What Skills Matter in Data Quality?

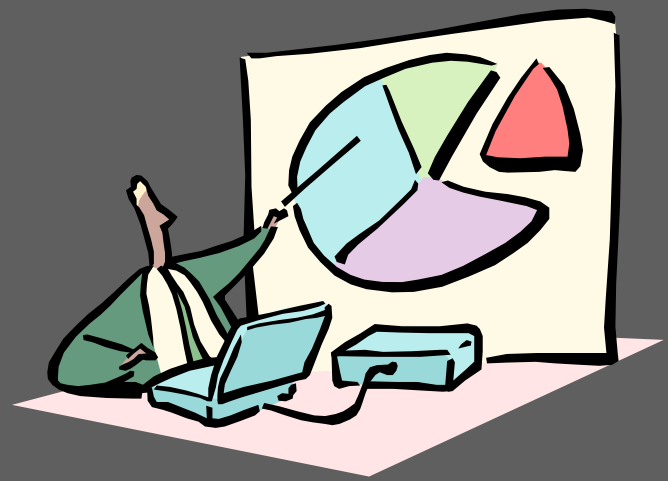
Research Team

WooYoung Chung
Case Western Reserve University

Craig Fisher
Marist College

Yang Lee
Northeastern University

Richard Wang
MIT & U.C. Berkeley

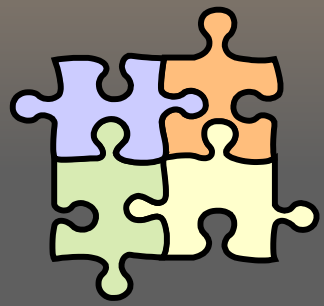


Presentation
The 7th International Conference on Information Quality

November 10, 2002

Presentation Roadmap

Challenge:
DQ/IQ is an inter-disciplinary field.



1. How should we conceptualize the DQ skills?
2. How do different sets of DQ skills complement one another?

3. What areas of DQ need more research and education?

- Conceptual Framework
  - General System Systems Theory
  - Classification of DQ skills
  - Literature Review
- Exploratory Survey Study
  - The skills survey at the 6th ICIQ
  - Preliminary Findings
  - Conclusions

General Systems Theory

Boulding (1956) General Systems Theory	Chaffee (1985) Classification of Business strategies	Morgan (1986) Classification of organizational theories	DQ/IQ Classification of DQ/IQ research and skills
Mechanical System (levels 1-3) Ex.: physics, economics	Linear Strategies Classic management theory, MBO	The machine metaphor	Technical (analytic) capabilities
Open Systems (levels 4-6) Ex.: biology, botany	Adaptive Strategies Contingency theory, OIP	The organism metaphor	Adaptive Capabilities
Human Systems (levels 7-9) Ex: sociology, anthropology	Interpretive Strategies Symbol management, social meanings	The brain metaphor	Interpretive Capabilites

