

# Data Migration for the BACS Infrastructure Renewal

Data Management Specialist Group: 30<sup>th</sup> November 2011

**Vince Burr**

VocaLink Consultant

[vince.burr@vocalink.com](mailto:vince.burr@vocalink.com)



**VOCALINK**  
safer payments, smarter partner

# Agenda

- Introduction and Background
- Context
  - The Programme
  - The Challenge
  - The Outcome
- Strategy
  - Identify Data Sources
  - Understand relationship with Implementation
  - Analyse each data source
  - Migration Process
- Data Cleansing
- Summary
- Q&A

# Introduction

- Vince Burr
- Worked in IT since graduation (1982), predominantly in the Payments Industry
- Roles: Programme, Application Design, Business and Systems Analysis, Project Manager, Consultancy
- Currently working in the VocaLink consulting team, with a focus on regulatory change in the market, Cash ISA Transfers, and Current Account Switching.

# Context – The Programme

- Bacs Renewal (2002-2006)
- A major UK wide programme of change to refresh the Bacs infrastructure.
- Dimensions:
  - 13 Member Banks, up to 500 Agency Banks
  - Circa 100,000 Corporate and Government Users
  - Circa 500 Service Bureaux
  - Circa 20 solution suppliers
  - 6 billion payments a year, 100 million payments on a peak day
  - £4 trillion per annum by value
- Four releases:
  - Release 1: Channels and Security
  - Release 2: Reference Data
  - **Release 3: Payment processing and exception management**
  - Release 4: Non-payment Messaging services
- Today, I look at the Data Migration strategy for release 3, and also look at a specific issue from release 2.

# Context – The Challenge

- How to achieve the implementation at a low level of risk
- Principle risks were:
  - Software failure
  - Failure to provide same service level on the Monday as was provided on the Friday
  - Catastrophic failure – extended loss of service
  - Data Failure – incorrect or missing payment
  - Settlement failure – incorrect settlement positions
- The potential impact of worse case scenarios was extremely high – hard even to reliably estimate

# Context – The Outcome

- The risks were mitigated to a point where they were no higher than BAU operation of the old services.
- The success of the strategy meant that the Go/no-go decision taken by the Industry prior to cutover was a non-event, with no issues, concerns or qualms expressed.
- The outcome was a non-event – service continued as before.
- Since implementation some 30 billion payments have been processed

