

Monetizing Data Quality Management

Peter Aiken

peter@datablueprint.com



Peter Aiken

- BS VCU 1981 Information Systems & Management
- MS VCU 1985 Information Systems
- PhD GMU 1989 Information Technology Engineering
- Full time in information technology since 1981
- IT engineering research and project background
- University teaching experience since 1979
- Seven books and dozens of articles
- **Research Areas**
 - reengineering, data reverse engineering, software requirements engineering, information engineering, human-computer interaction, systems integration/ systems engineering, strategic planning, and DSS/BI
- **Director**
 - George Mason University/Hypermedia Laboratory (1989-1993)
- **DoD Computer Scientist**
 - Reverse Engineering Program Manager/Office of the Chief Information Officer (1992-1997)
- **Visiting Scientist**
 - Software Engineering Institute/Carnegie Mellon University (2001-2002)
- **DAMA International President** (<http://dama.org>)
 - 2001 DAMA International Individual Achievement Award (with Dr. E. F. "Ted" Codd)
 - 2005 DAMA Community Award
- **Founding Advisor/International Association for Information and Data Quality** (<http://iaidq.org>)
- **Founding Advisor/Meta-data Professionals Organization** (<http://metadataprofessional.org>)
- **Founding Director Data Blueprint** 1993



Monetizing Data Quality Management

- Why is it important?
 - *Concretizing*
- State Agency Time & Leave Tracking
 - \$1 million USD annually
- ERP Implementation
 - \$1 million USD on a large project
- Data Warehouse Quality Analysis
 - \$5 billion USD US DoD (prevention)
- MDM British Telecom rollout
 - £ 250 (small investment)
- Non-Monetized Example
 - Different measures
- ERP Implementation Legal Case
 - \$ 5,355,450 CAN damages/penalties



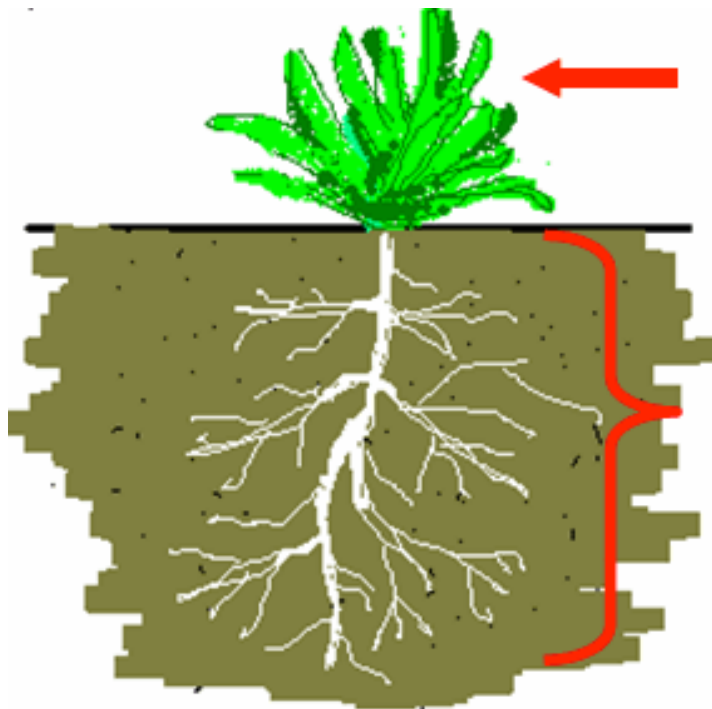
IT Project Failure Rates

Recent IT project failure rates statistics can be summarized as follows:

- *Carr 1994*
 - 16% of IT Projects completed on time, within budget, with full functionality
- *OASIG Study (1995)*
 - 7 out of 10 IT projects "fail" in some respect
- *The Chaos Report (1995)*
 - 75% blew their schedules by 30% or more
 - 31% of projects will be canceled before they ever get completed
 - 53% of projects will cost over 189% of their original estimates
 - 16% for projects are completed on-time and on-budget
- *KPMG Canada Survey (1997)*
 - 61% of IT projects were deemed to have failed
- *Conference Board Survey (2001)*
 - Only 1 in 3 large IT project customers were very "satisfied"
- *Robbins-Gioia Survey (2001)*
 - 51% of respondents viewed their large IT implementation project as unsuccessful
- *MacDonalds Innovate (2002)*
 - Automate fast food network from fry temperature to # of burgers sold-\$180M USD write-off
- *Ford Everest (2004)*
 - Replacing internal purchasing systems-\$200 million over budget
- *FBI (2005)*
 - Blew \$170M USD on suspected terrorist database-"start over from scratch"



Root Cause Analysis

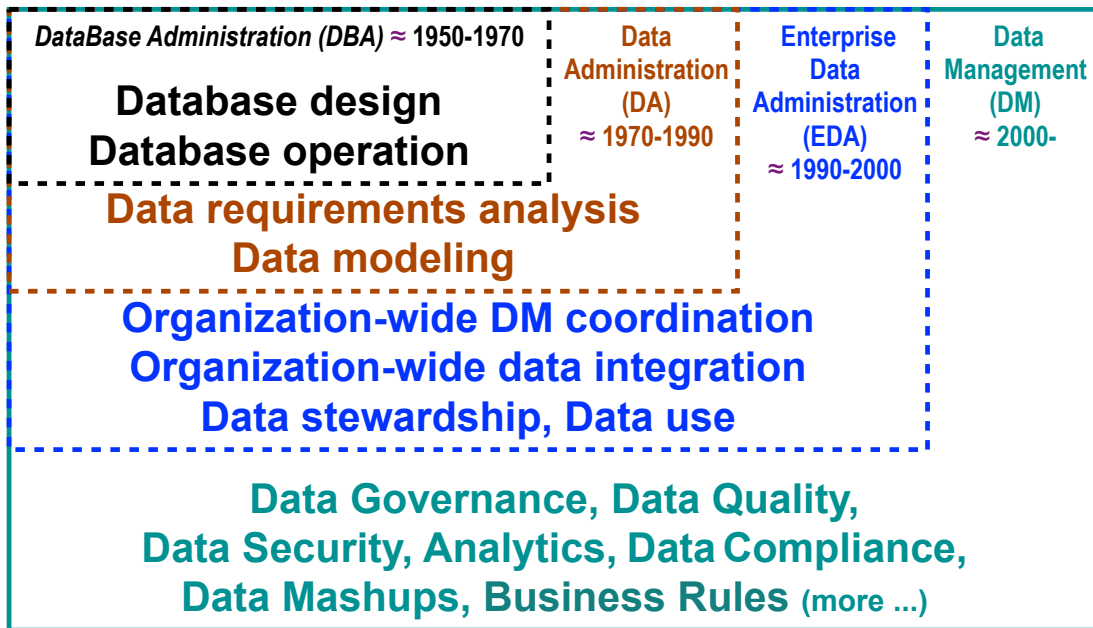


- Symptom of the problem
 - The weed
 - Above the surface
 - Obvious
- The underlying Cause
 - The root
 - Below the surface
 - Not obvious
- Poor Data Quality Management Practices

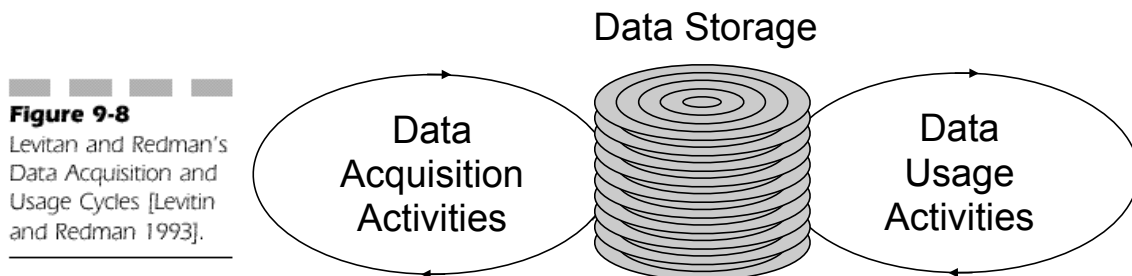
Monetizing - from Wikipedia

- Monetization is the process of converting or establishing something into legal tender.
- It usually refers to the printing of banknotes by central banks, but things such as gold, diamonds and emeralds, and art can also be monetized.
- Even intrinsically worthless items can be made into money, as long as they are difficult to make or acquire.

