Transformation & Innovation Conference

Business Data Transformation

Prof. Paul A. Strassmann, George Mason University
May 22, 2006

How to Save $20 Billion through Transformation

	IBM in 1990	IBM in 2004
Purchase Order Process Time	One month	One day
Procurement Sources	300	3
Electronic Purchases	<20%	95%
E-Enabled Suppliers	<500	35,000
Electronic Catalogs	0	280

Indicators of Transformation Potential

	IBM - 2004	Navy - 2004
Revenue	$ 96.3 Billion	$ 103.7 Billion
Personnel	319,000	550,000
Management Layers	6	27
Total Number of Networks	1	850
IT Appllications	4,100	23,755
Financial Management Applications	406	1,083
HR Applications	300	708

A Historical Perspective

1771 - Factory mechanization.

1829 - Steam and Railways

1875 - Steel and Electricity

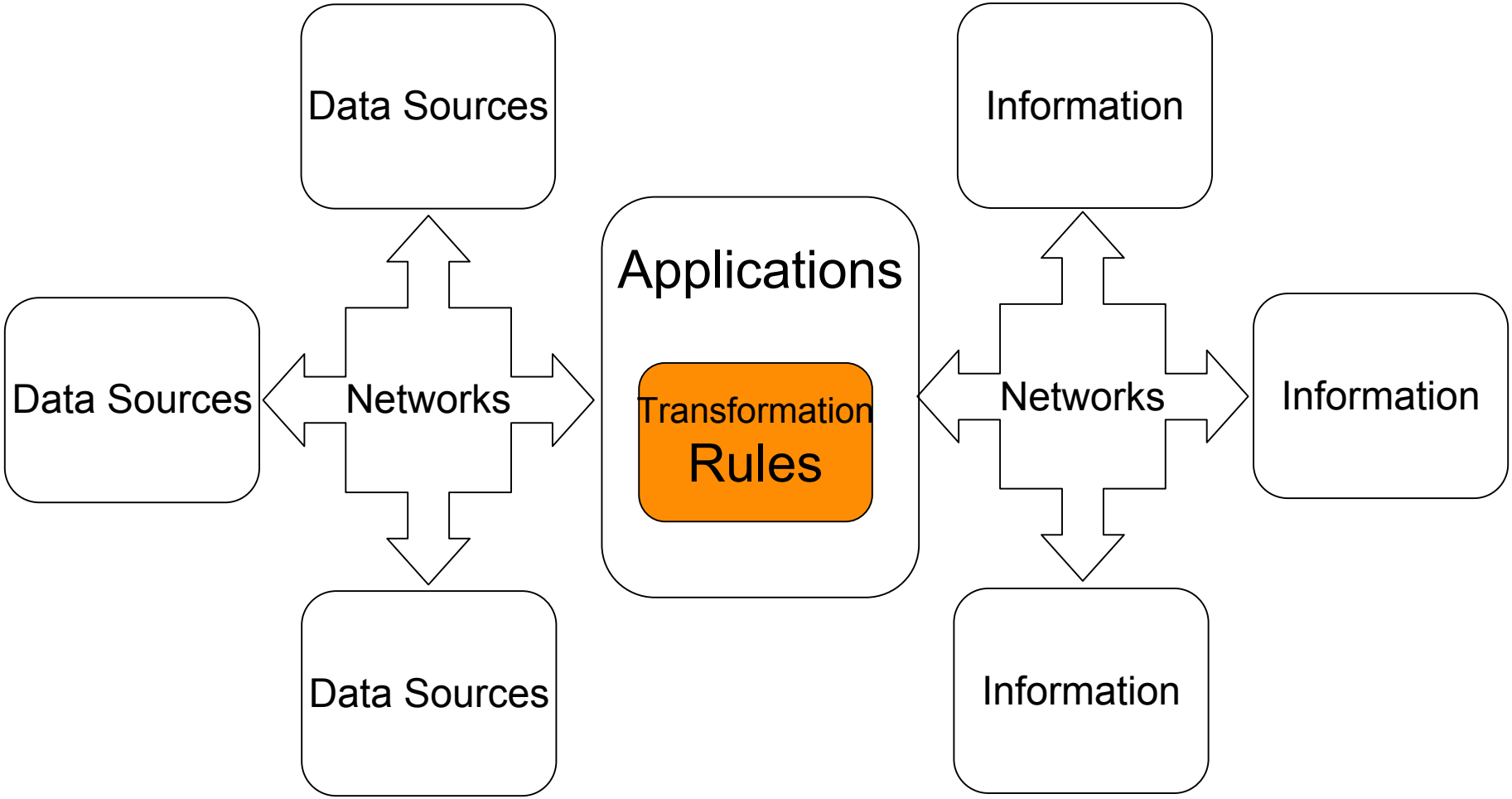
1908 - Automobile, Mass Production

1971 - Information and Telecommunications

The Context of Transformation

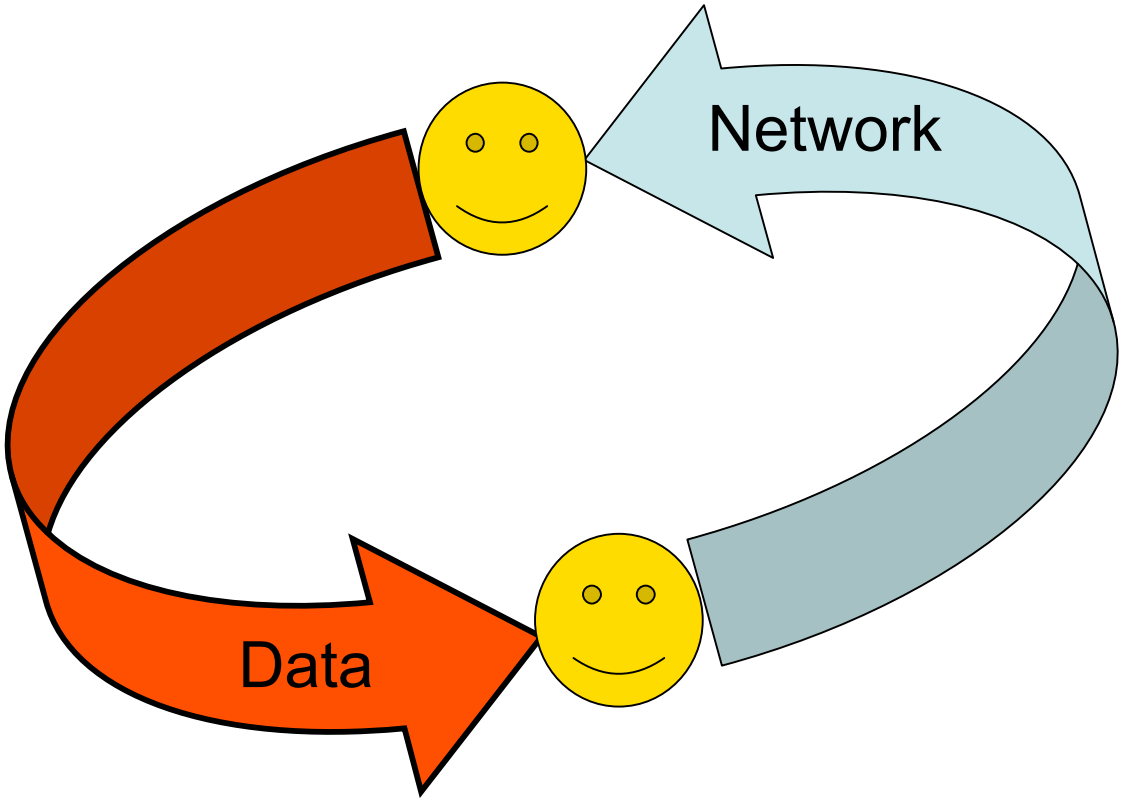
- Irruption: Invention/development of the new technological paradigm, decay of the preceding paradigm (1971).
- Frenzy: Rapid adoption of the new paradigm and intensive financial investment, a financial bubble (1998-2000).
- **Transformation: The rationalization of the new paradigm and renewed economic expansion after a purging of the excesses of the bubble (2005-?).**
- Maturity: Saturation and the gradual exhaustion of the potential of the new technology setting the stage for the next cycle (?).

How to Avoid the Possibility of 10 Trillion Errors?