# Methodology – Define your strategy

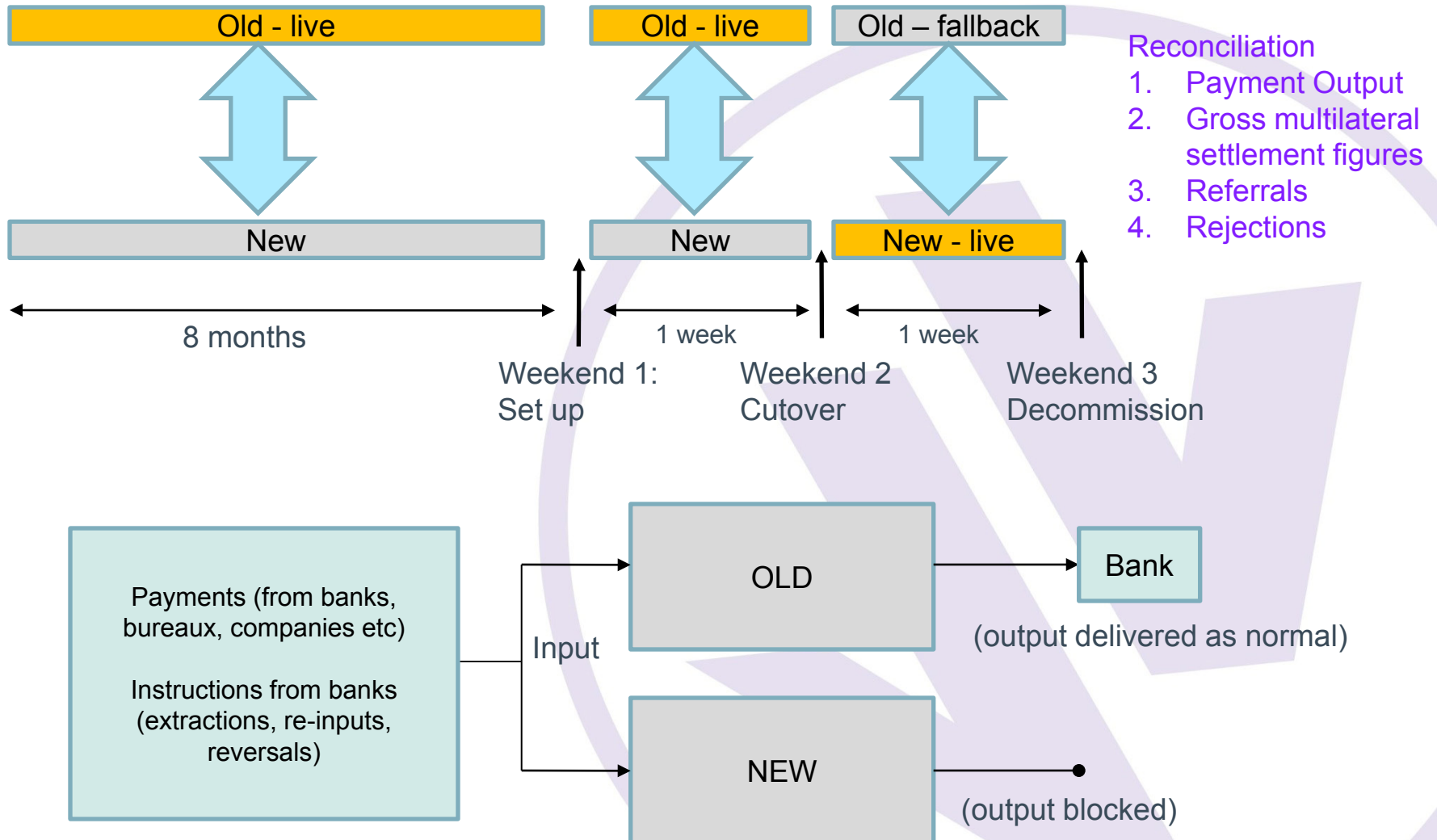
1. Define data sources in old system(s) that are potentially relevant for new system(s)
2. Understand synergy / relationship between Implementation and Data Migration
3. Analyse: determine requirements in relationship to each data source:
  - Urgency / Priority
  - Usage
  - Data Volatility
  - Timing (earliest availability / latest need-by)
  - Qualitative verification
  - Quantitative Verification
4. Define Data Migration process

# (1) Data Sources

- Future Dated Payments (payments in Bacs for clearing on a future date) – 5m to 70m payments
- Referrals (historic conditions referred to Banks relating to credit risk or alerting of possible customer error) – 13 month history – 500,000 records
- Submission History – historic payment and payment profile data used by Banks in their analysis of referrals – 13 months history – 10m records
- 6 day payment history – payments up to 6 days old may be “reversed” to back out errors – 300m to 800m records
- Advanced Warnings and Monitors – alert conditions set by banks against customers – 100s
- Limit History – analysis by customer of near or actual credit limit breach – used by banks to maintain sensible credit limits



## (2) Methodology – Implementation Synergy



## (3) Methodology – Analyse

- **Future Dated Payments** (payments in Bacs for clearing held for a future date)
  - 5m to 70m payments
  - Critical / high priority
  - Customers submit payments in advance of payment day – notably for end of month (salary, direct debit)
  - Volumes increase through the month, up to month end, when they are flushed out over a peak period of processing, and reduce to a much lower number for next month start.
  - Earliest availability: end of day Friday prior to weekend 1
  - Required by: Start of day Monday
  - Quantitative verification: Confirm correct numbers and aggregate values migrated, broken down by Service User
  - Qualitative verification: via Parallel Run

## (3) Methodology – Analyse

- **Referrals** (historic conditions referred to Banks relating to credit risk or alerting of possible customer error)
  - 13 month history, 500,000 records
  - Critical / high priority
  - Banks look at past referral conditions, and use previous responses to help direct how they should respond this time, most notably if they cannot contact the customer.
  - Referrals (and responses) are static as soon as they are written to the database.
  - Earliest availability: end of day Friday prior to weekend 1
  - Required by: Start of day Monday
  - However, the workflow implementation for new and old diverged to such an extent, that the data structures were infeasible to migrate reliably (old system had a lower level of data retention)
  - Strategy: Replicate old system screens (content and interaction model) on top of an access database, and circulate static copy to each bank containing only their customers data. Distribute on physical media over weekend 1.
  - Quantitative / Qualitative verification: Normal testing stages, plus trial copies to banks for acceptance.

