

# Information Flows

## A Tool for Enterprise Architecture Reengineering

Dr. Jerry Rosenbaum  
The Rose Tree Group

410-764-1843

[jrosenba@ix.netcom.com](mailto:jrosenba@ix.netcom.com)

# Outline

- Background – Approaches to System Change
- Introduction to Information Flows
- Initial Observations
- Organizing Information Flows
- Data Transformation Questions
- Using Information Flows to Reengineer Systems
- Concluding Remarks

# Two Approaches to System Change

- For system change, one needs a target architecture
  - **Approach 1 – Clean Slate**: Determine the requirements and build an entirely new system. If an old program is usable, you reuse it.
  - **Approach 2 – Evolutionary Migration**: Determine the requirements and migrate step by step from the existing system to the new target architecture
- But one must always know the starting point

# Key Questions

- How will the new system (or changed system) fit into the current environment
- What data does the new system need and from where will it get it.
- Will other systems need data from the new system, and how will they get it
- Is there a way to migrate from the current situation to an improved situation



# Not Atypical List of Documentation

- Tends to be rather sparse (and often outdated)

# So, Why Does the CIO Care

- COST
- Typical companies spend between 50-90% of their IT people and software budget on maintenance
- Gartner groups estimates that 60-80% of the IT people and software budget is the “bridges”

# There Goes Your Budget

- If we take 75% as the maintenance figure and 75% as the “bridge” maintenance then
    - You have just committed 52.5% of the people and software budget to “bridges”
  - Example
    - \$2B manufacturer spends 2% on IT
      - 60% of that is the IT personal (\$40M)
      - So \$21 M +software costs goes to bridges
    - \$4B insurance company spends 10% in IT
      - So \$210 M+software costs goes to bridges
- (And that is using conservative estimates)

# Words of Wisdom

If you do not properly analyze and understand the current system(s), your probability of success is significantly reduced

ETL is portrayed as very simple by the vendors, but that is not the whole story

# Introduction to Information Flows

# Information Flow Basics

- Sender of the data
- Receiver of the data
- Mechanism over which the data is sent
- What data is sent/received and is it transformed

# Part 1 – The Sending Mechanism



# Part 2 – Receiving Mechanism



# Part 3 – Send It Over the Wall



# Goals of Information Flows

- To gain a better understanding of the current spider webs of systems and their interactions and interrelationships and the overall business flow
- To improve data quality
- To reduce the risk failure of system modifications, maintenance, and migrations
- To rationalize the
  - Data
  - Systems
  - Flow
- To determine the contents of a “core” centralized database