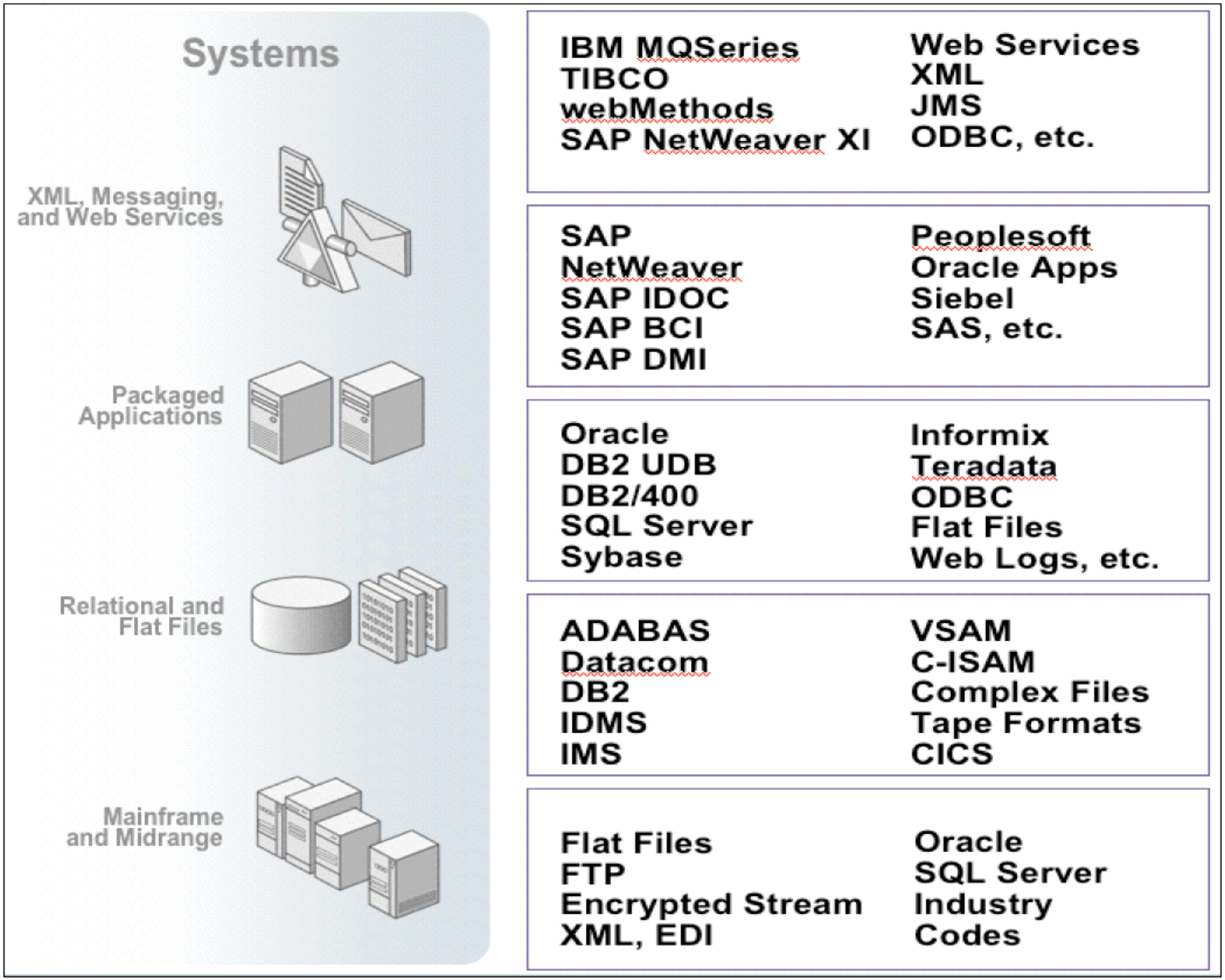


Example: I million Data Sources; 1,000 Networks; 10,000 Applications; 10 million Rules

A Data-Centric Perspective on Transformation



Where is the Data?



