not present in framework, this is the phase in which articles that are designed to present statistical results (paper, cd-rom or online) are written.) 5. Publication (C7 and to some extent C8, though we do not perform post survey evaluations on a regular basis)
9.	1. Design & tune 2. acquire data 3. process inputs 4. transform inputs into statistics (estimation, aggregation, output editing etc), analysis & explanation 5. assemble & disseminate 6. decision support 7. evaluate & tune