# What are Information Flows?

An Information Flow has two major components

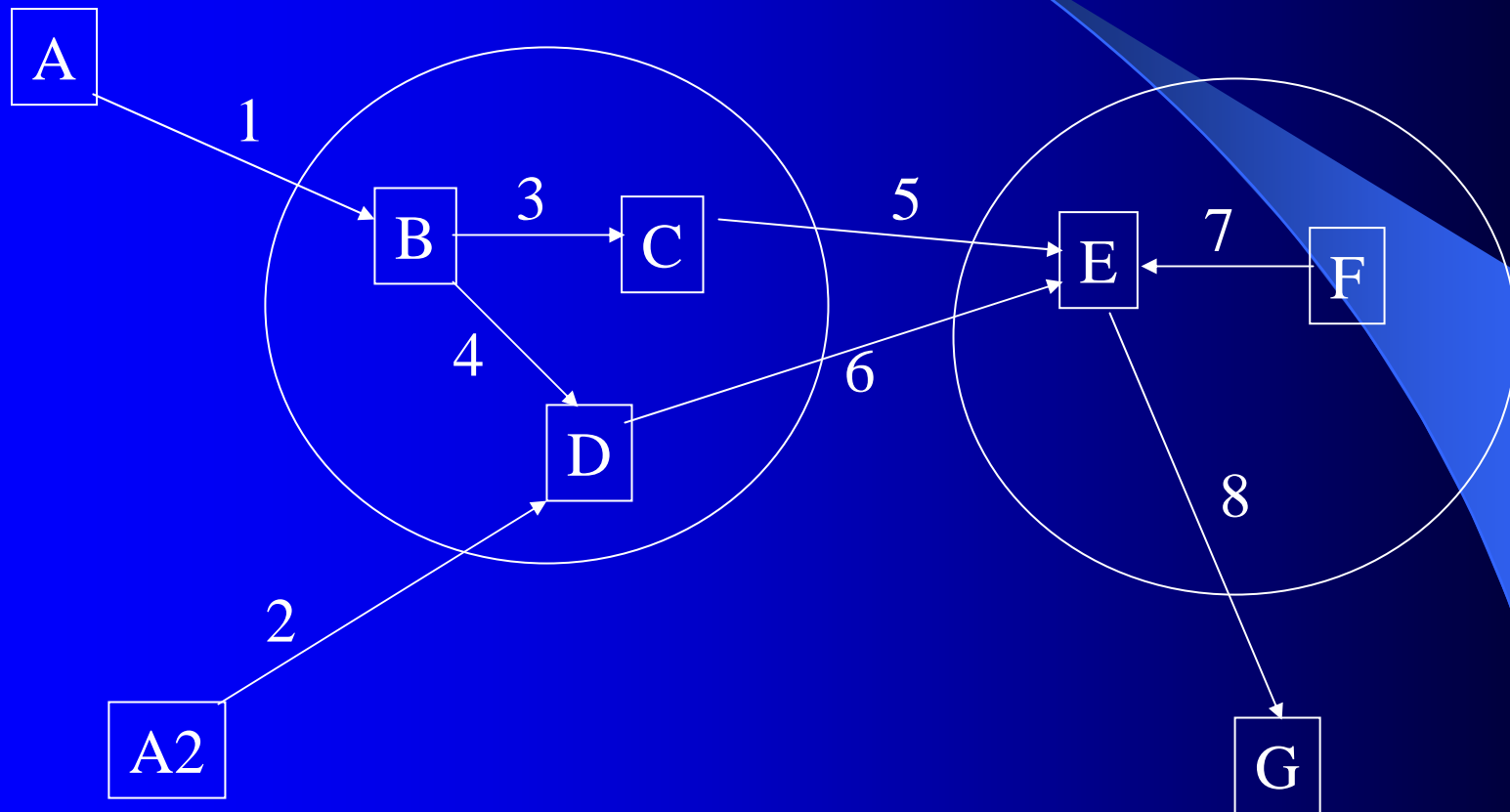
- Diagram(s) that depicts
  - Systems
  - Inputs to a systems from other internal systems
  - Outputs from a system to other internal systems
  - Inputs from the outside world
    - Human inputs
    - Automated feeds from external systems
  - Outputs to the outside world
    - Outputs for humans (reports, web pages, etc)
    - Automated feeds to external system

# What are Information Flows? (cont.)

- Plus all the associated metadata
  - Description of a system including
    - Including responsible business and IT organizations
  - Description of a flow
    - Data descriptions
    - Transformations
  - Responsible organization
    - Business organization
    - IT organization

Inputs and outputs from databases internal to a system are generally not shown

# Simple Information Flow Diagram



# Basic Info Flow Metadata

System Name	System Description

Flow Name	From System	To System

Flow Name	Data Group Name	Metadata for sender data	Metadata for receiver data	Data Transformation

# Information Flow Levels

- Business Level
  - Describes the business meaning of the system and data, business user, and business criticality
- Logical Level
  - Describes the entities and attributes, of the data passing from one system to another
- Physical / Implementation Level
  - Describes the layout of the data elements as well as communications link, volume, and frequency.

# Key Points

- Building information flows is relatively straight forward
  - Getting all the details correctly can be challenging
  - Eliminating inconsistencies is more challenging
- What is most important is to
  - Maintain the information flows as systems change
  - Use them appropriately (or they can become expensive wall paper)
- Tools can be invaluable.