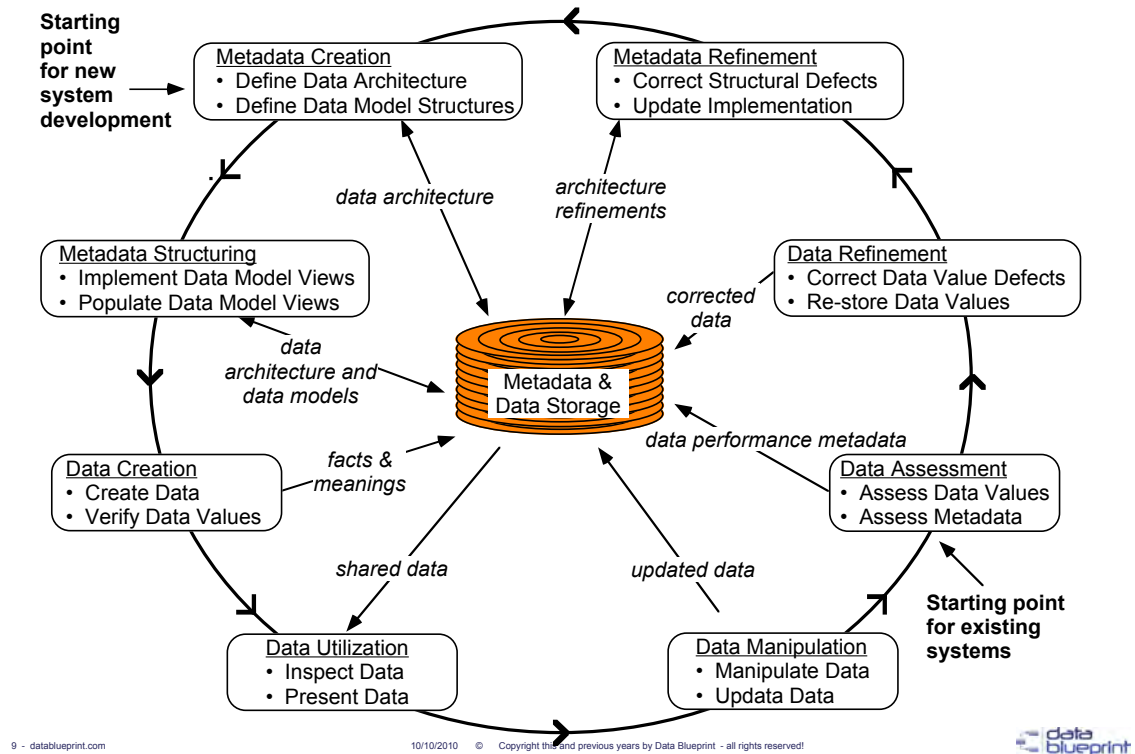
Expanding DM Scope



Traditional Quality Life Cycle



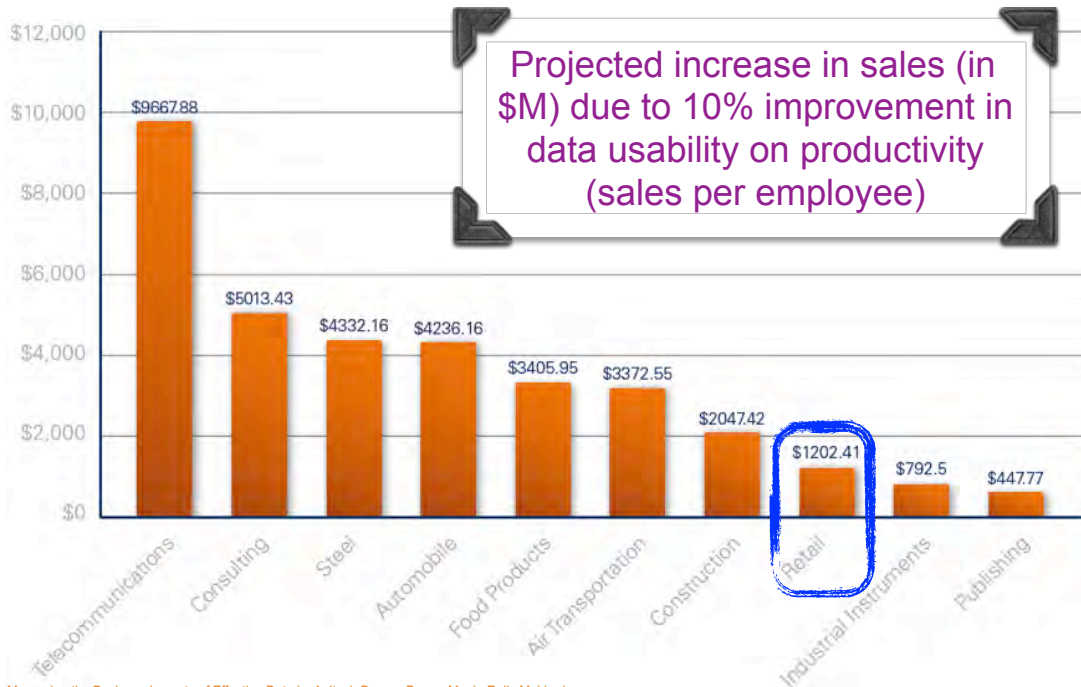
Extended data life cycle model with metadata sources and uses



