



Inland Revenue

Business Transformation Programme

Data Conversion, Data Cleansing & Data Enrichment Approach Release 3

Prepared by:
Date:

Data Lead – Business Data Stream
6 July 2018

About this Document

This document explains the overall data conversion, data cleansing and data enrichment approach that will be taken for the conversion of data for Release 3 as approved by the Data Owners Forum.

Document Control

File Name and Path	
Contact Person	
Status	Final
Version	1.0
BT UiD	10825

Document Signoff

Formal Review Area	Name	Signature	Date
Responsible person – work stream 1			
Responsible person – work stream 2			
Responsible people – work stream 3			
Responsible person – work stream 4			
Accountable Person			
The following people have supported the development of this document:			
The following people and groups have been consulted:	Data Owners Forum		
The following people and groups have been informed of the development and completion of this document.	Data Owners Forum Design Authority Transformation Executive Working Group		

Contents

1 Executive Summary 5

2 Data Cleansing Approach 7

2.1 Introduction 7

2.2 Roles and Responsibilities..... 7

2.2.1 Business Data Stream Team roles..... 7

2.2.2 Business Data Stream Team Responsibilities 7

2.2.2.1 Work stream 1 - Everyday BAU data cleansing carried out by CCSI and CCSB . 7

2.2.2.2 Work stream 2 – Known R3 data issues - data cleansing work carried out by CCSI, CCSB & IT&C..... 8

2.2.2.3 Work stream 3 – Converted data cleansing activity 8

2.2.2.4 Work stream 4 – Data Enrichment 9

2.2.3 Business Transformation Managers 10

2.2.4 Data Stewards..... 10

2.2.5 Business Stakeholders..... 10

2.2.6 Data Owners Forum 10

2.3 Data Cleansing Approach 12

2.4 Process for Data Cleansing / Remediation for work stream 3 12

2.5 Data Cleansing Guiding Principles..... 13

2.6 Lessons Learned from Release 2 13

3 Data Conversion Methodology and Approach 14

3.1 Work stream 3 – Converted data cleansing activity..... 14

3.2 Inventory Data Resources 14

3.2.1 Creating the Inventory 15

3.2.2 Review the Inventory for Integrity and Quality 15

3.3 Data Purification 15

3.4 Detailed Conversion Plan..... 16

3.4.1 Approach to Historic Data 16

3.4.2 Types of Data to be converted..... 16

3.4.3 Establish a System of Record 16

3.4.4 Live vs. FIRST Reference 16

3.4.5 Approach to Selection Criteria 17

3.4.5.1 Impact on Daily Operations..... 17

3.4.5.2 Impact on Planned System Processes..... 17

3.4.5.3 Impact on Data Purification..... 17

3.4.5.4 Impact on Cutover Timeline 17

3.4.6 Automatic vs. Manual Processes 17

3.4.7 Approach to Work in Progress..... 18

3.5	Conversion Process.....	18
3.5.1	Extract Data from FIRST Applications	18
3.5.2	Translation of Extracted Data from FIRST Applications.....	19
3.5.2.1	Mapping and Reformatting vs. Transforming Data	19
3.5.2.2	Avoid Production Copies	19
3.5.2.3	Reduce Risk by Reducing the Toolset.....	19
3.5.2.4	Leverage IR Expertise	19
3.5.3	Loading Data into START	19
3.5.3.1	Validate Transformed Data.....	19
3.5.3.2	Customer is the logical Unit of Work.....	19
3.5.3.3	Use Standard Business Objects.....	19
3.5.3.4	Suspending Some Core Functionality	20
3.5.4	Reconciliation	20
3.5.4.1	FIRST Report.....	20
3.5.5	User Verification	20
3.5.6	Mock Conversion Runs.....	20
3.5.7	Business Function Definitions	21
3.6	System Cutover	21
3.6.1	Cut-off for Processing Items in FIRST Systems	21
3.6.2	FIRST Enters a Read-Only State	22
3.6.3	Final Conversion Process.....	22
3.6.3.1	Suspend Maintenance.....	22
3.6.4	Going live	22
4	Data Enrichment Approach	22
5	Reporting	23
	Appendix A – Business Data Stream Team – The 4 Work streams	23
	Appendix B – High Level Data Cleansing Plan/Timeline	25
	Appendix G – Definitions	26

1 Executive Summary

The primary goal of the data conversion and cleansing approach is to deliver the data conversion and necessary data cleansing activities required to successfully deliver Release 3 and enable the realisation of benefits from the Business Transformation Programme.

Three main issues are top of mind. They are: size – R3 is substantially larger than R1 or 2; complexity – this is significantly higher; and timeframe – ensure data is in an acceptable condition to enable a clean conversion from FIRST to START in April 2019.

