

# Information is not Data

**Curtis Pierce**

*Transit IT Program Manager*

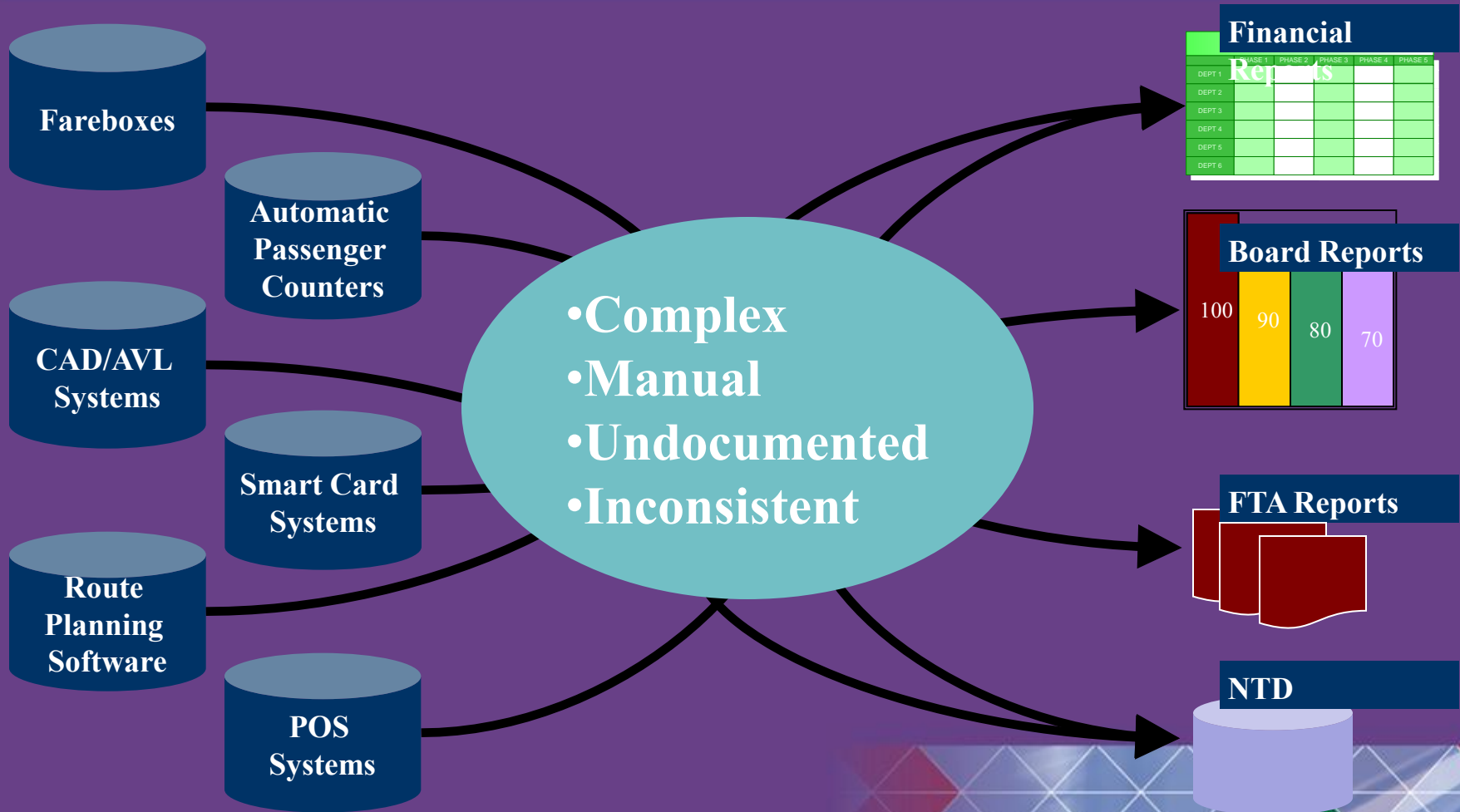
CH2M HILL

*San Francisco, CA*



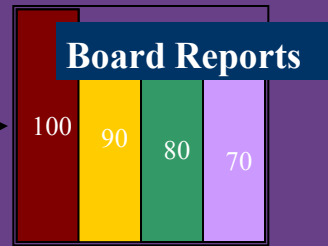
**2012 FARE COLLECTION WORKSHOP**

# Transit Data Streams are Varied, Fragmented and Incomplete



**Financial Reports**

	PHASE 1	PHASE 2	PHASE 3	PHASE 4	PHASE 5
DEPT 1					
DEPT 2					
DEPT 3					
DEPT 4					
DEPT 5					
DEPT 6					



# Data is Too Often Stovepiped

System	Data	Department
<ul style="list-style-type: none"><li>Fare Systems</li></ul>	<ul style="list-style-type: none"><li>Fares paid</li><li>Revenue collected</li></ul>	<ul style="list-style-type: none"><li>Revenue</li></ul>
<ul style="list-style-type: none"><li>Ticket Vending Machine (TVM)</li><li>Point of Sale (POS)</li></ul>	<ul style="list-style-type: none"><li>Ticket and pass sales by time and location</li></ul>	<ul style="list-style-type: none"><li>Revenue</li></ul>
<ul style="list-style-type: none"><li>Automatic People Counter (APC)</li></ul>	<ul style="list-style-type: none"><li>Boardings</li><li>Alightings</li></ul>	<ul style="list-style-type: none"><li>Service Planning</li></ul>
<ul style="list-style-type: none"><li>CAD/AVL</li><li>SCADA</li></ul>	<ul style="list-style-type: none"><li>vehicle location</li><li>route assignments</li><li>timing</li></ul>	<ul style="list-style-type: none"><li>Operations</li></ul>

# Other Common Data Issues

- Data is viewed as proprietary by system vendors
- Data is available but unused
- Data is not held
- Excel is not a system of record
- Data is mismatched
  - GIS Schema
  - Vehicle ID
  - Time Notation



# Opportunities in Data Integration

- Data richness – Two systems combined are more than the sum of the parts