Academic Research Findings

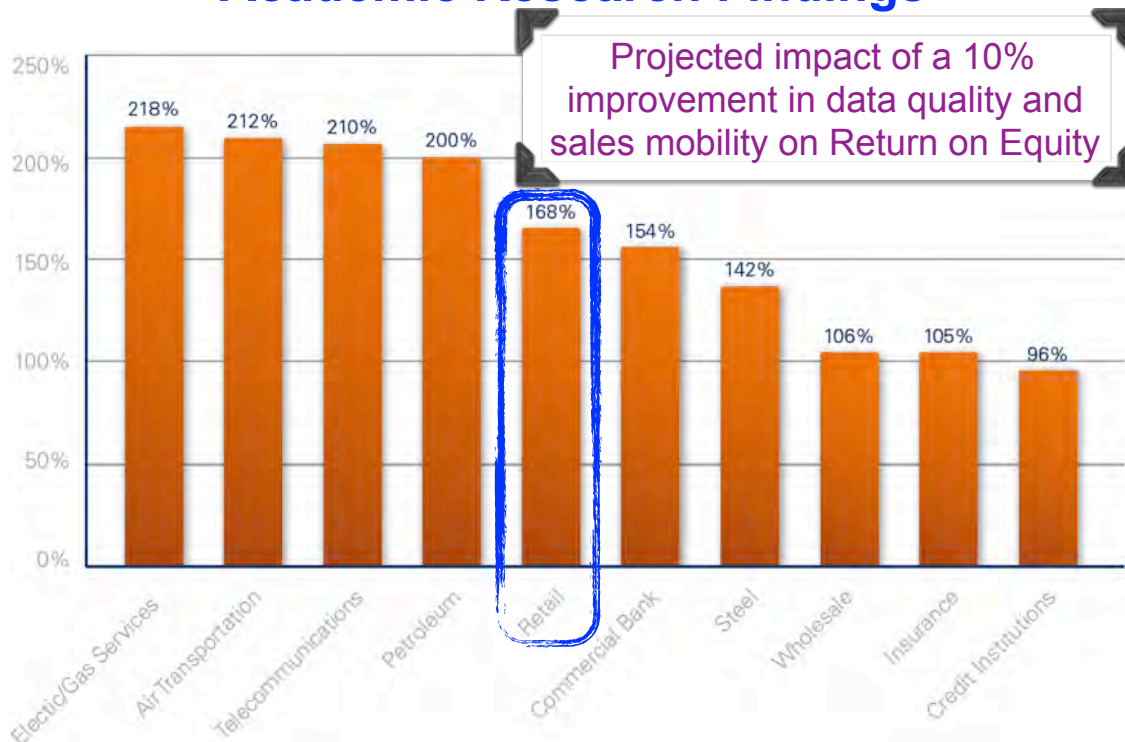


Academic Research Findings



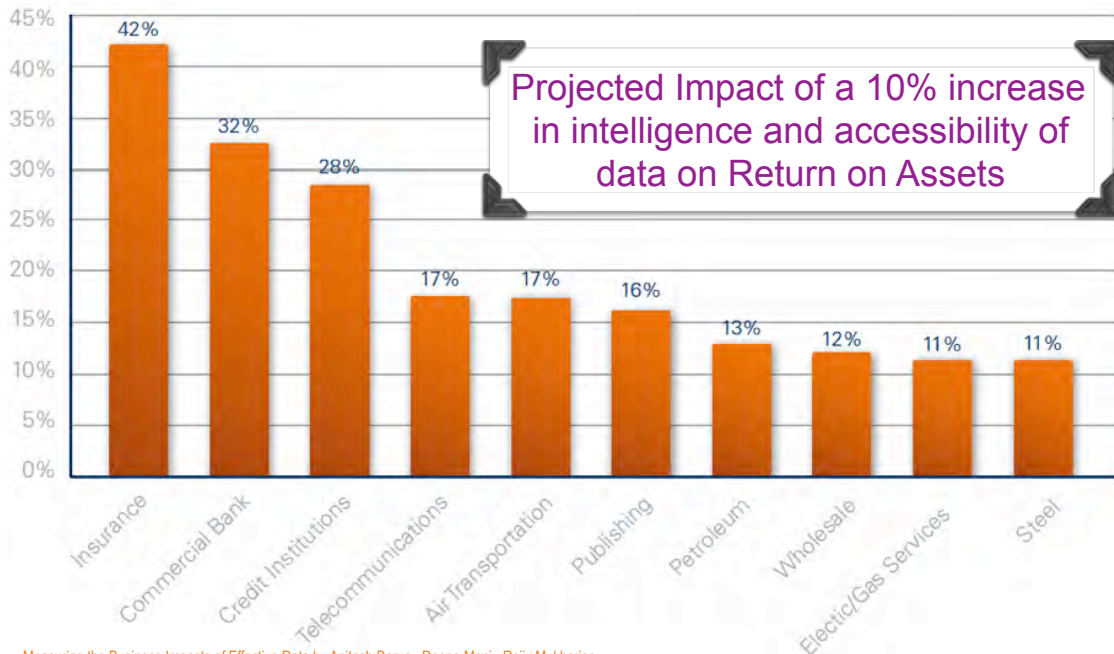
Measuring the Business Impacts of Effective Data by Anitesh Barua,, Deepa Mani,, Rajiv Mukherjee

Academic Research Findings



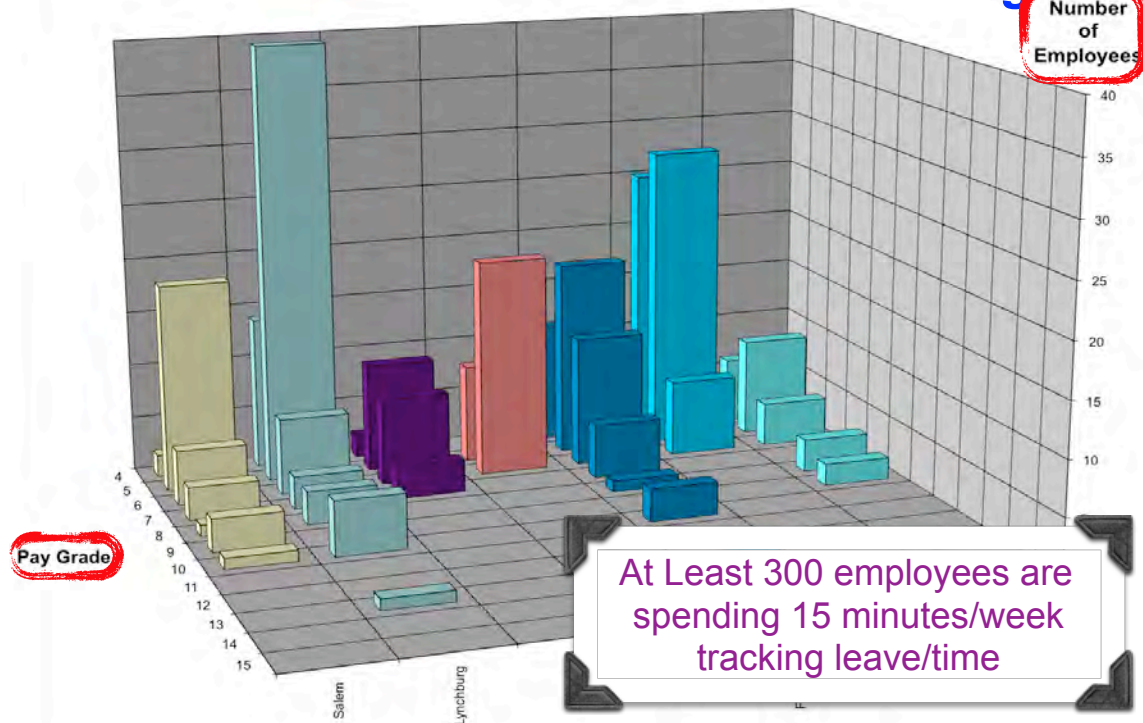
Measuring the Business Impacts of Effective Data by Anitesh Barua,, Deepa Mani,, Rajiv Mukherjee

Academic Research Findings



Measuring the Business Impacts of Effective Data by Anitesh Barua, Deepa Mani, Rajiv Mukherjee