A successful Go Live will require a collaborative and coordinated approach to data cleansing across all of IR. We propose categorising the work into the following 4 work streams:

1. *Everyday BAU Data cleansing carried out by CCSI and CCSB & IT&C*
2. *Known data issues - data cleansing work carried out by CCSI, CCSB & IT&C - driven by the Business Data Stream Team*
3. *Converted Data cleansing activity - driven by the Data Conversion Team*
4. *Data Enrichment to enable Transformation benefits realisation – driven by Business Lead, Data Enrichment*

Section 2 does contain some release 3 specific data issues as well as the general guidelines and approach for data cleansing for work streams 1 & 2.

Section 3 is not release specific. It contains the general guidelines and principles of the conversion methodology and approach for work stream 3. It is not intended to identify release specific risks or data quality issues.

Section 4 is new for Release 3 and contains high level general guidelines and principles of our approach to data enrichment which are still being developed.

At this early stage the known data issues that have been prioritised as “Must Do”, “Should Do” and “Could Do” in the 4 work streams number as shown here:

Data Cleansing via BAU (WS1)		BDST Initiated BAU Work (WS2)		Converted Data Cleansing (WS3)		Data Enrichment (WS4)	
Must Do	Should Do	Must Do	Should Do	Must Do	Should Do	Must Do	Should Do
 4	 8	0	 4	 14	 10	 2	 5
Could Do	Won't Do	Could Do	Won't Do	Could Do	Won't Do	Could Do	Won't Do
 4		 4		 9		 5	

Key

“Must Do” – critical for current delivery timeframe.

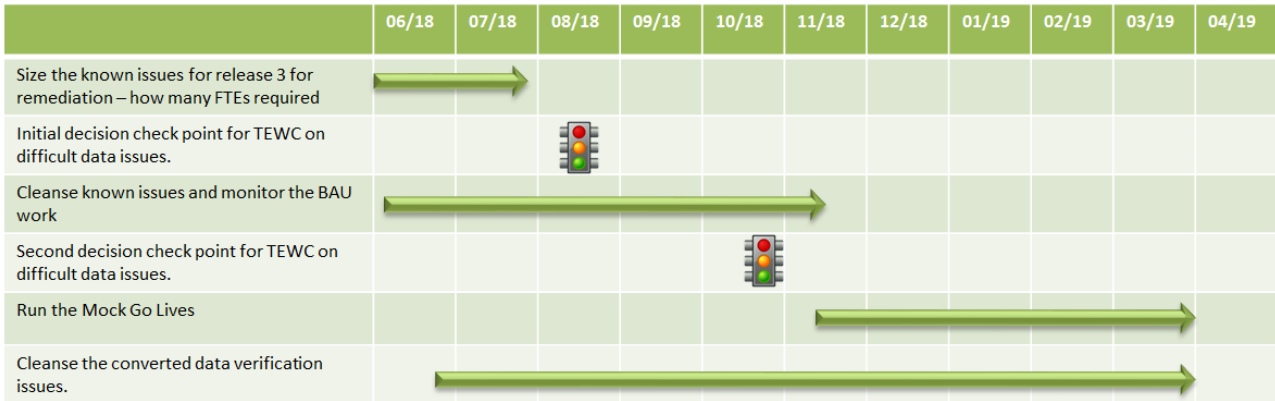
“Should Do” – important, we will attempt to deliver as many as possible in the current timeframe.

“Could Do” – desirable but not necessary for conversion.

“Won't Do” – least critical and not appropriate at this time. Currently there are no issues in this category but it may be that decisions are required at the CIR/ELT decision checkpoints proposed in the high level action plan to move issues into this category.

Transparency of the state of data issues will be vital at the Commissioner and Executive Leadership levels to enable important conversations and making difficult decisions such as moving issues into the “won’t do” category. At the same time a decision would need to be made on how to manage those issues once they are converted to START so that START does not inherit issues from FIRST that are then not resolved.