## (3) Methodology – Analyse

- **Submission History** – historic payment and payment profile data used by Banks in their analysis of referrals – 13 months history – 10m records
  - Critical / high priority
  - Static as each day completes
  - Volumes remain relatively stable.
  - Earliest availability: 13 months prior to weekend 1
  - Required by: Start of day Monday
  - Infeasible to migrate over a weekend
  - Strategy:
    - Extract and Transform in units of 1 month in the month prior to weekend 1
    - Final month extracted at end of Friday prior to weekend 1
    - Load during weekend 1
  - Quantitative verification: Confirm correct numbers and aggregate values migrated, broken down by submission, file, day section, account section aggregations.
  - Qualitative verification: Sampling during Parallel Run, and sampling over the weekend 1. Note. Bank staff engaged on site at VocaLink at key times for this purpose.

## (3) Methodology – Analyse

- **6 day payment history** – payments up to 6 days old may be “reversed” to back out errors – 300m to 800m records
  - 300m to 800m payments; highest at month start (due to previous end of month peak). Note how this the reverse of Future dated payments profile.
  - Critical / high priority
  - Earliest availability: end of day Friday prior to weekend 1
  - Required by: Start of day Monday.
  - Reversals are unusual events – perhaps 10 or so a month.
  - Strategy:
    - Concluded not possible to migrate in time during weekend 1 (elapsed time estimate 20 to 40 hours).
    - Optimal migration timing conflicts with future dated payments migration.
    - “On-demand” migration – special procedure to migrate ONLY those transactions required to be reversed in the event that they need to be reversed.
    - Only available during week 1, and Monday of week 2.
  - Quantitative verification: Confirm correct numbers and aggregate values migrated.
  - Qualitative verification: via Parallel Run

## (3) Methodology – Analyse

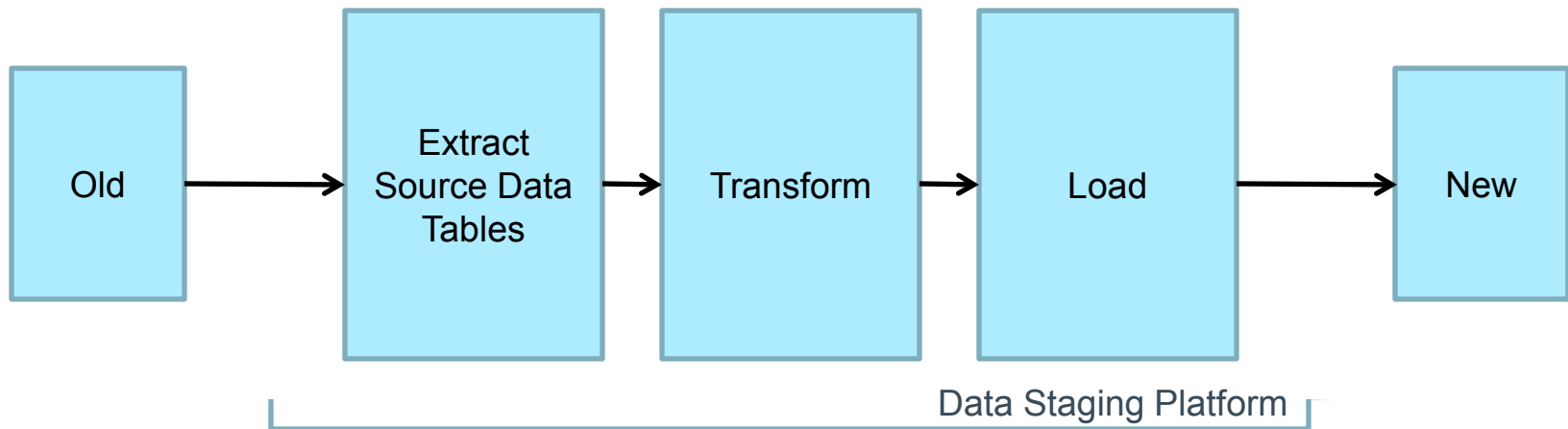
- **Advanced Warnings and Monitors** – alert conditions set by banks against customers – 100s
  - Medium/high priority
  - Low enough volume, such that banks could manually set them up again.
  - Advanced warning – associates a message with a future event, e.g. “please ring Mr Smith at British Gas, on 0293 456 7869, when direct debit file received”.
  - Monitors – set against companies that are a credit risk (e.g. Imminent bankruptcy), so that credit files received will generate a referral to the bank.
  - Volumes static
  - Earliest availability: end of day Friday prior to weekend 1
  - Required by: Start of day Monday
  - Quantitative verification: Confirm correct numbers.
  - Qualitative verification: via Parallel Run, bank sampling weekend 1.