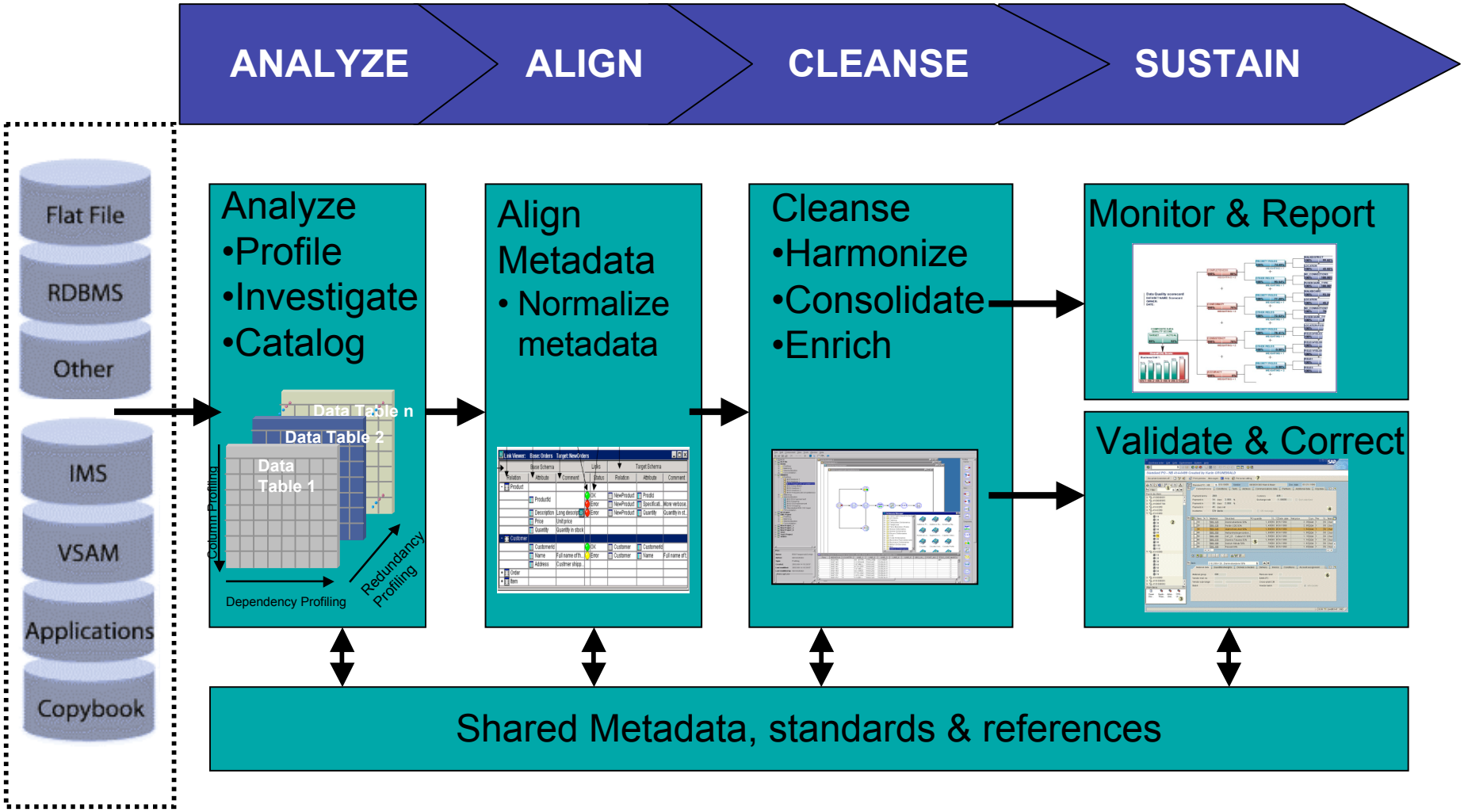
What Data?

UNSTRUCTURED	SEMI-STRUCTURED	XML
Microsoft Word	XML	LegalXML
Microsoft Excel	HL7	IFX
PowerPoint	HIPAA	cXML
PDF	ASTM	ebXML
Star Office	EDI–X12	HL7 V3.0
Word Perfect	EDI-Fact	ACORD (AL3,
ASCII reports	FIX	GJXDM
HTML	Cargo IMP	TWPDES
EBCDIC	MVR	
Undocumented	AFP	
Flat files	Post Script	
RPG	DJDE	
ANSI		

Example of Data Pollution

Citizen?	Frequency
US	9,451
Yes	228
USA	158
U.S.A.	128
U.S.	88
United States	20
(US)	5
Green Card	2
Naturalized	1
Applied	1
-	323

