


Introductory Remarks on Metadata – Paradata – Data and Quality in Comparative Survey Research



Lars Lyberg
Stockholm University
and
Peter Ph. Mohler
Mannheim University
CSDI 2011 Workshop
London March 2011

Part 1

Meta-Data



Peter Ph. Mohler



Overview

- Definitions
 - Data, meta-, para-
 - Meta-Data
- DDI Alliance - a mega story
- DDI and Comparative Survey Methodology
 - Towards a Meta-Data UFO
 - Towards quality indicator driven comparative survey documentation



Data: Definitions from Dictionary

- facts and **statistics** collected together for reference or analysis: *there is very little data available*
- the **quantities**, characters, or symbols on which operations are performed by a computer...
- philosophy: **things** known or assumed as **facts**, making the basis of reasoning or calculation.

Oxford Online Dictionary 2011



Two Types of Data in Survey Research – A First Approach

- Data

- Extensive Data

- Indicate the amount or extension of a fact
- **Numerals**
- Example: frequency of a response in a population

- Qualitative Data

- Indicate qualitative facts
- **Non-numerals**
- Example: Job descriptions (ISCO items)

Note: questions often ask for a qualitative fact (education) that is then coded on the spot into a numeral



Meta- & Para- Dictionary Entries

- “Meta”
 - “denoting something of higher or second-order kind”
- “Para”
 - “besides, adjacent to”
 - “something that protects or wards off”