# Initial Observations

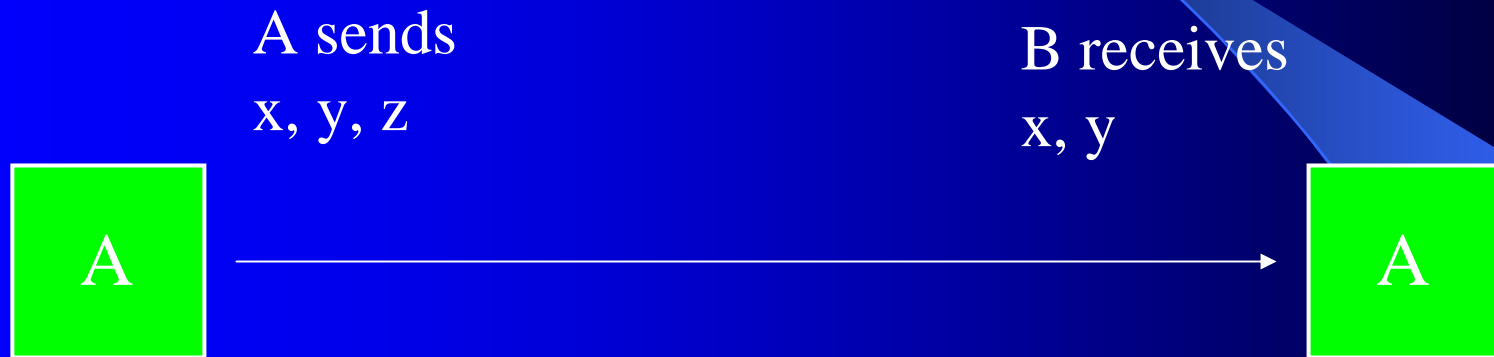
(or the documentation said that)

# Data Centric View

# Non-Matching Inputs & Outputs

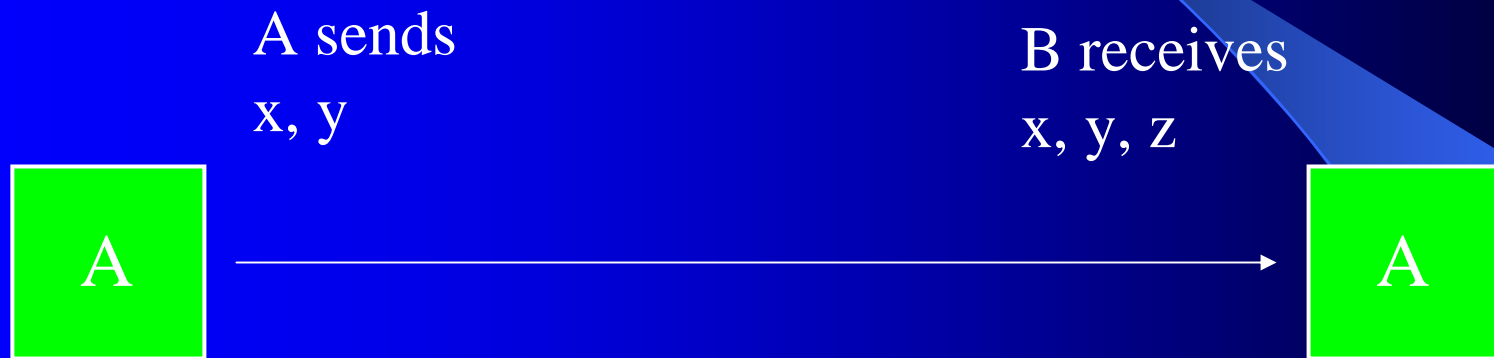
- Every arrow represents a reflexive relationship between the sender and the receiver
  - If A sends information to B then B receives information from A
  - What if the documentation of the two systems disagrees
- Compatible Definitions
  - Do both A and B agree on the definition of each data component of the flow
  - If A sends the same data to both B and C do all 3 agree
- Transformation Integrity
  - Does the transformation compromise the data integrity?