Monitization: Time & Leave Tracking



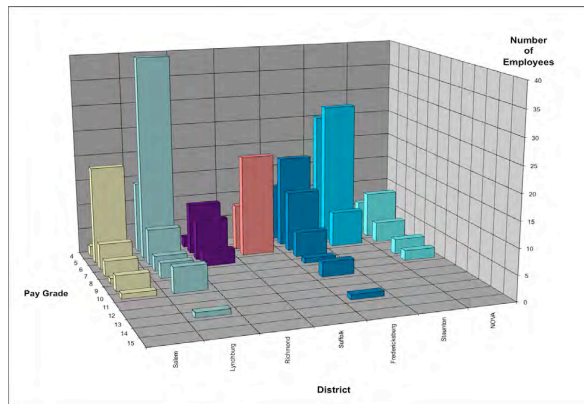
Annual Grade	Step																
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	\$ 12,020.00	\$ 12,290.00	\$ 12,567.00	\$ 12,850.00	\$ 13,140.00	\$ 13,436.00	\$ 13,738.00	\$ 14,048.00	\$ 14,364.00	\$ 14,688.00	\$ 15,018.00	\$ 15,357.00	\$ 15,703.00	\$ 16,056.00	\$ 16,418.00	\$ 16,788.00	\$ 17,166.00
2	\$ 13,140.00	\$ 13,436.00	\$ 13,738.00	\$ 14,048.00	\$ 14,364.00	\$ 14,688.00	\$ 15,018.00	\$ 15,357.00	\$ 15,703.00	\$ 16,056.00	\$ 16,418.00	\$ 16,788.00	\$ 17,166.00	\$ 17,552.00	\$ 17,948.00	\$ 18,352.00	\$ 18,765.00
3	\$ 14,364.00	\$ 14,688.00	\$ 15,018.00	\$ 15,357.00	\$ 15,703.00	\$ 16,056.00	\$ 16,418.00	\$ 16,788.00	\$ 17,166.00	\$ 17,552.00	\$ 17,948.00	\$ 18,352.00	\$ 18,765.00	\$ 19,188.00	\$ 19,620.00	\$ 20,062.00	\$ 20,514.00
4	\$ 15,703.00	\$ 16,056.00	\$ 16,418.00	\$ 16,788.00	\$ 17,166.00	\$ 17,552.00	\$ 17,948.00	\$ 18,352.00	\$ 18,765.00	\$ 19,188.00	\$ 19,620.00	\$ 20,062.00	\$ 20,514.00	\$ 20,976.00	\$ 21,449.00	\$ 21,932.00	\$ 22,426.00
5	\$ 17,166.00	\$ 17,552.00	\$ 17,948.00	\$ 18,352.00	\$ 18,765.00	\$ 19,188.00	\$ 19,620.00	\$ 20,062.00	\$ 20,514.00	\$ 20,976.00	\$ 21,449.00	\$ 21,932.00	\$ 22,426.00	\$ 22,931.00	\$ 23,447.00	\$ 23,975.00	\$ 24,515.00
6	\$ 18,765.00	\$ 19,188.00	\$ 19,620.00	\$ 20,062.00	\$ 20,514.00	\$ 20,976.00	\$ 21,449.00	\$ 21,932.00	\$ 22,426.00	\$ 22,931.00	\$ 23,447.00	\$ 23,975.00	\$ 24,515.00	\$ 25,068.00	\$ 25,632.00	\$ 26,209.00	\$ 26,800.00
7	\$ 20,514.00	\$ 20,976.00	\$ 21,449.00	\$ 21,932.00	\$ 22,426.00	\$ 22,931.00	\$ 23,447.00	\$ 23,975.00	\$ 24,515.00	\$ 25,068.00	\$ 25,632.00	\$ 26,209.00	\$ 26,800.00	\$ 27,403.00	\$ 28,020.00	\$ 28,652.00	\$ 29,297.00
8	\$ 22,426.00	\$ 22,931.00	\$ 23,447.00	\$ 23,975.00	\$ 24,515.00	\$ 25,068.00	\$ 25,632.00	\$ 26,209.00	\$ 26,800.00	\$ 27,403.00	\$ 28,020.00	\$ 28,652.00	\$ 29,297.00	\$ 29,956.00	\$ 30,630.00	\$ 31,320.00	\$ 32,027.00
9	\$ 24,515.00	\$ 25,068.00	\$ 25,632.00	\$ 26,209.00	\$ 26,800.00	\$ 27,403.00	\$ 28,020.00	\$ 28,652.00	\$ 29,297.00	\$ 29,956.00	\$ 30,630.00	\$ 31,320.00	\$ 32,027.00	\$ 32,752.00	\$ 33,495.00	\$ 34,257.00	\$ 35,038.00
10	\$ 26,800.00	\$ 27,403.00	\$ 28,020.00	\$ 28,652.00	\$ 29,297.00	\$ 29,956.00	\$ 30,630.00	\$ 31,320.00	\$ 32,027.00	\$ 32,752.00	\$ 33,495.00	\$ 34,257.00	\$ 35,038.00	\$ 35,838.00	\$ 36,657.00	\$ 37,495.00	\$ 38,353.00
11	\$ 29,297.00	\$ 29,956.00	\$ 30,630.00	\$ 31,320.00	\$ 32,027.00	\$ 32,752.00	\$ 33,495.00	\$ 34,257.00	\$ 35,038.00	\$ 35,838.00	\$ 36,657.00	\$ 37,495.00	\$ 38,353.00	\$ 39,230.00	\$ 40,126.00	\$ 41,043.00	\$ 41,980.00
12	\$ 32,027.00	\$ 32,752.00	\$ 33,495.00	\$ 34,257.00	\$ 35,038.00	\$ 35,838.00	\$ 36,657.00	\$ 37,495.00	\$ 38,353.00	\$ 39,230.00	\$ 40,126.00	\$ 41,043.00	\$ 41,980.00	\$ 42,937.00	\$ 43,915.00	\$ 44,914.00	\$ 45,935.00
13	\$ 35,038.00	\$ 35,838.00	\$ 36,657.00	\$ 37,495.00	\$ 38,353.00	\$ 39,230.00	\$ 40,126.00	\$ 41,043.00	\$ 41,980.00	\$ 42,937.00	\$ 43,915.00	\$ 44,914.00	\$ 45,935.00	\$ 46,977.00	\$ 48,041.00	\$ 49,127.00	\$ 50,235.00
14	\$ 38,353.00	\$ 39,126.00	\$ 40,126.00	\$ 41,043.00	\$ 41,980.00	\$ 42,937.00	\$ 43,915.00	\$ 44,914.00	\$ 45,935.00	\$ 46,977.00	\$ 48,041.00	\$ 49,127.00	\$ 50,235.00	\$ 51,366.00	\$ 52,520.00	\$ 53,698.00	\$ 54,899.00
15	\$ 41,980.00	\$ 42,783.00	\$ 43,747.00	\$ 44,732.00	\$ 45,740.00	\$ 46,770.00	\$ 47,823.00	\$ 48,900.00	\$ 50,002.00	\$ 51,128.00	\$ 52,280.00	\$ 53,457.00	\$ 54,661.00	\$ 55,892.00	\$ 57,151.00	\$ 58,439.00	\$ 59,755.00
16	\$ 45,740.00	\$ 46,770.00	\$ 47,823.00	\$ 48,900.00	\$ 50,002.00	\$ 51,128.00	\$ 52,280.00	\$ 53,457.00	\$ 54,661.00	\$ 55,892.00	\$ 57,151.00	\$ 58,439.00	\$ 59,755.00	\$ 61,101.00	\$ 62,477.00	\$ 63,884.00	\$ 65,323.00
17	\$ 50,002.00	\$ 51,128.00	\$ 52,280.00	\$ 53,457.00	\$ 54,661.00	\$ 55,892.00	\$ 57,151.00	\$ 58,439.00	\$ 59,755.00	\$ 61,101.00	\$ 62,477.00	\$ 63,884.00	\$ 65,323.00	\$ 66,794.00	\$ 68,299.00	\$ 69,837.00	\$ 71,410.00
18	\$ 54,661.00	\$ 55,892.00	\$ 57,151.00	\$ 58,439.00	\$ 59,755.00	\$ 61,101.00	\$ 62,477.00	\$ 63,884.00	\$ 65,323.00	\$ 66,794.00	\$ 68,299.00	\$ 69,837.00	\$ 71,410.00	\$ 73,019.00	\$ 74,663.00	\$ 76,345.00	\$ 78,064.00
19	\$ 59,755.00	\$ 61,101.00	\$ 62,477.00	\$ 63,884.00	\$ 65,323.00	\$ 66,794.00	\$ 68,299.00	\$ 69,837.00	\$ 71,410.00	\$ 73,019.00	\$ 74,663.00	\$ 76,345.00	\$ 78,064.00	\$ 79,820.00	\$ 81,613.00	\$ 83,444.00	\$ 85,313.00
20	\$ 65,323.00	\$ 66,794.00	\$ 68,299.00	\$ 69,837.00	\$ 71,410.00	\$ 73,019.00	\$ 74,663.00	\$ 76,345.00	\$ 78,064.00	\$ 79,820.00	\$ 81,613.00	\$ 83,444.00	\$ 85,313.00	\$ 87,221.00	\$ 89,168.00	\$ 91,155.00	\$ 93,182.00
21	\$ 71,410.00	\$ 73,019.00	\$ 74,663.00	\$ 76,345.00	\$ 78,064.00	\$ 79,820.00	\$ 81,613.00	\$ 83,444.00	\$ 85,313.00	\$ 87,221.00	\$ 89,168.00	\$ 91,155.00	\$ 93,182.00	\$ 95,250.00	\$ 97,359.00	\$ 99,509.00	\$ 101,700.00
22	\$ 78,064.00	\$ 79,820.00	\$ 81,613.00	\$ 83,444.00	\$ 85,313.00	\$ 87,221.00	\$ 89,168.00	\$ 91,155.00	\$ 93,182.00	\$ 95,250.00	\$ 97,359.00	\$ 99,509.00	\$ 101,700.00	\$ 103,933.00	\$ 106,208.00	\$ 108,525.00	\$ 110,884.00
23	\$ 85,313.00	\$ 87,221.00	\$ 89,168.00	\$ 91,155.00	\$ 93,182.00	\$ 95,250.00	\$ 97,359.00	\$ 99,509.00	\$ 101,700.00	\$ 103,933.00	\$ 106,208.00	\$ 108,525.00	\$ 110,884.00	\$ 113,286.00	\$ 115,732.00	\$ 118,223.00	\$ 120,759.00