Transformation Process Engineering



Data Pollution Diagnostics

1-4GZUD	Bank One	456542345	1 Bank One Plaza Mail Suite Il	Chicago	Il	60670	USA	Current Cl
1-KRVV	Prudential Securities	23467T665	100 Mulberry Street	Newark	Nj	7102	usa	Current Cl
1-KRWP	Herzog Heine Geduld	7893434344	525 Washington Blvd	Jersey City	Nj	7310	USA	Current Cl
1-KRYD	Scudder Kemper Investments	7658786	222 South Riverside	Chicago	Il	60606	USA	Current Cl
1-KS2P	Pioneer Management		60 State Street	Boston	MD	2109	USA	Current Cl
1-KS4D	Donald & Company	999999999	512 Seventh Avenue	New York	Ny	10018	USA	Current Cl
1-KS55	Merrill Lynch	98756543	233 S. Wacker Drive	Chicago	Il	60606	USA	Current Cl
1-KS5S	Dni Holdings Inc Co Exodus Communications		300 Boulevard East	Weehawker	Nj	7087	USA	Current Cl
1-KS6K	Raymond James Financial Services-Br	# 4af01	5980 Lake Michigan Drive	Allendale	Mi	49401	USA	Current Cl
1-KS7F	Raymond James Financial Services-Br	# 4ky	Beckwith Street	Frenchtown	Mt	59834	USA	Current Cl
3D-50KX	3Com Corporation	999999999	11 Penn Plaza, Suite 1710	New York	New Yo	10001	USA	Prospect
JG-3FLC	Aquilla Corporation	9.9999E+10	Unknown	Kansas City		Unknown	USA	Prospect
JG-5GSE	MICHAEL FITZGERALD		Lake Union Building Suite 320	Seattle		98109	USA	Prospect
Y8+7V-51Y	Bank One, N.A.	1E+11	1 Bank One Plaza, IL1-0047	Chicago		60670	USA	Current Cl

COMPLETENESS CONFORMITY CONSISTENCY DUPLICATION INTEGRITY

Policy

Policy: Synchronization Through Metadata

- Data assets visible by associating metadata ("tagging") for each data element.
- Metadata conforms to Metadata Specifications.
- Metadata discoverable, searchable, and retrievable government-wide.

Rationalization Through Shared Services and Registry

- Data assets shall be accessible as shared services.
- Data assets shall conform to methods consistent with GIG technologies.
- Data assets shall be made understandable by semantic metadata in a DoD metadata registry.
- Data interoperability shall be supported by making data assets available and reused.

Example of a Metadata Registry

Mission of Metadata Registry

- The Metadata Registry provides software developers access to XML data and metadata components.
- This enhances capabilities for transforming shared data and to deploy "plug-and-play" software.

Economics

Excessive Costs for an Enterprise Environment

Number of Projects	Budget (BEA) Classification	FY 2006 Budget - $000	% of total
539	Warfighter	$7,887,228	26.2%
62	Acquisition	$159,932	0.5%
1,786	Enterprise Environment	$16,250,971	53.9%
241	Finance, Accounting	$481,541	1.6%
492	Human Resource Management	$1,859,608	6.2%
35	Installations and Environment	$117,901	0.4%
394	Logistics	$1,653,823	5.5%
77	Strategic Planning and Budgeting	$90,555	0.3%
362	Unspecified General Support	$1,647,667	5.5%
3,449	Subtotal - Support Costs	$22,261,998	73.8%
3,988	Total	$30,149,226	100.0%

Example of a Transformation Case

Data Management Functions	Prior to Transformation	After Transformation
Number of Applications	185	42
Number of Databases	40	2

Changing Scope After Transformation

Data Management Functions	Prior to Transformation	After Transformation
SCOPE OF DATA MANAGEMENT		
Data Sources	22,800	18,800
Median Number of Data Elements Entered / Transaction	8	10
Median Number of Transactions / Source / Day	800	750
Data Elements Entered / Year - Millions	53,261	51,465