- Rider information – Exposing your data can create a better riding experience
- Historical record – Planning and reporting can improve significantly
- One version of the truth

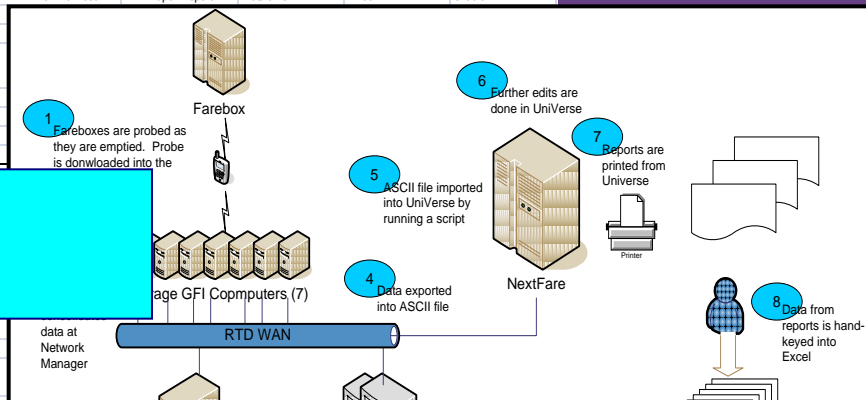


# Building a Data Integration Plan

Start by cataloging data sources and their usage

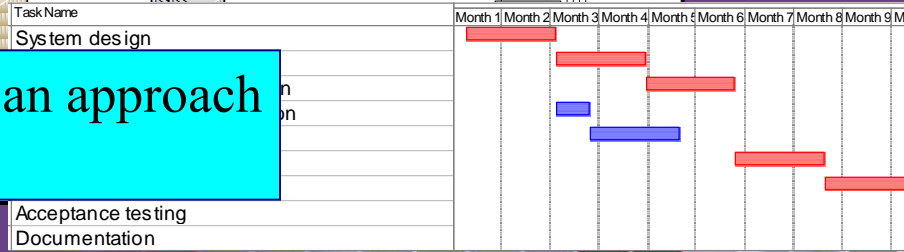
Reporting Category	RTD Service Name	NTD Mode Code	NTD Service Code	Source System	Source Data Type	Intermediate ETL	Intermediate System	Intermediate Data type	
	seniorShopper (Monday-Friday)	DR	DO	GFI Fareboxes	Sybase	Automated	RDR v. 1.0	Oracle	
	Bus (Local, Express, Regional)		DO	GFI Fareboxes	Sybase		RDR v. 1.0	Oracle	
	Mall Ride		DO	Init APCs	Database		RideCheck+	Oracle	
		MB	DO	GFI Fareboxes	Sybase		RDR v. 1.0	Oracle	
			PT	GFI Fareboxes	Sybase		RDR v. 1.0	Oracle	
			PT	GFI Fareboxes	Sybase		RDR v. 1.0	Oracle	
		LR	DO	Init APCs	Database		RideCheck+	Oracle	
		DR	DO	GFI Fareboxes	Database		RDR v. 1.0	Oracle	
				Driver Manifest	Paper Report		Mathews	Excel	Oracle
			DO						
	access-a-Ride (Paratransit)	DR	PT						
	Vanpool		PT						
	Seasonal Service (Sporting Events)	LR	DO						
		MB	DO						

Next Document the current process



Then develop an approach

Reporting Category	RTD Service Name	NTD Mode Code	NTD Service Code	Source System	Source Data Type	Intermediate ETL	Intermediate System	Intermediate Data type
Passenger Miles and Revenue Vehicle Miles	Misc., Shopper Special (Saturday)	DR	DO	data at Network Manager				
	seniorRide		DO					
	call-n-Ride	DR	DO					
	access-a-Ride (Paratransit)		PT					
	Vanpool		PT					
	Seasonal Service (Sporting Events)		PT					
Operators	Bus (Local, Express, Regional)							