Monthly Grade	Step																
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	\$ 1,001.67	\$ 1,024.17	\$ 1,047.25	\$ 1,070.83	\$ 1,095.00	\$ 1,119.87	\$ 1,144.83	\$ 1,170.67	\$ 1,197.00	\$ 1,224.00	\$ 1,251.50	\$ 1,279.75	\$ 1,308.58	\$ 1,338.00	\$ 1,368.17	\$ 1,398.00	\$ 1,428.50
2	\$ 1,095.00	\$ 1,119.87	\$ 1,144.83	\$ 1,170.67	\$ 1,197.00	\$ 1,224.00	\$ 1,251.50	\$ 1,279.75	\$ 1,308.58	\$ 1,338.00	\$ 1,368.17	\$ 1,398.00	\$ 1,428.50	\$ 1,459.67	\$ 1,491.50	\$ 1,524.00	\$ 1,557.17
3	\$ 1,197.00	\$ 1,224.00	\$ 1,251.50	\$ 1,279.75	\$ 1,308.58	\$ 1,338.00	\$ 1,368.17	\$ 1,398.00	\$ 1,428.50	\$ 1,459.67	\$ 1,491.50	\$ 1,524.00	\$ 1,557.17	\$ 1,591.00	\$ 1,625.50	\$ 1,660.67	\$ 1,696.50
4	\$ 1,308.58	\$ 1,338.00	\$ 1,368.17	\$ 1,398.00	\$ 1,428.50	\$ 1,459.67	\$ 1,491.50	\$ 1,524.00	\$ 1,557.17	\$ 1,591.00	\$ 1,625.50	\$ 1,660.67	\$ 1,696.50	\$ 1,733.00	\$ 1,770.00	\$ 1,807.50	\$ 1,845.50
5	\$ 1,430.50	\$ 1,462.67	\$ 1,495.67	\$ 1,529.33	\$ 1,563.75	\$ 1,599.00	\$ 1,635.00	\$ 1,671.83	\$ 1,709.50	\$ 1,748.00	\$ 1,787.42	\$ 1,827.67	\$ 1,868.63	\$ 1,910.32	\$ 1,952.75	\$ 1,995.92	\$ 2,039.83
6	\$ 1,563.75	\$ 1,599.00	\$ 1,635.00	\$ 1,671.83	\$ 1,709.50	\$ 1,748.00	\$ 1,787.42	\$ 1,827.67	\$ 1,868.63	\$ 1,910.32	\$ 1,952.75	\$ 1,995.92	\$ 2,039.83	\$ 2,084.50	\$ 2,129.92	\$ 2,186.17	\$ 2,243.25
7	\$ 1,709.50	\$ 1,748.00	\$ 1,787.42	\$ 1,827.67	\$ 1,868.63	\$ 1,910.32	\$ 1,952.75	\$ 1,995.92	\$ 2,040.00	\$ 2,084.92	\$ 2,130.67	\$ 2,177.25	\$ 2,224.67	\$ 2,272.92	\$ 2,322.00	\$ 2,371.92	\$ 2,422.67
8	\$ 1,868.63	\$ 1,910.32	\$ 1,952.75	\$ 1,995.92	\$ 2,040.00	\$ 2,084.92	\$ 2,130.67	\$ 2,177.25	\$ 2,224.67	\$ 2,272.92	\$ 2,322.00	\$ 2,371.92	\$ 2,422.67	\$ 2,474.25	\$ 2,526.67	\$ 2,579.92	\$ 2,634.00
9	\$ 2,042.92	\$ 2,089.00	\$ 2,136.00	\$ 2,184.00	\$ 2,233.33	\$ 2,283.58	\$ 2,335.00	\$ 2,387.67	\$ 2,441.42	\$ 2,496.25	\$ 2,552.25	\$ 2,609.50	\$ 2,668.00	\$ 2,727.75	\$ 2,788.83	\$ 2,850.25	\$ 2,913.00
10	\$ 2,233.33	\$ 2,283.58	\$ 2,335.00	\$ 2,387.67	\$ 2,441.42	\$ 2,496.25	\$ 2,552.25	\$ 2,609.50	\$ 2,668.00	\$ 2,727.75	\$ 2,788.83	\$ 2,850.25	\$ 2,913.00	\$ 2,977.17	\$ 3,042.75	\$ 3,109.75	\$ 3,178.17
11	\$ 2,441.42	\$ 2,496.25	\$ 2,552.25	\$ 2,609.50	\$ 2,668.00	\$ 2,727.75	\$ 2,788.83	\$ 2,850.25	\$ 2,913.00	\$ 2,977.17	\$ 3,042.75	\$ 3,109.75	\$ 3,178.17	\$ 3,248.00	\$ 3,319.25	\$ 3,391.92	\$ 3,466.00
12	\$ 2,668.92	\$ 2,729.00	\$ 2,790.00	\$ 2,853.33	\$ 2,919.00	\$ 2,986.00	\$ 3,054.25	\$ 3,123.75	\$ 3,194.50	\$ 3,266.50	\$ 3,339.75	\$ 3,414.25	\$ 3,490.00	\$ 3,566.92	\$ 3,645.00	\$ 3,724.25	\$ 3,804.67
13	\$ 2,917.67	\$ 2,983.33	\$ 3,050.58	\$ 3,119.25	\$ 3,189.50	\$ 3,261.33	\$ 3,334.83	\$ 3,409.92	\$ 3,486.50	\$ 3,564.67	\$ 3,644.42	\$ 3,724.75	\$ 3,805.75	\$ 3,887.50	\$ 3,970.00	\$ 4,053.25	\$ 4,137.25
14	\$ 3,189.50	\$ 3,261.33	\$ 3,334.83	\$ 3,409.92	\$ 3,486.50	\$ 3,564.67	\$ 3,644.42	\$ 3,724.75	\$ 3,805.75	\$ 3,887.50	\$ 3,970.00	\$ 4,053.25	\$ 4,137.25	\$ 4,222.00	\$ 4,307.50	\$ 4,393.75	\$ 4,480.75
15	\$ 3,486.75	\$ 3,565.25	\$ 3,645.58	\$ 3,727.67	\$ 3,811.67	\$ 3,897.50	\$ 3,985.25	\$ 4,075.00	\$ 4,166.83	\$ 4,260.67	\$ 4,356.67	\$ 4,454.75	\$ 4,555.00	\$ 4,656.42	\$ 4,759.00	\$ 4,862.75	\$ 4,967.67
16	\$ 3,811.67	\$ 3,897.50	\$ 3,985.25	\$ 4,075.00	\$ 4,166.83	\$ 4,260.67	\$ 4,356.67	\$ 4,454.75	\$ 4,555.00	\$ 4,656.42	\$ 4,759.00	\$ 4,862.75	\$ 4,967.67	\$ 5,073.75	\$ 5,181.00	\$ 5,289.33	\$ 5,398.75
17	\$ 4,166.83	\$ 4,260.67	\$ 4,356.67	\$ 4,454.75	\$ 4,555.00	\$ 4,656.42	\$ 4,759.00	\$ 4,862.75	\$ 4,967.67	\$ 5,073.75	\$ 5,181.00	\$ 5,289.33	\$ 5,398.75	\$ 5,508.25	\$ 5,618.92	\$ 5,730.75	\$ 5,843.75
18	\$ 4,555.00	\$ 4,656.42	\$ 4,759.00	\$ 4,862.75	\$ 4,967.67	\$ 5,073.75	\$ 5,181.00	\$ 5,289.33	\$ 5,398.75	\$ 5,508.25	\$ 5,618.92	\$ 5,730.75	\$ 5,843.75	\$ 5,957.00	\$ 6,071.50	\$ 6,187.25	\$ 6,304.25
19	\$ 4,979.58	\$ 5,091.75	\$ 5,206.42	\$ 5,323.67	\$ 5,443.58	\$ 5,566.17	\$ 5,691.50	\$ 5,819.50	\$ 5,950.83	\$ 6,084.92	\$ 6,221.92	\$ 6,362.00	\$ 6,505.33	\$ 6,650.92	\$ 6,800.00	\$ 6,951.67	\$ 7,105.00
20	\$ 5,443.58	\$ 5,566.17	\$ 5,691.50	\$ 5,819.50	\$ 5,950.83	\$ 6,084.92	\$ 6,221.92	\$ 6,362.00	\$ 6,505.33	\$ 6,650.92	\$ 6,800.00	\$ 6,951.67	\$ 7,105.00	\$ 7,260.92	\$ 7,419.50	\$ 7,580.75	\$ 7,744.67
21	\$ 5,950.83	\$ 6,084.92	\$ 6,221.92	\$ 6,362.00	\$ 6,505.33	\$ 6,650.92	\$ 6,800.00	\$ 6,951.67	\$ 7,105.00	\$ 7,260.92	\$ 7,419.50	\$ 7,580.75	\$ 7,744.67	\$ 7,910.25	\$ 8,078.50	\$ 8,249.50	\$ 8,423.25
22	\$ 6,505.33	\$ 6,650.92	\$ 6,800.00	\$ 6,951.67	\$ 7,105.00	\$ 7,260.92	\$ 7,419.50	\$ 7,580.75	\$ 7,744.67	\$ 7,910.25	\$ 8,078.50	\$ 8,249.50	\$ 8,423.25	\$ 8,599.00	\$ 8,776.25	\$ 8,955.25	\$ 9,137.00
23	\$ 7,111.58	\$ 7,271.75	\$ 7,435.50	\$ 7,603.00	\$ 7,774.25	\$ 7,949.33	\$ 8,128.42	\$ 8,311.50	\$ 8,498.67	\$ 8,690.17	\$ 8,885.83	\$ 9,084.75	\$ 9,286.92	\$ 9,492.33	\$ 9,701.00	\$ 9,912.00	\$ 10,125.25