## (3) Methodology – Analyse

- **Limit History** – analysis by customer of near or actual credit limit breach – used by banks to maintain sensible credit limits
  - Circa 100,000 sets of data, showing accrual of payments against daily/weekly/monthly/periodic limits
  - Critical / high priority
  - Earliest availability: end of day Friday prior to weekend 1
  - Required by: Start of day Monday
  - Quantitative verification: Confirm correct number of records migrated.
  - Qualitative verification: via Parallel Run

## (4) Migration Process (1)

- Requirements of the process:
  - Robust and fail safe
  - Must decouple from old system as soon as data is captured
  - Reconcilable; assurance of no data loss
  - Validate “old” data against “new” system rules; may result in data cleansing activity (see later slide)
  - Audit trail; support retrospective analysis (for defects), and data corrections.
  - Must meet performance requirements



Extract: Load raw data.  
No data manipulation.  
Minimises possibility of errors at this stage.

Transform: Need a mapping (see next slide)

Load tables: same definition as “New”, again no data manipulation.



## (4) Migration Process (2) - Mapping

- Can be the most challenging part.
- May not be possible to complete until “new” service has a final database design.
- A cross discipline exercise. Expertise to cover new and old implementation, business and perhaps operational (dedicated migration team).
- Physical inspection of real data supports the process.
- We used a simple excel spreadsheet:
  - Defines source tables and attributes.
  - Define target tables and attributes.
  - Relate the two.
  - Specify how to deal with:
    - Lack of corresponding attributes in old (generation / defaulting rules)
    - Lack of corresponding attributes in new (does it matter?)
    - Matching attributes, but mismatching definitions(e.g. allowable enumerations may differ.
  - Opportunity to validate that nothing is missing from new service!!
- Lets look at an example!!!