Definitions from Oxford Online Dictionary c2011

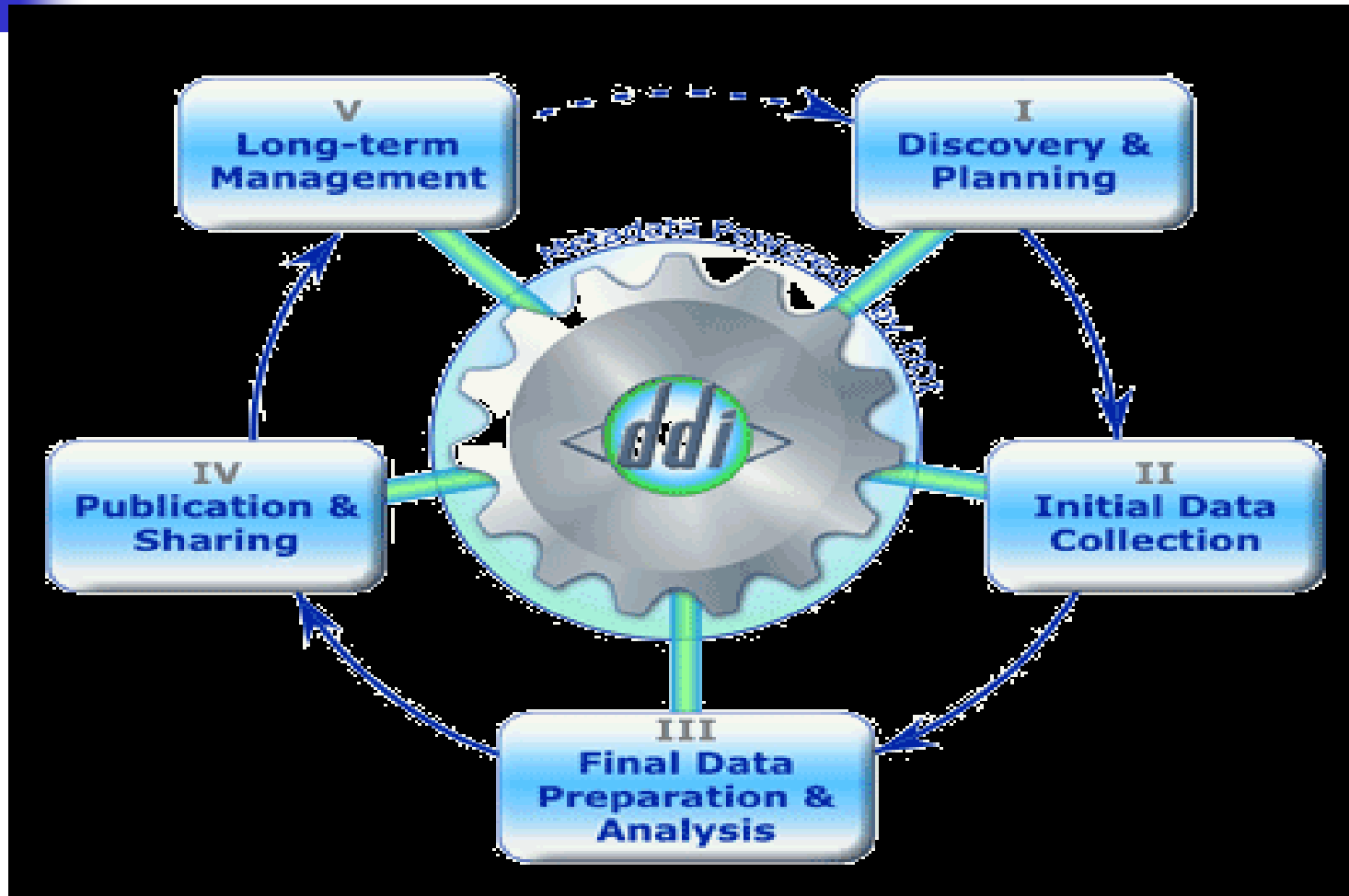


Metadata

- “a set of **data** that describes or gives information about other **data**”
- Extensive & Qualitative Metadata
 - Example extensive: response rate
 - Example qualitative: name of PI, description of quality procedures

DDI – Data Documentation Alliance

www.ddialliance.org





The Mega-Story of DDI

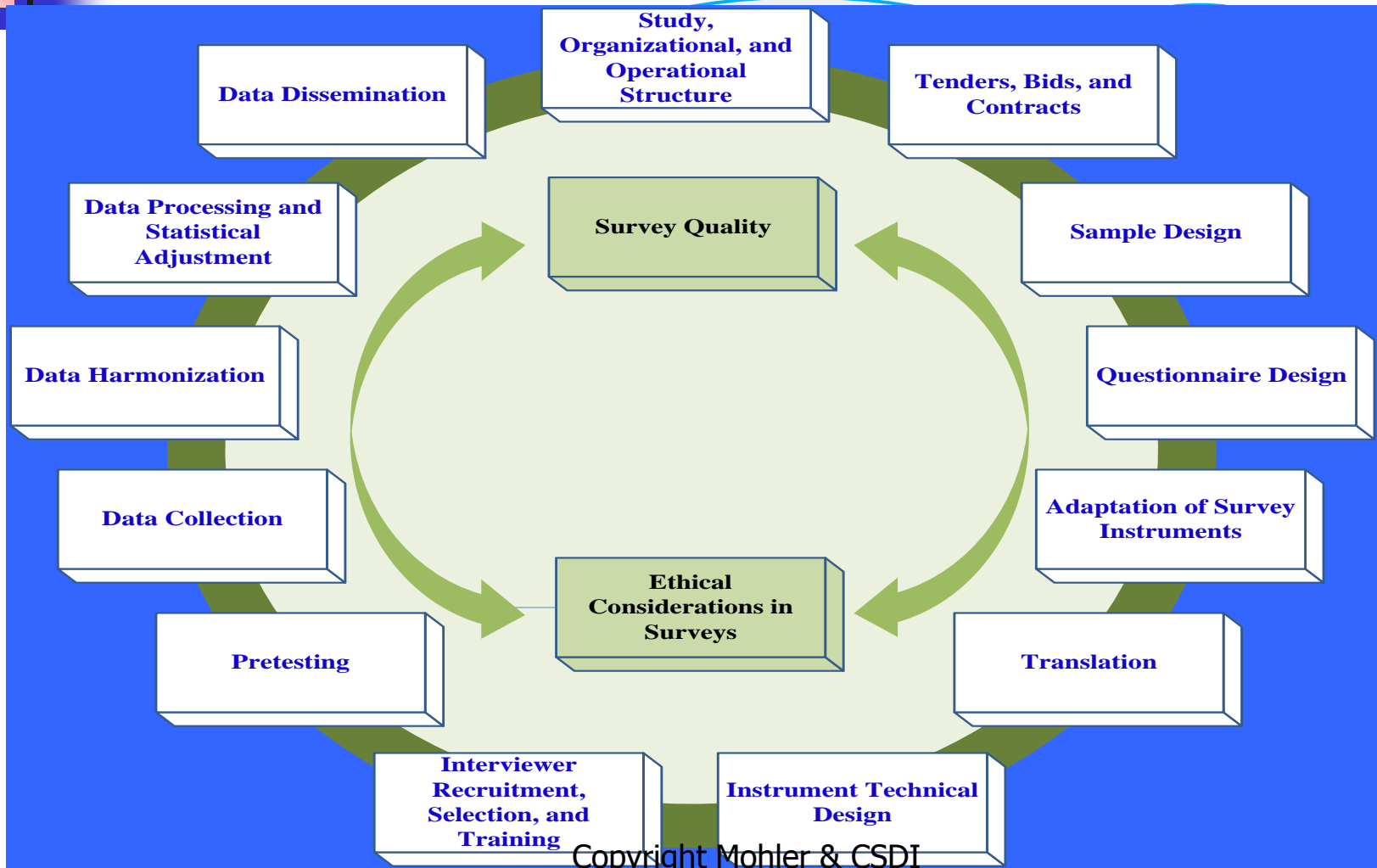
- 1999 Miller, Kenneth, and Pasqualino Titto Assini *"The XML Files: Developing Data UFO"*
- 2001 Ryssevik, Jostein, and Simon Musgrave *"From Data Graveyards to Knowledge Greenhouses"*
- 2006 Thomas, Wendy *"Locating the Geographic Center of DDI 3.0 - Part of DDI for the Next Decade: Toward Version 3.0, to Presentations from the 2nd Annual European DDI Users Group Meeting, Urecht, Netherlands, December 2010"*.
- Gregory, Arofan and Pascal Heus. "Metadata." **German Data Forum (RatSWD)** Working Paper no. 57/2009, March 2009.