Pay Period Grade	Step																
	5	6	7	8	9	10	11	12	13	14							

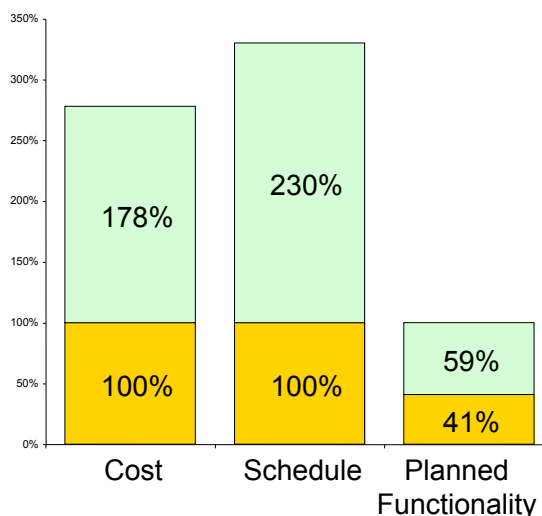
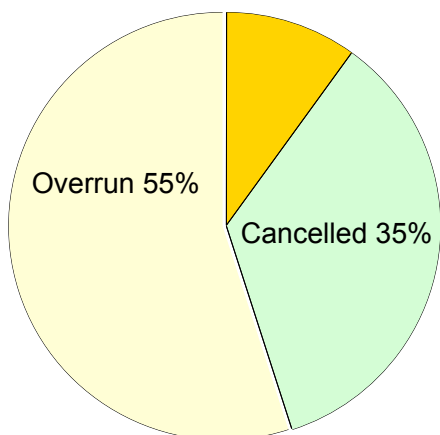
Annual Organizational Totals

- Range \$192,000 - \$159,000/month
- \$100,000 Salem
- \$159,000 Lynchburg
- \$100,000 Richmond
- \$100,000 Suffolk
- \$150,000 Fredericksburg
- \$100,000 Staunton
- \$100,000 NOVA
- \$800,000/month or **\$9,600,000/annually**
- *Awareness of the cost of things considered overhead*



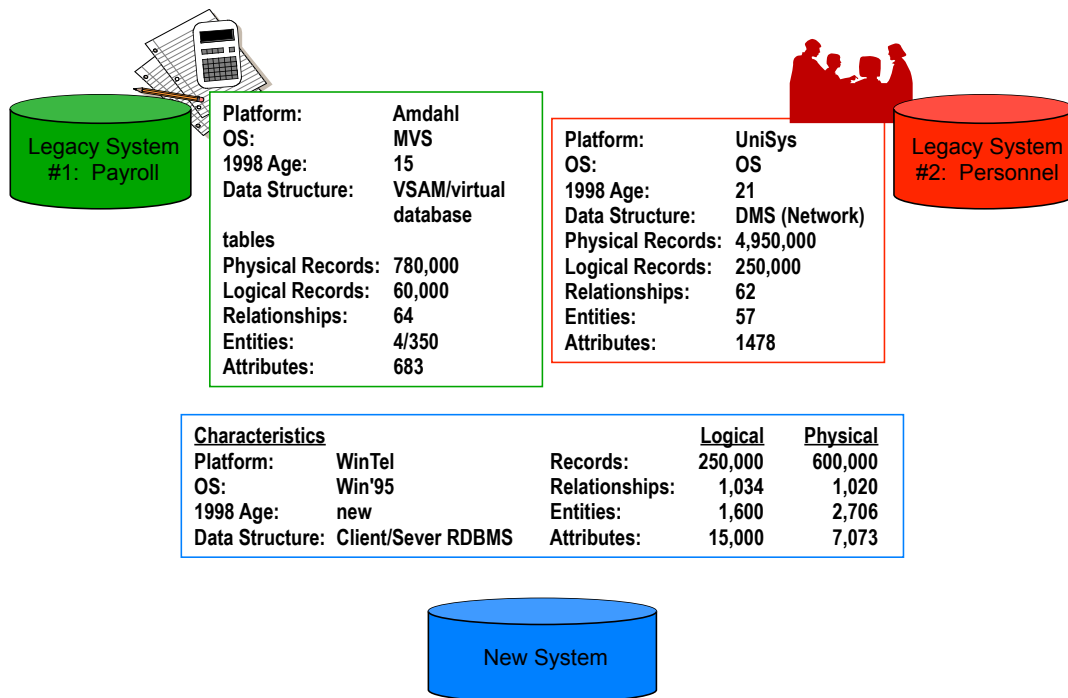
ERP Implementation Success

On time, within budget, as planned 10%



- Most ERP implementations today result in cost and schedule overruns; courtesy of the Standish Group

Predicting Engineering Problem Characteristics



"Extreme" Data Engineering

- 2 person months = 40 person days
- 2,000 attributes mapped onto 15,000
- 2,000/40 person days = 50 attributes per person day
or 50 attributes/8 hour = 6.25 attributes/hour
- and
- 15,000/40 person days = 375 attributes per person day
or 375 attributes/8 hours = 46.875 attributes/hour
- Locate, identify, understand, map, transform, document, QA at a rate of -
- 52 attributes every 60 minutes or **.86 attributes/minute!**



Why ETL and Data Migration Projects Fail

— Joseph R. Hudicka
Information Architecture Team



- Assessed 1200 migration projects!

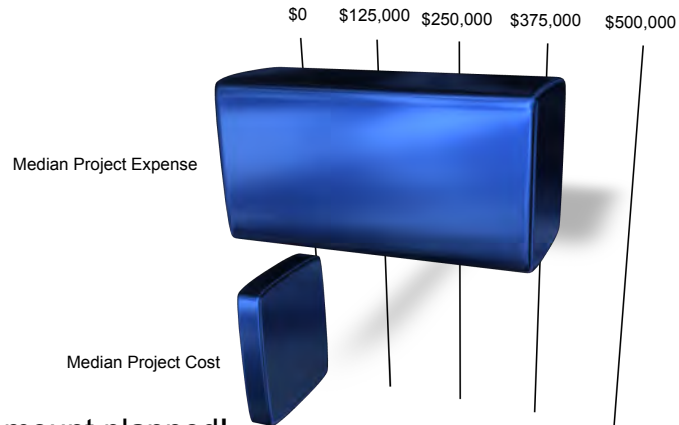
- Surveyed only experienced migration specialists who have done at least four migration projects

- The median project costs over 10 times the amount planned!