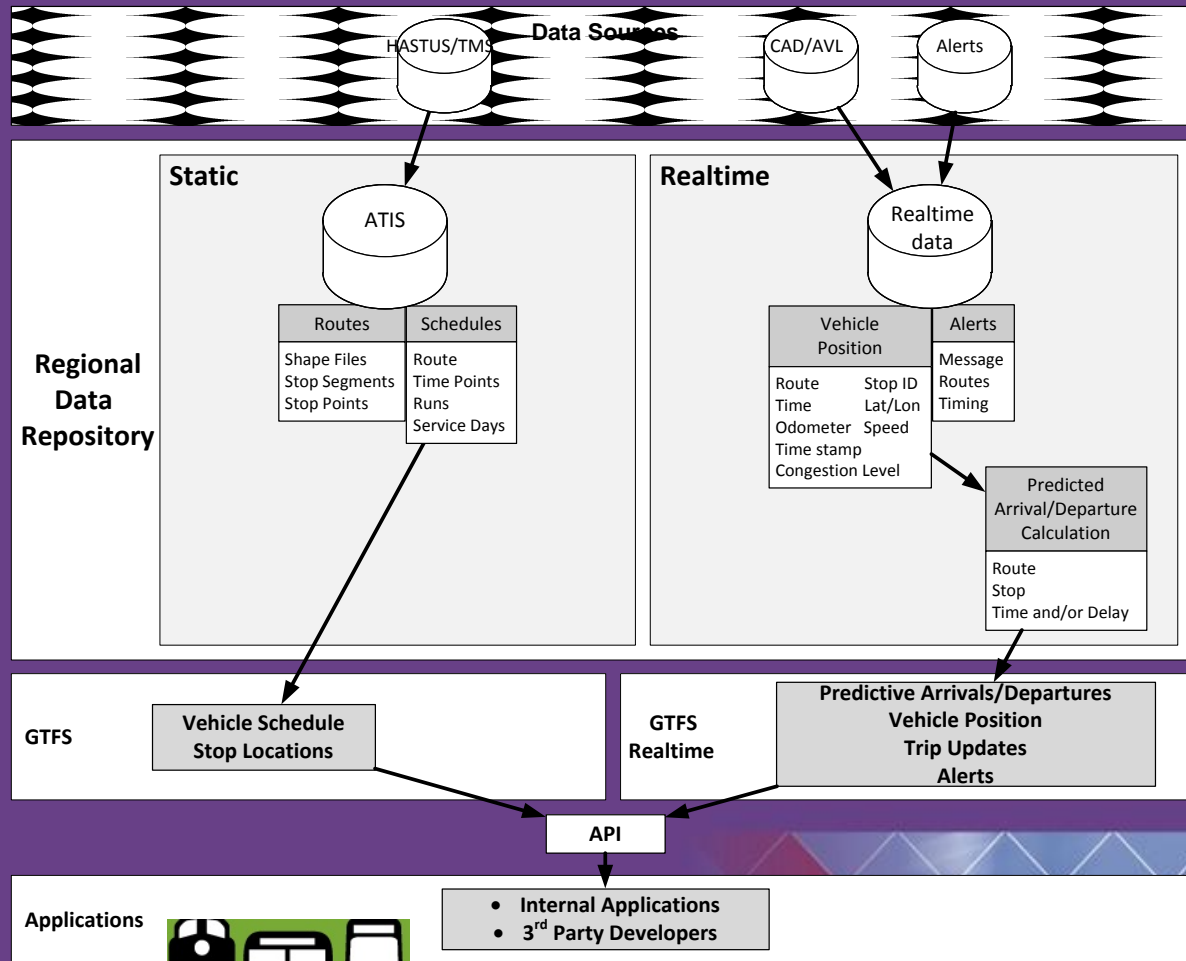


# Discovery

- It is critical to meet with as many stakeholders as possible
  - the view from above does not always match the view on the ground
- Inventory data sources
  - Where is the data generated?
  - In what form?
  - Where is it held and for how long?
- Map data uses
  - Who uses the data?
  - How do they manipulate it for reporting?
  - What other uses are possible?



# Design Data Flow at a High Level





# Executing a Data Integration Plan

- Design your execution around existing projects – incrementalism works
- Expose your data
  - Public agency = public data
  - Leverage outside developers
- Metrics are critical
  - How accurate
  - How manual
  - How timely



# Address Data in Procurements

- Avoid proprietary systems
  - Data not accessible
  - Data can't be merged with other data streams.
- Specify that data be held in an open database
  - Published schema
  - Access through standard tools



# Better Data Leads to Better Operations



Integrated data will be more accurate and more trusted throughout the agency



# ThankYou

**Curtis Pierce**

*Transit IT Program Manager*

CH2M HILL

*San Francisco, CA*

*Curtis.Pierce@CH2M.com*

*510.541.2799*