# Vanishing Data

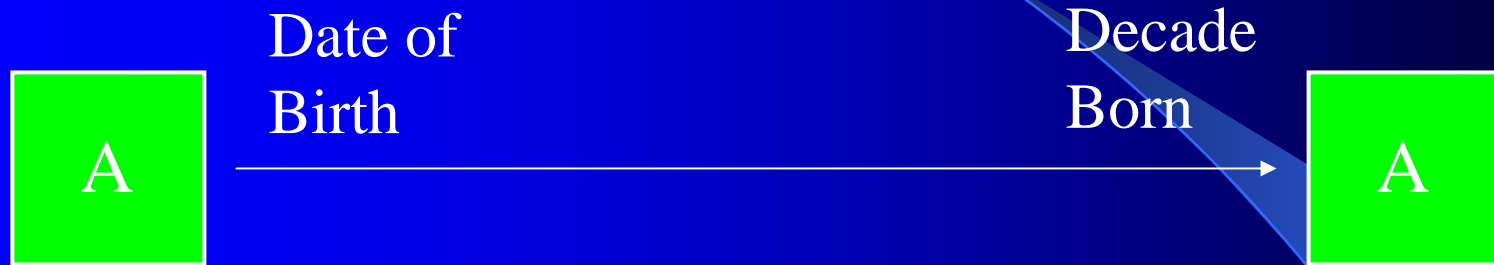


At least according  
to the documentation

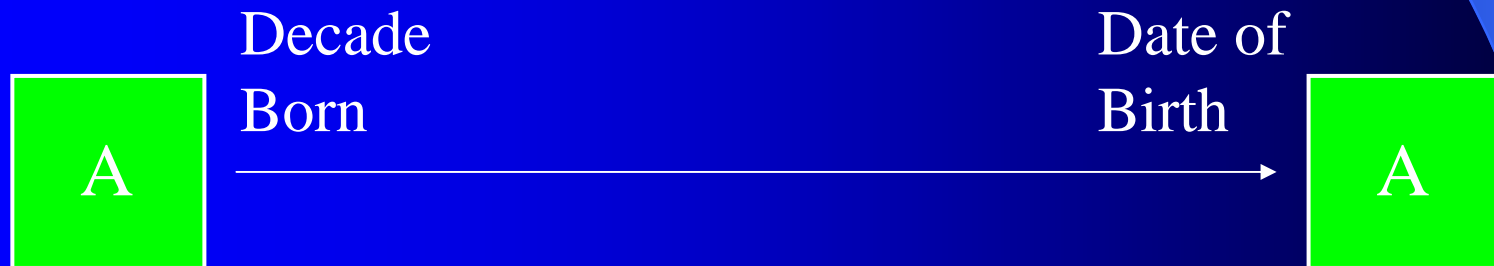
# Spontaneous Data Creation



# Transformations

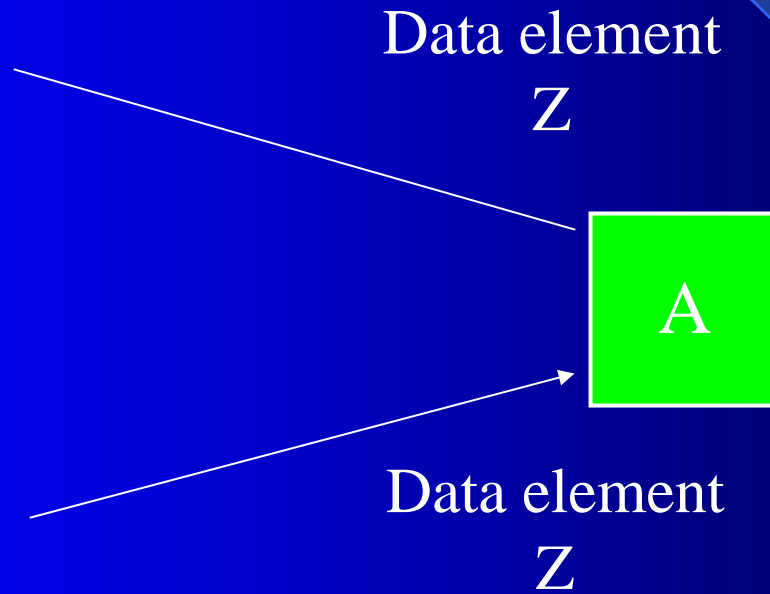


VS



If a transformation is not reversible  
be sure to ask more questions.