- Biggest Challenges: Bad Data; Missing Data; Duplicate Data

- The survey did not consider projects that were cancelled largely due to data migration difficulties

- "... problems are encountered rather than discovered"



Legacy System Migration to ERP: MDM

- Challenge

- Millions of NSN/SKUs
- Key and other data stored in clear text/comment fields
- Original suggestion was manual approach to text extraction
- Left the MDM structuring problem unsolved

- Solution

- Proprietary, improvable text extraction process
- Converted non-tabular data into tabular data
- Saved a minimum of \$5 million
- Literally person centuries of work



An Iterative Approach to MDM Structuring

	Unmatched Items	Unmatched Items (% Total)	Ignorable NSNs	Ignorable Items (% Total)	Items Matched	Avg Extracted Per Item	Items Matched (% Total) Items Extracted	
1	329948	31.47%	14034	1.34%	N/A	N/A	N/A	264703
2	222474	21.22%	73069	6.97%	N/A	N/A	N/A	286675
3	216552	20.66%	78520	7.49%	N/A	N/A	N/A	287196
4	340514	32.48%	125708	11.99%	582101	1.1000222	55.53%	640324
...
14	94542	9.02%	237113	22.62%	716668	1.1142914	68.36%	798577
15	94929	9.06%	237118	22.62%	716276	1.1139282	68.33%	797880
16	99890	9.53%	237128	22.62%	711305	1.1153008	67.85%	793319
17	99591	9.50%	237128	22.62%	711604	1.1154392	67.88%	793751
18	78213	7.46%	237130	22.62%	732980	1.2072812	69.92%	884913



Quantitative Benefits

Time needed to review all NSNs once over the life of the project:	
NSNs	2,000,000
Average time to review & cleanse (in minutes)	5
Total Time (in minutes)	10,000,000
Time available per resource over a one year period of time:	
Work weeks in a year	48
Work days in a week	5
Work hours in a day	7.5
Work minutes in a day	450
Total Work minutes/year	108,000
Person years required to cleanse each NSN once prior to migration:	
Minutes needed	10,000,000
Minutes available person/year	108,000
Total Person-Years	92.6
Resource Cost to cleanse NSN's prior to migration:	
Avg Salary for SME year (not including overhead)	\$60,000.00
Projected Years Required to Cleanse/Total DLA Person Year	93
Total Cost to Cleanse/Total DLA Savings to Cleanse NSN's:	\$5.5 million

Business Implications

- National Stock Number (NSN) Discrepancies
 - If NSNs in LUAF, GABF, and RTLS are not present in the MHIF, these records cannot be updated in SASSY
 - Additional overhead is created to correct data before performing the real maintenance of records
- Serial Number Duplication
 - If multiple items are assigned the same serial number in RTLS, the traceability of those items is severely impacted
 - Approximately \$531 million of SAC 3 items have duplicated serial numbers
- On-Hand Quantity Discrepancies
 - If the LUAF O/H QTY and number of items serialized in RTLS conflict, there can be no clear answer as to how many items a unit actually has on-hand
 - Approximately **\$5 billion** of equipment does not tie out between the LUAF and RTLS



Date: Tue, 26 Mar 2002 10:47:52 -0500
 From: Jamie McCarthy <jamie@mccarthy.vg>
 Subject: Friendly Fire deaths traced to dead battery

In one of the more horrifying incidents I've read about, U.S. soldiers and allies were killed in December 2001 because of a stunningly poor design of a GPS receiver, plus "human error."

<http://www.washingtonpost.com/wp-dyn/articles/A8853-2002Mar23.html>

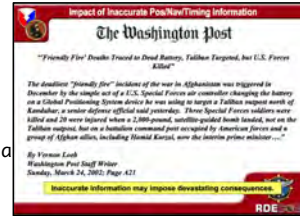
A U.S. Special Forces air controller was calling in GPS positioning from some sort of battery-powered device. He "had used the GPS receiver to calculate the latitude and longitude of the Taliban position in minutes and seconds for an airstrike by a Navy F/A-18."

According to the *Post* story, the bomber crew "required" a "second calculation in 'degree decimals'" -- why the crew did not have equipment to perform the minutes-seconds conversion themselves is not explained.

The air controller had recorded the correct value in the GPS receiver when the battery died. Upon replacing the battery, he called in the degree-decimal position the unit was showing -- without realizing that the unit is set up to reset to its *own* position when the battery is replaced.

The 2,000-pound bomb landed on his position, killing three Special Forces soldiers and injuring 20 others.

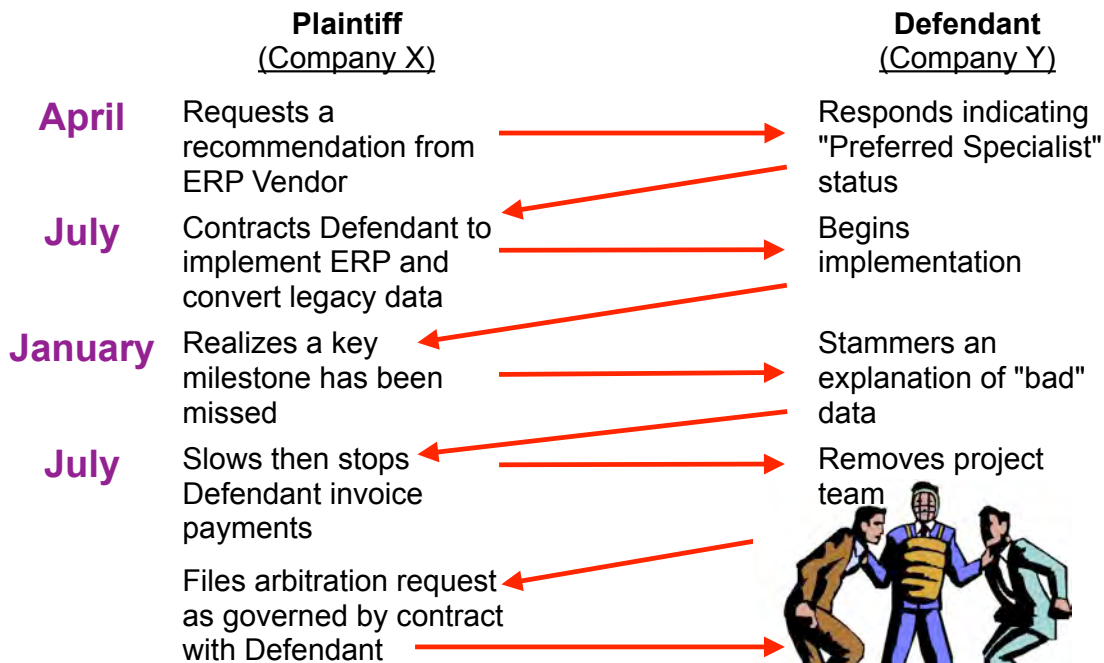
If the information in this story is accurate, the RISKS involve replacing memory settings with an apparently-valid default value instead of blinking 0 or some other obviously-wrong display; not having a backup battery to hold values in memory during battery replacement; not equipping users to translate one coordinate system to another (reminiscent of the Mars Climate Orbiter slamming into the planet when ground crews confused English with metric); and using a device with such flaws in a combat situation



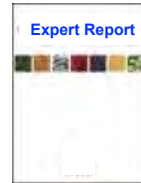
Friendly Fire deaths traced to Dead Battery



Messy Sequencing Towards Arbitration
















Expert Reports



Ours provided evidence that :

1. Company Y's conversion code introduced errors into the data
2. Some data that Company Y converted was of measurably lower quality than the quality of the data before the conversion
3. Company Y caused harm by not performing an analysis of the Company X's legacy systems and that that the required analysis was not a part of any project plan used by Company Y
4. Company Y caused harm by withholding specific information relating to the perception of the on-site consultants' views on potential project success

FBI & Canadian Social Security Gender Codes

-  1. Male
-  2. Female
-   3. Formerly male now female
-   4. Formerly female now male
-  5. Uncertain
-  6. Won't tell
-  7. Doesn't know
-   8. Male soon to be female
-   9. Female soon to be male

If column 1 in source = "m"

- *then set value of target data to "male"*
- *else set value of target data to "female"*

Hypothesized extensions contributed by a Chicago DAMA Member

10. Psychologically female, biologically male
11. Psychologically male, biologically female
12. Both soon to be female
13. Both soon to be male

AJHR0213_CAN_UPDATE.SQR

```
!-----  
! Procedure Name: 230-Assign-PS-EmpId  
!  
! Description : This procedure generates a PeopleSoft Employee ID  
! (EmpId) by incrementing the last EmpId processed by 1  
! First it checks if the applicant/employee exists on  
! the PeopleSoft database using the SSN.  
!-----  
Begin-Procedure 230-Assign-PS-EmpId  
  
move 'N' to $found_in_PS          !DAR 01/14/04  
move 'N' to $found_on_XXX        !DAR 01/14/04  
  
BEGIN-SELECT -Db'DSN=HR83PRD;UID=PS_DEV;PWD=psdevelopment'  
NID.EMPLID  
NID.NATIONAL_ID  
  
move 'Y' to $found_in_PS          !DAR 01/14/04  
move &NID.EMPLID to $ps_empId  
  
FROM PS_PERS_NID NID  
!WHERE NID.NATIONAL_ID = $ps_ssn  
WHERE NID.AJ_APPL_ID = $applicant_id  
END-SELECT  
  
if $found_in_PS = 'N'            !DAR 01/14/04  
do 231-Check-XXX-for-EmpId      !DAR 01/14/04  
if $found_on_XXX = 'N'          !DAR 01/14/04  
add 1 to #last_empId  
let $last_empId = to_char(#last_empId)  
let $last_empId = lpad($last_empId,6,'0')  
let $ps_empId = 'A' || $last_empId  
end-if  
end-if                            !DAR 01/14/04
```

The defendant knew to prevent duplicate SSNs

The exclamation point prevents this line from looking for duplicates, so no check is made for a duplicate SSN/National ID

Legacy systems business rules allowed employees to have more than one AJ_APPL_ID.