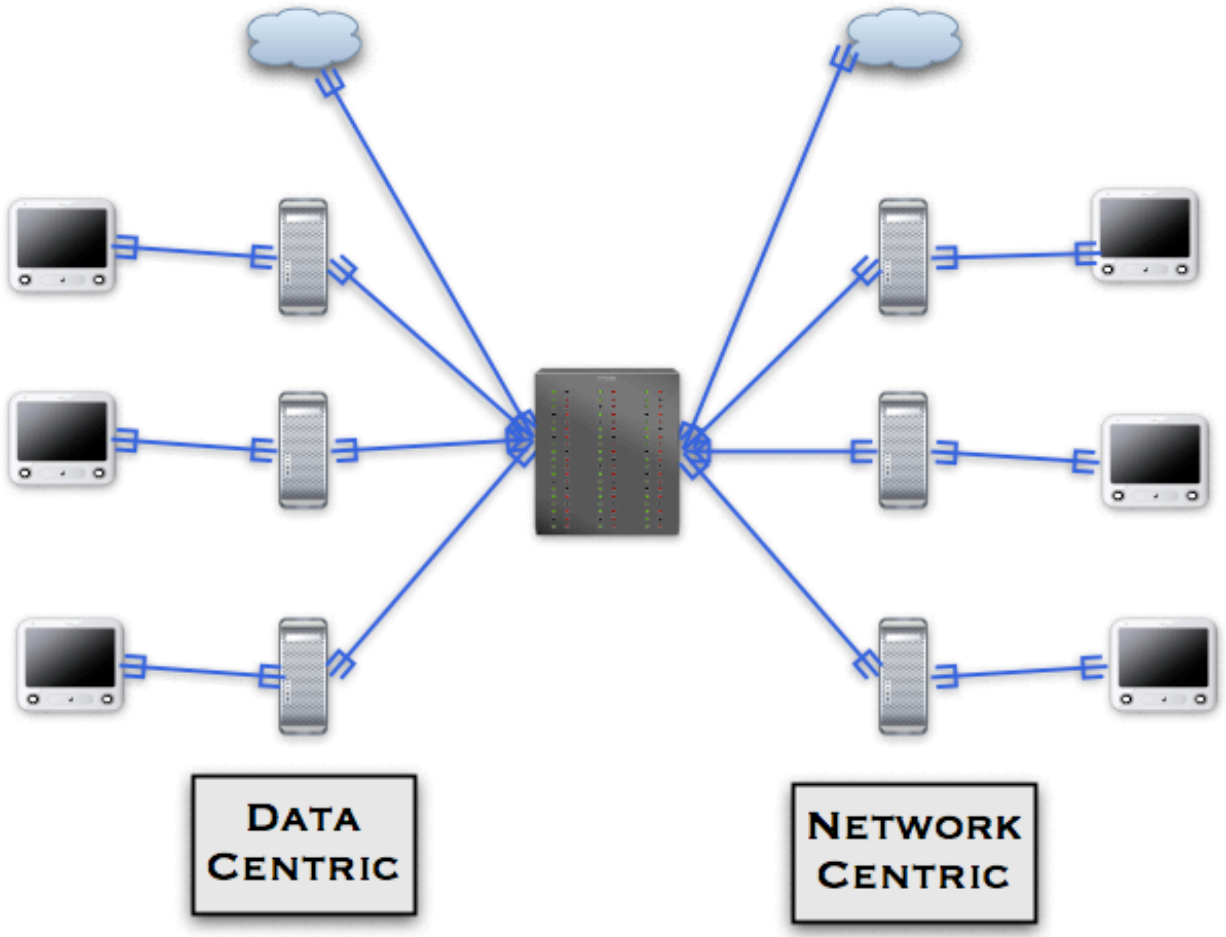
Changing Quality After Transformation

Data Management Functions	Prior to Transformation	After Transformation
DATA MANAGEMENT QUALITY MEASURES		
Data Element Entry Completeness per Requirement	97.00%	99.99%
Data Element Entry Verification against Data Dictionary	98.00%	99.99%
Inconsistent Duplicate Data Element Entered in Database	5.00%	0.50%
Conformity of Database Entries with Defined Rules	90.00%	99.00%

The Payoff from Data Transformation for Quality

Data Management Functions	Prior to Transformation	After Transformation
DATA MANAGEMENT METRICS		
Data Requiring Correction - Millions/Year	2,663.0	10.0
Defective/Inconsistent additions to Database - Millions/Year	266.0	2.6
DATA INTEGRITY MANAGEMENT COSTS		
Estimated cost of data audit and remediation - Cost per automated intervention for errors -$	$0.10	$0.20
Estimated administrative costs caused by data base defects - Cost per defect -$	$0.08	$0.03
Estimated Annual Costs for Defective Data Management - $Millions	$287.60	$2.14

Data Centricity at Enterprise Level



Summary: Enterprise Data Control is First Step in Transformation