# Pick One



# Competing Inputs

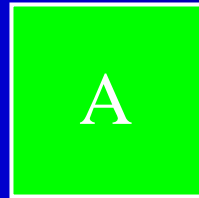
- If a system receives the “same” data from two other systems.
  - Is there one ultimate source of the data or two distinct sources?
    - If only one ultimate source, are the definitions and transformations on both value chains compatible?
  - Is this a case where the “same” data comes from one source or else the other source, but never from both (potential value chain problem)?
  - If two sources, how do we compare the data, and what do we do if they disagree?

# Murphy

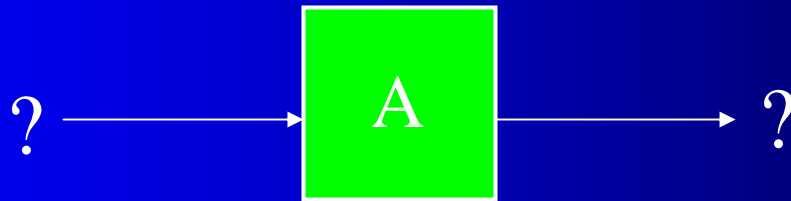
If you think you have seen it all, don't worry  
Someone out there will have yet another  
new case you have yet to see.

# System Centric View

# The Wall Flower

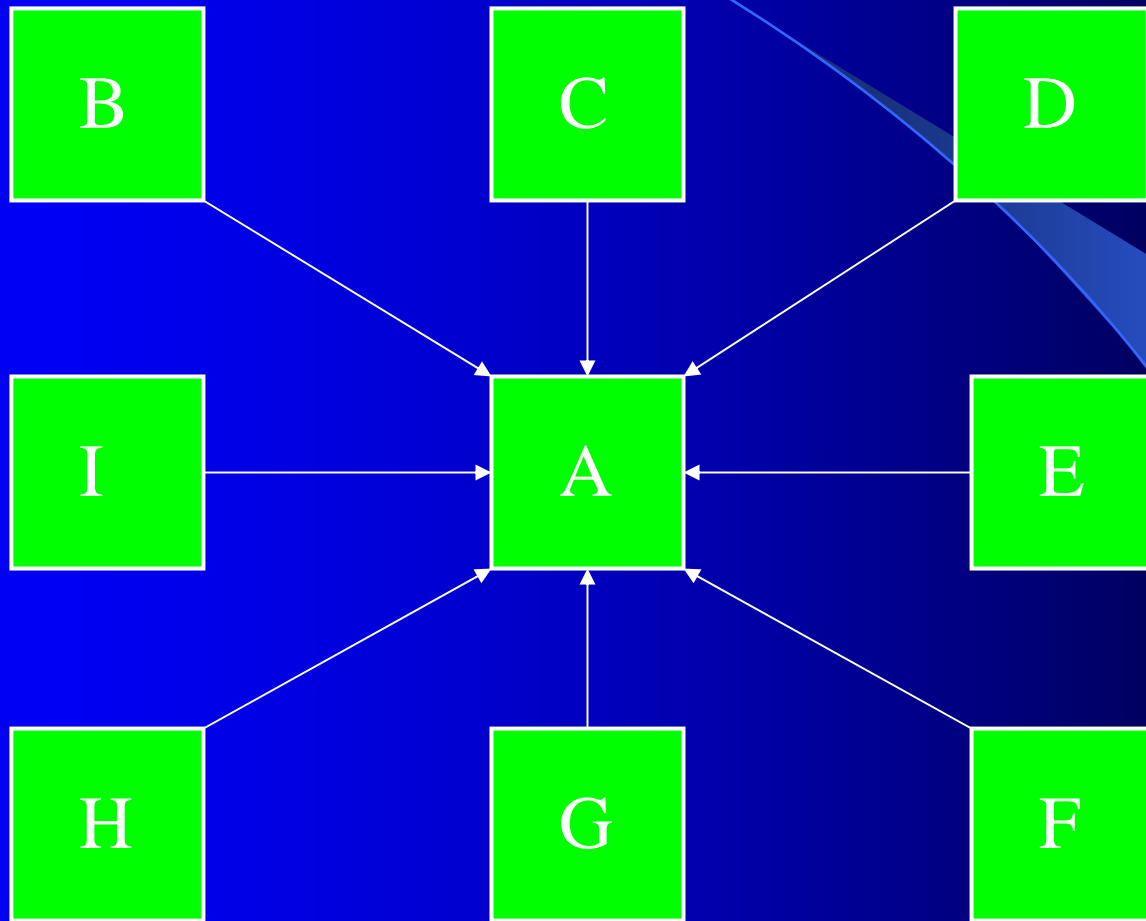


# Mystery Data



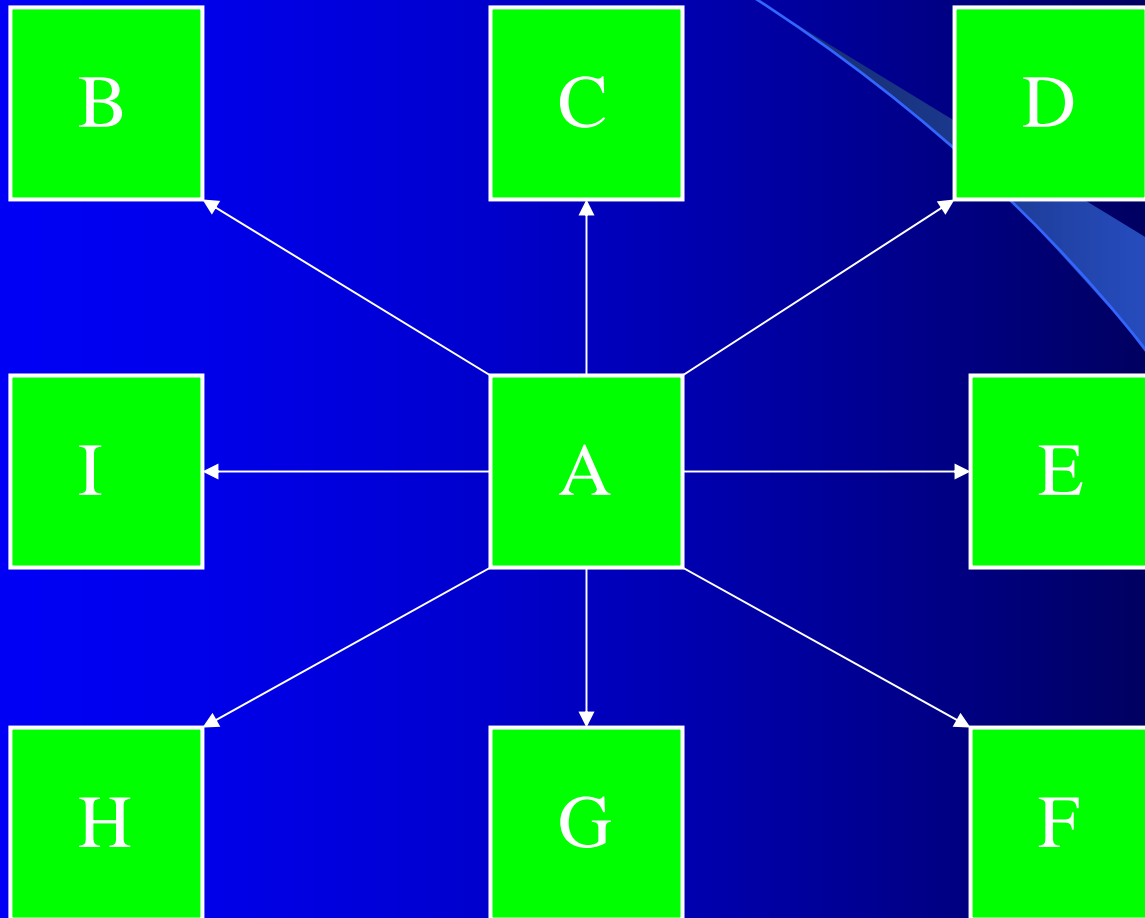
Documentation gives a source or target, but the system doesn't exist  
If we send data to another system, does it really have a receiver