SQL Query Analyzer - [Query - PSDBCLUSTER.FS84PRD.PS_READ - Untitled2*]

```
SELECT COUNT(*) FROM PS_CUSTOMER  
SELECT COUNT(DISTINCT EMPLID) FROM HR83PRD..PS_EMPLOYMENT
```

(No column name)
63131
100236

Query batch completed. PSDBCLUSTER (8.0) PS_READ (252) FS84PRD 0:00:00 2 rows Ln 2, Col 58
Connections: 2 NUM

Identified & Quantified Risks

Risk	High	Med	Low	Notes
User knowledge transfer – training	X			The lack of permanent PeopleSoft technical staff will make knowledge transfer a greater challenge as compared to a project where the PeopleSoft technical staff members are permanent employees. Also, the Sales Colleagues are spread throughout the country and will need to do inquiries upon the system. Training and Buy-in will be challenging for them.
User Buy-In	X			End users are focused on the current environment. Change introduces overhead to their daily operations that affects how willingly they adapt to any new application and the necessary related process changes.
implementation and rollout not timely or complete.	X			
Quality of Conversion Data	X			has customers in 4 different databases, each of which may contain the same customer multiple times with active open balances.

Risk Response

“Risk response development involves defining enhancement steps for opportunities and threats.”

Page 119, Duncan, W., *A Guide to the Project Management Body of Knowledge*, PMI, 1996

Tasks	Hours
New Year Conversion	120
Tax and payroll balance conversion	120
General Ledger conversion	80
Total	320

"The go-live date may need to be extended due to certain critical path deliverables not being met. This extension will require additional tasks and resources. The decision of whether or not to extend the go-live date should be made by Monday, November 3, 20XX so that resources can be allocated to the additional tasks."

Resource	Hours
G/L Consultant	40
Project Manager	40
Recievables Consultant	40
HRMS Technical Consultant	40
Technical Lead Consultant	40
HRMS Consultant	40
Financials Technical Consultant	40
Total	280

Delay	Weekly Resources	Weeks	Tasks	Cumulative
January (5 weeks)	280	5	320	1720
February (4 weeks)	280	4		1120
			Total	2840

Professional & Workmanlike Manner

warrants that the services it provides hereunder will be performed in a professional and workmanlike manner in accordance with industry standards.

Defendant warrants that the services it provides hereunder will be performed in a professional and workmanlike manner in accordance with industry standards.

The Defense's "Industry Standards"

- Question:
 - What are the industry standards that you are referring to?
- Answer:
 - There is nothing written or codified, but it is the standards which are recognized by the consulting firms in our (industry).
- Question:
 - I understand from what you told me just a moment ago that the industry standards that you are referring to here are not written down anywhere; is that correct?
- Answer:
 - That is my understanding.
- Question:
 - Have you made an effort to locate these industry standards and have simply not been able to do so?
- Answer:
 - I would not know where to begin to look.



Published Industry Standards Guidance

- Examples from the:
- IEEE (365,000 members)
 - Institute of Electrical and Electronic Engineers
 - 150 countries, 40 percent outside the United States
 - 128 transactions, journals and magazines
 - 300 conferences
- ACM (80,000+ members)
 - Association of Computing Machinery
 - 100 conferences annually
- ICCP (50,000+ members)
 - Institute for Certification of Computing Professionals
- DAMA International (3,500+ members)
 - Data Management Association
 - Largest Data/Metadata conference



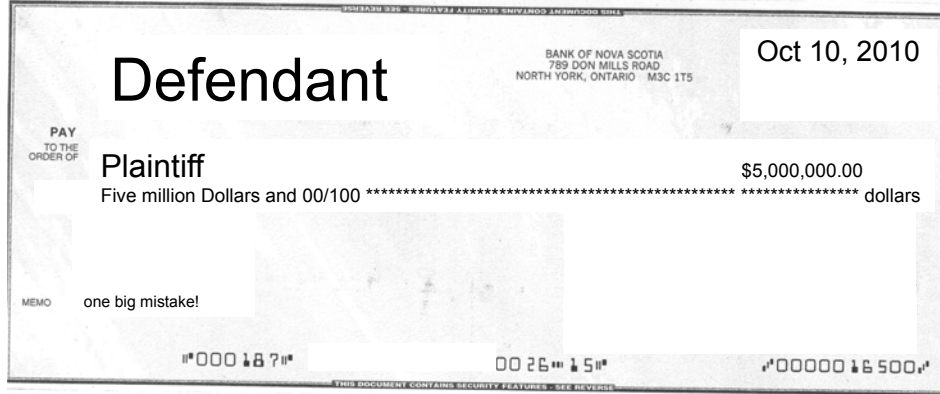
IEEE Code of Ethics

We, the members of the IEEE, in recognition of the importance of our technologies in affecting the quality of life throughout the world, and in accepting a personal obligation to our profession, its members and the communities we serve, do hereby commit ourselves to the highest ethical and professional conduct and agree:

1. To accept responsibility in making engineering decisions consistent with the safety, health and welfare of the public, and to disclose promptly factors that might endanger the public or the environment;
2. To avoid real or perceived conflicts of interest whenever possible, and to disclose them to affected parties when they do exist;
3. **To be honest and realistic in stating claims or estimates based on available data;**
4. To reject bribery in all its forms;
5. **To improve the understanding of technology, its appropriate application, and potential consequences;**
6. To maintain and improve our technical competence and to undertake technological tasks for others only if qualified by training or experience, or after full disclosure of pertinent limitations;
7. To seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, and to credit properly the contributions of others;
8. To treat fairly all persons regardless of such factors as race, religion, gender, disability, age, or national origin;
9. To avoid injuring others, their property, reputation, or employment by false or malicious action;
10. To assist colleagues and co-workers in their professional development and to support them in following this code of ethics. [Approved by the IEEE Board of Directors, August 1990]

Outcome

- Three days after the hearing, the panel issued a one-page decision awarding damages of \$5 million to Company X



<http://peteraiken.net>



Contact Information:
 Peter Aiken, Ph.D.
Department of Information Systems
 School of Business
 Virginia Commonwealth University
 Snead Hall Room B4217
 301 West Main Street
 Richmond, Virginia 23284-4000

Data Blueprint
 10124C West Broad Street
 Glen Allen VA 23060
 804.521.4056
<http://datablueprint.com>

office: +1.804.883.7594
 cell: +1.804.382.5957

e-mail: peter@datablueprint.com
 web: <http://peteraiken.net>

Do more with your data Visit DataBlueprint.com

Peter Aiken
 Associate Professor of Information Systems at Virginia Commonwealth University, Author, Speaker, Musician, and Founder of **Data Blueprint, Inc.**

Professional Personal

My background is a combination of real world experience and **academically based research**. My career positions have ranged from the retail and other businesses to computer systems engineering and more recently to **research-based consulting**.
Peter Aiken

Professional
 Data Blueprint Biography Publications VCU business Speaker
 DAMA International Data-Ed Online

Personal
 Contact Affiliations Music Philosophies Did you see this?

News
 Ask Peter Where in the world is Peter? Newflash

10/10/2010 © Copyright this and previous years by Data Blueprint - all rights reserved!