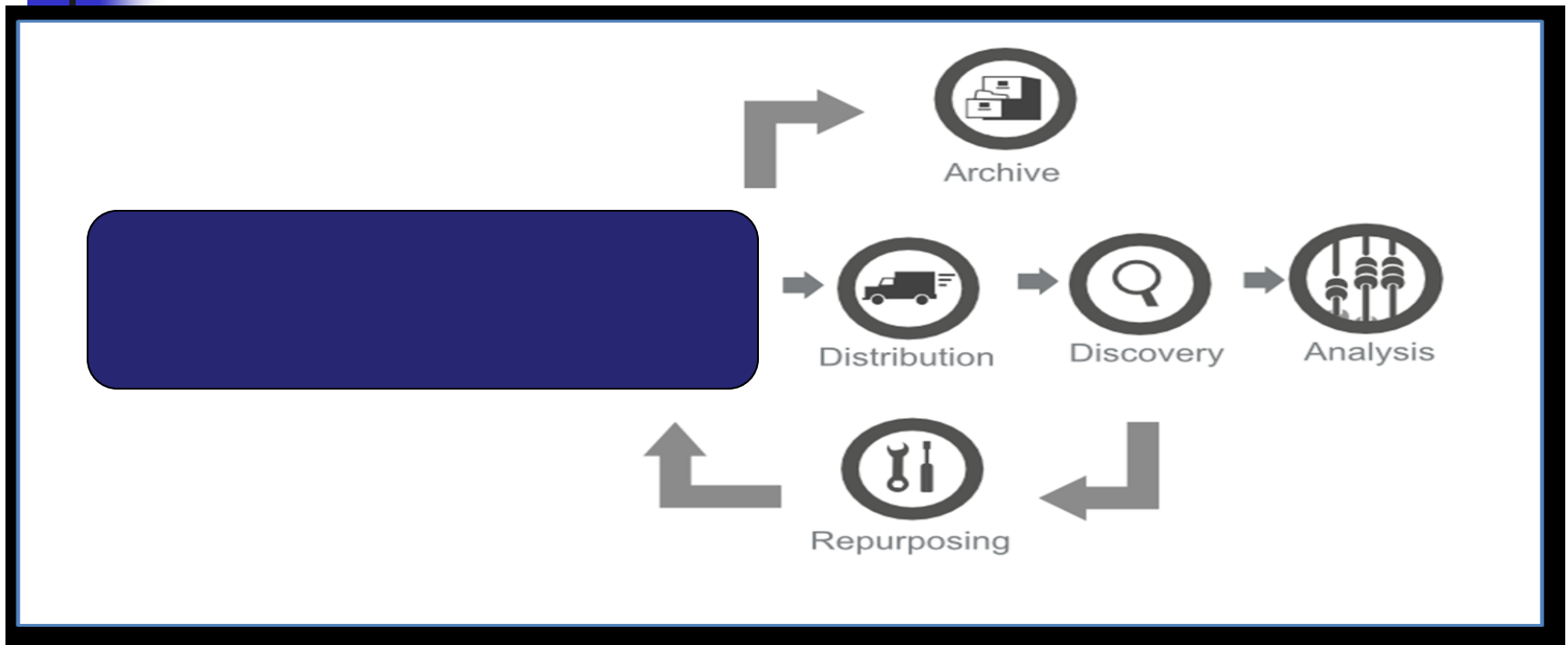
New in Comparative Survey Methodology

- Integration of production process quality into the survey life cycle paradigm
- Paradigm change from life cycle to complex production process systems
- Extensive guidelines since 2010
 - <http://ccsg.isr.umich.edu/>