# The Black Hole

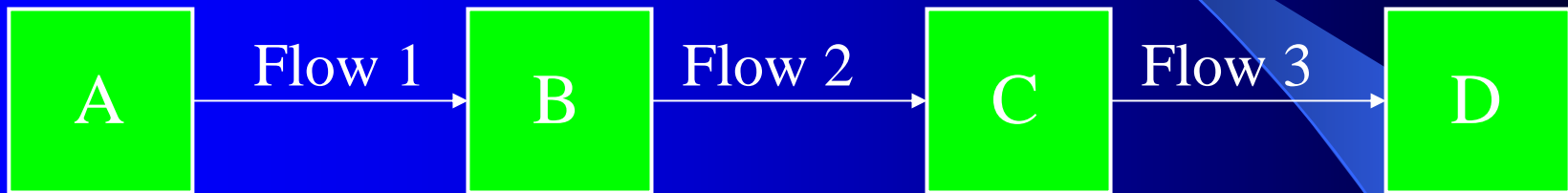


Alias the Data Prison

# Starburst



# High Tech Telephone Game



- What if the same data group appears on Flow 1 and also on Flow 2 and Flow 3, etc. ?
- What if B chances a piece of data sent by A, is A told (and so on down the chain) ?
- How do you spell SOX ?

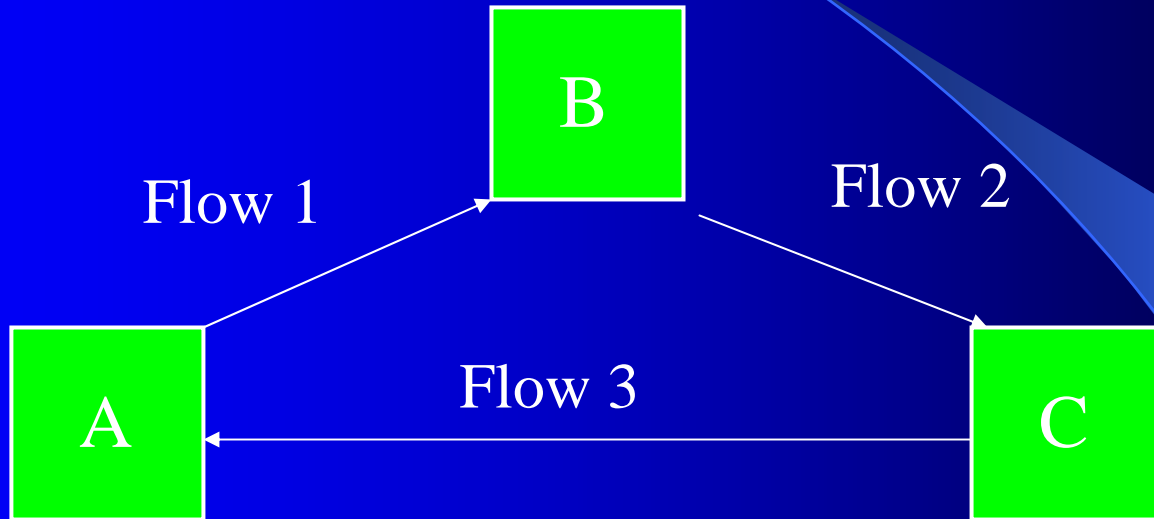
# Low Tech Telephone Game

System A outputs a report and a person enters data from the report into system B, etc., etc., etc.



This is the real telephone game

# Wheel of Karma



What happens if the same data group appears on all three Information Flows?

Is anyone in control?