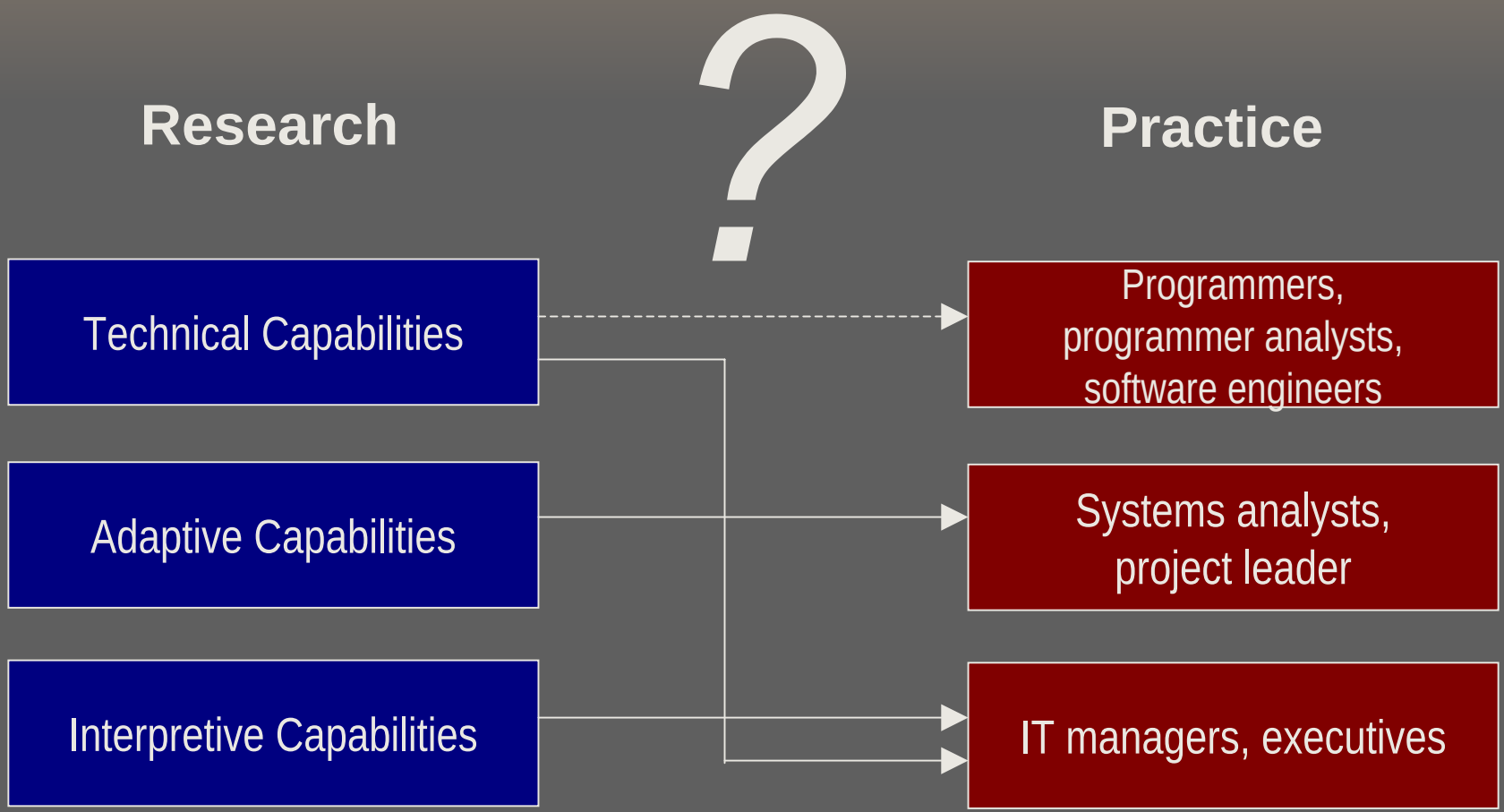
Research Examples

Classification	Examples
Technical Capabilities	Bohm and Jacopini, 1966; Chen, 1976; Coad and Yourdon, 1991; Codd, 1970; Dijkstra, 1968; Gane and Sarson, 1979
	Ballou and Pazer, 1985, 1995
Adaptive Capabilities	Adams, Nelson, and Todd, 1992; Davis, Bagozzi, and Warshaw, 1989; Mathieson, 1991; Taylor and Todd, 1995
	Baroudi, Olsen, and Ives, 1986; Chin and Newsted, 1995
	Brancheau and Wetherbe, 1990; Moore and Benbasat, 1991
	DeLone and McLean, 1992; Seddon, 1997
	Barki and Hartwick, 1994; Newman and Robey, 1992; Salaway, 1987;
	Orr, 1998; Wang and Strong, 1996

Research Examples

Classification	Examples
Interpretive Capabilities	Barley, 1986; DeSanctis and Poole, 1994; Orlikowski, 1992, 2000; Orlikowski and Yates, 1994;
	Constant, Sproull, and Kiesler, 1996; Pickering and King, 1995; Zack, 1993
	Weisband, 1987
	Davenport and Short, 1990; Hammer, 1990; Hammer and Champy, 1993

Research and Practice



Descriptive Statistics

Table 1

No.	Description	Mean	Std Dev	Based on T-Test	Based on Tukey	Subsets Based on Tukey			
1.	DQ measurement	6.33	0.96	1-2	1-10	A			
2.	DQ implications	6.10	1.08	1-6, 9	1-13	A	B		
3.	TQM	5.90	1.09	2-10	1-15	A	B	C	
4.	Data entry improvement	5.84	1.20	2-11	1-15	A	B	C	
5.	Org. policies	5.79	1.27	2-13	1-15	A	B	C	
6.	DB error detection	5.77	1.35	2-12	1-15	A	B	C	
7.	DQ dimensions	5.75	1.04	3-12	1-15	A	B	C	
8.	Change process	5.72	1.07	3-12	1-15	A	B	C	
9.	DQ cost/benefit	5.70	1.35	2-12	1-15	A	B	C	
10.	User requirements	5.67	1.14	3-14	1-15	A	B	C	
11.	Info. overload	5.49	1.21	4-15	2-15		B	C	
12.	DQ audit	5.46	1.18	5-15	2-15		B	C	
13.	Statistical techniques	5.30	1.46	5,10-15	2-18		B	C	D
14.	Data mining skills	5.23	1.50	10-15	3-18			C	D
15.	Data warehouse setup	5.18	1.43	11-15	3-18			C	D
16.	Analytic models	4.54	1.59	16-18	13-18				D
17.	Relational algebra	4.54	1.52	16-18	13-18				D
18.	Software tools	4.54	1.41	16-18	13-18				D

Factor Analysis

Table 2

	Factor 1	Factor 2	Factor 3
Factor 1 (Alpha = .798)			
14. Data Mining Skills	.783		
16. Analytic Models	.765		
15. Data Warehouse Setup	.726		
17. Relational Algebra	.588	.508	
13. Statistical Techniques	.577	.442	
Factor 2 (Alpha = .644)			
1. DQ Measurements		.813	
3. TQM		.675	.413
4. Data Entry Improvement		.612	
10. User Requirements		.553	
Factor 3 (Alpha = .699)			
8. Change Process			.795
2. DQ Implications			.694
9. DQ Cost/Benefit			.644
6. DB Error Detection	.466		.561

Skills Ratings by Job Titles

Table 5 – abridged

	Professors (N=10)		Executives Managers (N=11)		Consultants (N=10)		Project Managers Analysts (N=15)	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std Dev	Mean	Std Dev
Interpretive Capabilities	6.08	0.69	5.95	0.66	5.38	1.32	5.97	0.82
Adaptive Capabilities	5.68	0.83	5.64	1.10	6.30	0.60	6.17	0.60
Technical Capabilities	4.66	1.13	4.38	1.10	4.84	1.48	5.21	0.93

Conclusions

- Practitioners value diverse skills differently depending on their job situations.
- IS educators should design a data quality curriculum to fit the need of long-term and short-term career objectives of their student.
- Both academic researchers and executives reported Interpretive Capabilities—the ability of identifying and articulating organizational implications of data quality—as most important in improving and maintaining the data quality.
- Future research efforts should be directed to a systematic investigation of how data quality would affect the way in which people are organized and jobs are structured.