Turning the Life Cycle into a Survey Production Process System

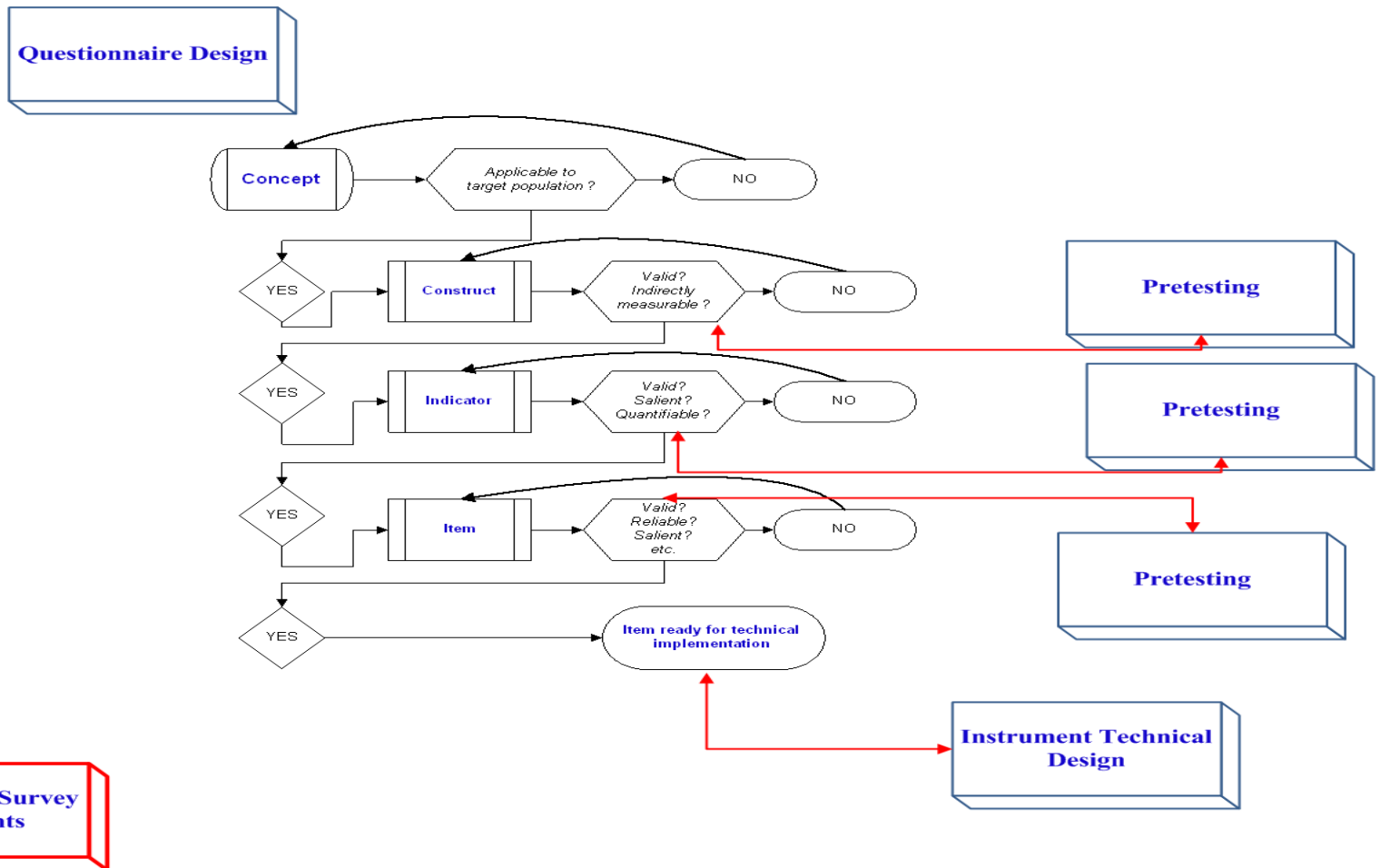


Simplistic View on Survey Production Process



Iverson, Jeremy "Metadata Driven Survey Design"
Presentations from the 2nd Annual European DDI Users Group Meeting,
Utrecht, Netherlands, December 2010. (*Collectia Sales Talk, PM*)

Example Monocultural Integrated Process Quality Monitoring and Control





Meta-Data Comparative UFO

- Acknowledgment of survey process complexity
- Concentration on process stages incl. their quality
- Simultaneous 50+ surveys documentation (and control)



Quality Indicator Documentation

- Quality – see Lars Lybergs presentation
- Just one example of information reduction
- As always is the ESS plagued by “the enemy of the good is the better”...

E23 (ACQ30BS): Category 55 ("It depends") was not fielded in Croatia.

ESS4 - 2008 Summary and deviations

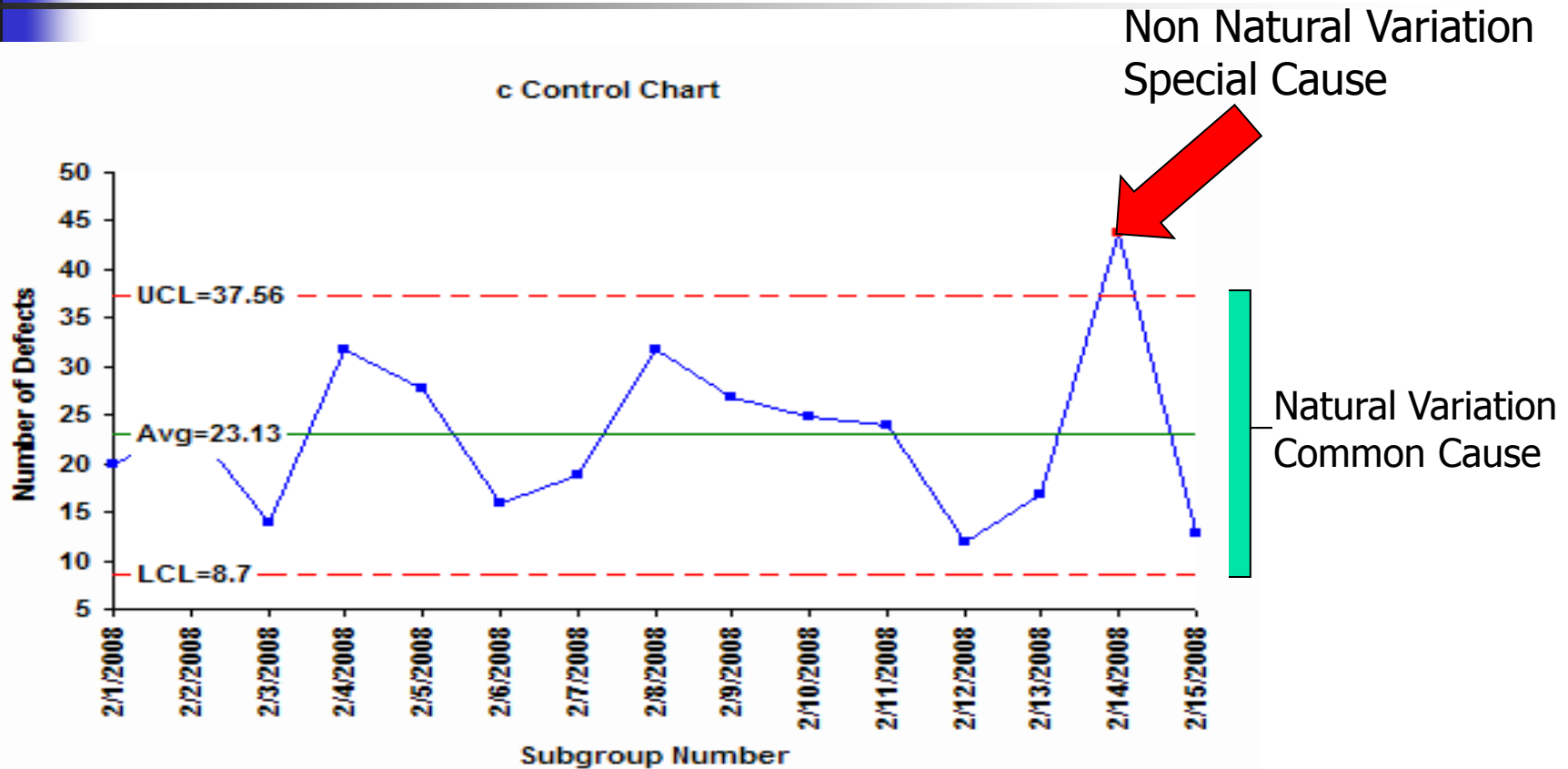
Target response rate: 70%.