# Organizing Information Flows

# Organizing Information Flows

- If there are a small number of systems, everything is depicted on one diagram.
- If there are a large number of systems, flows can be organized
  - By Organization
  - By Line of Business
  - By Business Function

# Look For

- The overall set of flows should be end-to-end from a business perspective
- Flows go from transaction systems to financial systems
- Flows that follow the CIF (or GIF)

# Major Sources & Sinks

- Only expect to see
  - a major source near the beginning of the business flow
  - A major sink near the end of the business flow
- Financial systems and Data Warehouses are typically major sinks as well as the source for reporting

# Data Transformation Questions

# First Question

- For each piece of data on the flow
  - Is it sent with no change or is there a transformation along the way?
  - Did the originating system transform the data before it was sent?
  - Did the receiving system transform the data after it was received.?
- If you answer yes, start analyzing

# Semantic Issues

- Do both systems have the same (or at least compatible) meaning for the data
- Do both systems use the data in the same way (context)
- If the data is a code value, do both systems have the same meaning

# Updates

- Does the receiving system change the data received
- If data is changed, is the sending system notified
- Does a change to the data have implications within the sending system

# Transformation

- Integers
- Decimal
- Date
- Character
  
- Issues
  - Precision (decrease or increase??)
  - Data type

# Using Information Flows to Reengineer Systems

# The Goal

- Understand the current “mess”
- Develop a strategy to rationalize the “mess”
- Reduce costs
- Figure out what are the candidates for a centralized DB
- Determine data lineage
  - Required for determining data quality
  - Required for audits and SOX

# Cost Reduction

- Gartner Group has estimated that maintenance of the bridges costs between 50 and 80% of the IT budget maintenance (building them is cheap!)
- Every data item on a flow is potentially replicated data which needs to be kept synchronized
- Every flow between two systems must be
  - Tested and/possibly
  - Modifiedwhen the system at either end is changed

# Data Dispersal

- If a data group (or element) is part of a flow, then at least two systems may maintain that data group (or element)
  - Data quality issues
    - Synchronization issues
    - Transformation issues
  - Data storage costs
  - Back up and recovery costs
  - Political costs
    - Is your data right or is mine right

# System Modification & Maintenance

- Knowing the Information Flows helps mitigate the risks of systems modification and maintenance
- Whenever any system is modified, the information flows give you a list of
  - “Adjacent” systems that may need modification
  - Data bridges, including their transformations, that may need modification
  - And everything needs to be tested.
- Guide for sequencing a series of changes

# COTS Products

- For a COTS product implementation the information flows indicate
  - Where the COTS product fits into the overall flow
  - The list of systems that need to interface with the COTS product
  - The data bridges that need to be built
  - The data transformations that need to be performed by the data bridge.
  - What about the next release of the COTS product

# Systems Replacement & Migration

- Whenever a system is replaced you know
  - All the data feeds to your replacement system
  - All the data feeds from your replacement system to other systems
- For systems migrated to another platform, every bridge must be sent across the network

(Risk Management)

# Data Quality

- To test data quality, one must know the full value chain for the data elements being tested.
- Information flows provide the
  - Original source for each data element
  - Chain of sources for use of the data
  - List of places where the data may have been changed or transformed.
- One must still go to the external sources for accuracy verification (usually statistical sampling)

# Information Flow Data Quality Example

- Researcher requested data about a sets of events from a federal agency that maintains that data
- Researcher searched agency's website for data regarding the same set of events and found the dates conflicted with dates provided to her directly from the agency
- Federal agency needed to find and correct data error
- Information Flow model revealed the problems and supported development of a remediation plan

# Observations

- Governance – Regulation of information flow and congruity - is a characteristic of information maturity. (MetaGroup) Without effective information and data governance, data quality suffers.
- Data Stewardship – Accountability and responsibility for information and data quality, content, security, and availability – will enable congruent information flows. Absence of data stewardship results in risk of conflicting, redundant, problematic information flows.

# “Core” Centralized Database

- A centralized database does not need to store every data item
- Only data items that are shared among the systems are prime candidates to be in the “core” centralized database.
  - Reduces cost
  - Improves performance
  - Possible easier buy in by owners

# Business Process Reengineering

- Information flows provide guidance for
  - Sequencing system changes or replacements
  - Testing
  - Reducing risk when making changes

# Concluding Remarks

Thank you

Questions