# (4) Migration Process (2) – Mapping Example

REFT	BCOLL	PE	Comment
RnnnnSPLDXVOLUME	R1117XSUBMISSION	SUBMISSION	
SPLD-VOL-JUL-DAY	QSP0-REFERENCE	Replaced by URN	GENERATED
SPLD-VOL-SEQ-NUM			
CHAR-CODE	None	None	REFT Media Handling
DECK-NUM	None	None	REFT Media Handling
REC-MODE-CODE	None	None	REFT Media Handling
VOL1-LABEL: vol-srl-num	VSN	SERIAL_NUM	REQUIRED
VOL1-LABEL: vol1-ax-flag	None	None	Not required
VOL1-LABEL: vol1-seq-num	None	None	Not required
VOL1-SUBMITTER-ID	USER-BUREAU-ID	SUBMITTER	REQUIRED
VOL1-LABEL: vol1-jul-day	None	None	Not required
TEL-U-B-SBM-FILE-CNT	NUMBER-OF-FILES	None	Used to record number of files received within a single BACSTEL submission.
TEL-U-B-SBM-FILE-RCV-CNT	None	None	
SPLD-VOL-IN-USE-IND	None	None	REFT implementation feature.
TYPE-OF-MED	MEDIA-TYPE	INPUT_CHANNEL	REQUIRED
SUB-TYPE-OF-MED	None		
TYPE-1-ERR-IND	VOLUME-STATUS	STATUS	REQUIRED
FILE-CNT	None		REFT – Instructions optimisation.
EARL-DAY-SEC-DATE	None		REFT – Instructions optimisation.
LAT-DAY-SEC-DATE	LATEST-PROC-DATE		REFT – Instructions optimisation.
CR-DAY-SEC-CNT	None	<i>Depends on Intervention Instructions State Model.</i>	REFT – Instructions optimisation.
DR-DAY-SEC-CNT	None		REFT – Instructions optimisation.
CR-AC-SEC-CNT	None		REFT – Instructions optimisation.
DR-AC-SEC-CNT	None		REFT – Instructions optimisation.
REV-CR-IND	None		REFT – Instructions optimisation.
REV-DR-IND	None		REFT – Instructions optimisation.
DATE-STARTED	None	None	
TIME-STARTED	None	None	
DATE-ENDED	None	None	Start/End date/time for SPOOLING
TIME-ENDED	None	None	
-	EXTRACT-IND	<i>Depends on Intervention Instructions State Model.</i>	Records that Submission contains extracted data.
-	SPOO-BAT-NUM	None	REFT batch – redundant
-	TCOM-SBM-NUM	None	BACSTEL batch - redundant
-	-	PROCESSING_STAGE	GENERATED
-	-	TEST_STATUS	GENERATED
R3320VALXPDY	-	ARRIVAL DATE	REQUIRED

# Data Cleansing (Release 2, Reference Data)

- Data Migrations so far relate to data created in a consistent and uniform manner by a batch application. Also helped as both new and old systems were processing identical messages as input (i.e. Bacs payments).
- This isn't always the case. The Bacs service holds reference data that records information about banks, bank branches and also banks corporate customers. This could not be migrated without first being "cleansed":
  - User Interfaces often allow a degree of variability as to the data that they allow to be recorded on the database. A good example is address information.
  - In the "old" service, Banks maintained their customer information locally, and provided VocaLink with a copy on a daily basis. In the "new" service, Banks maintain their customer information at VocaLink, and a copy is provided to the Banks when/as required.
  - The Bacs systems, over the years, had fragmented and duplicated reference data information, leading to inconsistency and data contradiction.
  - The "old" reference data did not comply with the more stringent integrity rules of the new service, and these errors could not be corrected in-flight as part of data migration.

# Data Cleansing (Release 2, Reference Data)

- **Solution:** Data Cleansing Process.
- Approach:
  1. Migrate Data (as per process defined previously)
  2. Test New UI or batch processes against migrated data
  3. Analyse problems that arise out of steps 1 and 2
  4. Agree Data Cleansing steps (e.g. standardise address representation)
  5. Repeat from step 1, until no further issues
  6. Final confidence test shortly prior to implementation

For Release 2, this process took some 6 months to complete.

# Summary

- Invest time in the strategy. Define and agree a clear strategy; consider what, when, how. Be creative, consider,
  - Direct Migration
  - Migration in advance
  - Migration just in time
  - Migration on-demand / partial migration
  - Not migrating – provide the information in some other way
  - Do nothing (always an option)
  - Take into account the Implementation strategy. In my case study, the shape of implementation constrained data migration. In other cases, the reverse may occur.
- Define a migration process that is,
  - Fail safe and resilient
  - Decouples from the old system
  - Provides an audit trail, and supports diagnosis and correction.
- Mapping – is the key detailed specification
  - Take into account missing or mismatched mappings
  - Remember: Proving your new systems works against migrated data is different to proving in works against representative combination of “normal” inputs.
  - Ensure all mappings are externalised (not hidden in toolsets), and are validated by multi-disciplinary teams.
- Data that cannot be mapped may need a “data cleansing” strategy.
- Use a dedicated team.



# Thank You, Questions welcome

Follow up queries welcome at:  
[vince.burr@vocalink.com](mailto:vince.burr@vocalink.com)

+44 (0)870 920 8232 (direct)  
+44 (0)779 201 7251 (mobile)



**VOCALINK**

safer payments, smarter partner