Target fieldwork period: 01.09.08 - 31.12.08.

The underlying national questionnaire items have changed since ESS3-2008.

COUNTRY	FIELDWORK PERIOD	N	RESP. RATE*	DEVIATIONS IN DATA (CLICK ON TEXT BELOW FOR DETAILS)
Belgium	13.11.08-20.03.09	1 760	58.9	F6 (EDULVLA), F6 (EISCED), F36 (EDULVLP), F49 (EDULVLF), F55 (EDULVLM)
Bulgaria	06.03.09-31.05.09	2 230	75.0	F32 (HINCTNTA), F6 (EDULVLA), F36 (EDULVLP), F49 (EDULVLF), F55 (EDULVLM), Gq (IPRSPOT)
Croatia**	22.12.08-31.03.09	1 484	45.7	E23 (ACQ30BS), E24 (ACQ70BS), F23 (TPORGWK)
Cyprus	29.09.08-21.12.08	1 215	78.7	F6 (EDULVLA), F32 (HINCTNTA), F36 (EDULVLP), F49 (EDULVLF), F55 (EDULVLM)
Czech Republic	08.06.09-08.07.09	2 018	69.5	C18 (RLGDNM), C20 (RLGDNME), F6 (EDULVLA) F36 (EDULVLP), F49 (EDULVLF), F55 (EDULVLM)
Denmark	01.09.08-11.01.09	1 610	53.9	E23 (ACQ30BS), E24 (ACQ70BS), F36 (EDULVLP), F49 (EDULVLF), F55 (EDULVLM), F70 (CHLDHHE)
Estonia	05.11.08-11.03.09	1 661	57.4	F6 (EDULVLA), F6 (EISCED), F62 (MARITALA), F63 (LVGHWA), F64 (LVGPTNA), F66 (LVGPTNE), F68 (DVRCDV)
Finland	19.09.08-05.02.09	2 195	68.4	B 11 (VOTE), C18 (RLGDNM), C20 (RLGDNME), E11 (P20CNEC), E13 (P70CUST)

Control Chart (example from Lars Lyberg)





Special Cause (Non-Natural) Variation Comparative Meta-Data Documentation

Country	Question	Non-Natural Variation		
My Country	Q5	Response Category "it depends" missing		
	Q200	Question not asked		
	Q313	Question did not pass translation quality assessment		
Your Country	Q1	Question not asked		
	Q25	Wrong filtering resulting in inadequate target population		
	Q32	Response categories reversed		
	Q212	Translation comment part of questionnaire		



Conclusions

- A Meta-Data UFO (or better Hubble) will be a quantum leap in transparency, measurement and survey quality
- It requires accepting survey production complexity
- Complexity of comparative surveys is a power function of number of populations or languages under observation