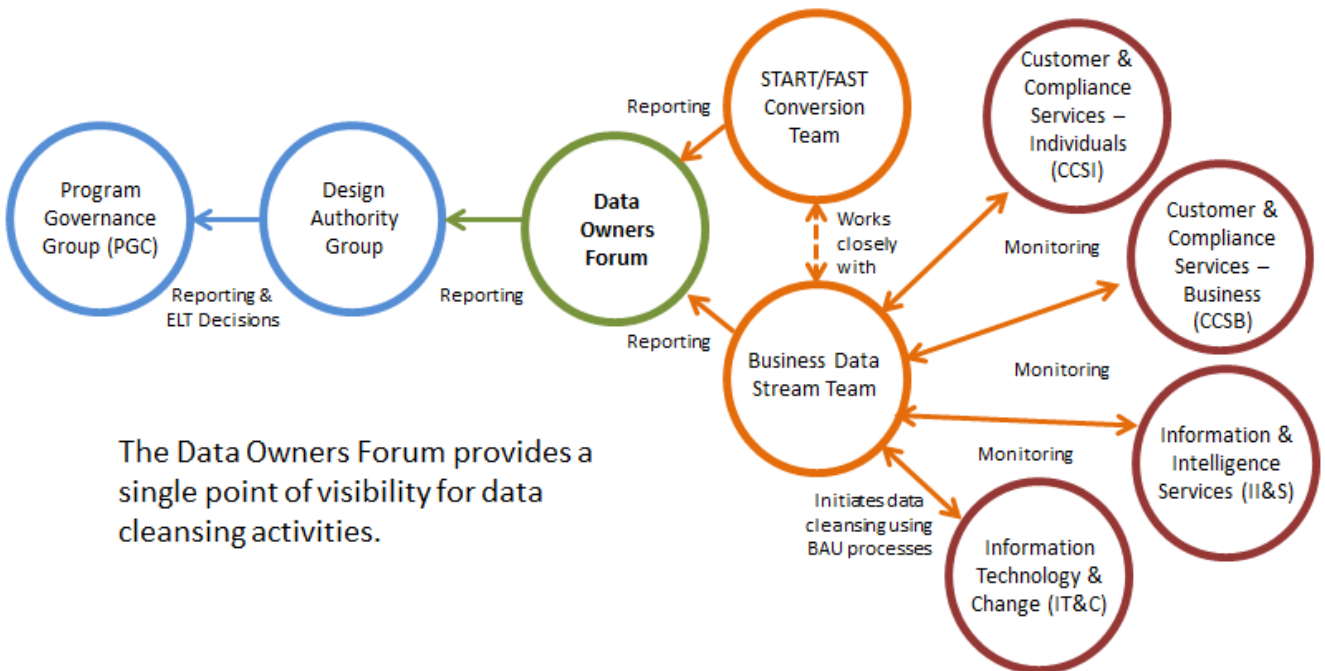
To enable that the following high level action plan is proposed:



- Questions to be asked at Data Owners Forum with decisions to be made at each checkpoint**
- Do we have the resource to resolve these issues within the given time frame?
 - Can we resolve these issues within the given time frame?
 - Should the data be locked down and not converted to START?
 - Which data issues can we not resolve? How are we going to treat them?
 - Are there any decisions required at a TEWC level?
 - Do those decisions need communicating to the Minister?

Successful delivery of these work streams requires a partnership approach with support given at a Deputy Commissioner level as well as at Customer Segment Lead, Segment Management Lead, Group Lead and Team Lead levels across CCSI, CCSB & IT&C.

A single point of visibility and governance for all data issues is a necessity. To enable this we recommend the following communication/reporting approach:



The Data Owners Forum provides a single point of visibility for data cleansing activities.



2 Data Cleansing Approach

The accountability for Data Cleansing is with the Business Data Stream Team.

2.1 Introduction

This section of the document outlines the Data Cleansing Approach for Release 3. It describes the processes and activities as well as the high level timeline for Data Cleansing.

It sets out the roles, responsibilities for Data Cleansing Release 3 including the role of the Data Owners Forum.

For key decision points and activities see the high level action plan in the Executive Summary and Appendix F for more details.

2.2 Roles and Responsibilities

2.2.1 Business Data Stream Team roles

The Business Data Stream Team is financed by the Business Deployment Work stream and reports to the Chair of Data Owners Forum. This group works closely with the FAST Data Conversion Team on work stream 3 to ensure that IRs data is in the best position, i.e. valid and remediated, for conversion and implementation of BT Release 3.

The Business Data Stream Team is made up of a Project Manager/Data Lead and three Data Remediation Analysts seconded into the team, plus a further seven Data Remediation Officers on assignment from CCSI/CCSB to work on work stream 3 data issues. The number of Data Remediation Officers will be reviewed and adjusted if required in December 2018.

The timeframe for achieving this is outlined in Appendix F the Data Cleansing Plan.

Data needs to be managed from an IR wide perspective with a single point of contact. The Business Data Stream Team provides a connection between Data Owners Forum and CCSI, CCSB & IT&C. This ensures transparency and provides a channel for Data Owners Forum to advise and give direction on the appropriateness of on data cleansing activities underway. For more details see the proposed reporting and communications pathway in the Executive Summary and Appendix A.

2.2.2 Business Data Stream Team Responsibilities

The Business Data Stream Team is responsible for the following streams of work:

2.2.2.1 Work stream 1 - Everyday BAU data cleansing carried out by CCSI and CCSB

The Business Data Stream Team holds the following responsibilities in relation to this Work stream:

- Monitor the progress of this everyday work that has an impact on conversion (i.e. must do's) in order to provide visibility of all data cleansing activities to Data Owners Forum.
- Communicate deadlines and restrictions on bulk BAU data cleansing work driven by the BT Release 3 Data Conversion Plan in conjunction with the Deployment Team to CCSI and CCSB by way of intranet news items and People Leaders updates.
- NOTE: While it may be preferable that the bulk of this work is completed prior to the first Mock Go Live in November 2018, this BAU work will in reality continue right up to go-live.



- Receive data remediation work from BT (either from the BTMs/SMEs or from the FAST Data Conversion Team). This will then be channelled into the appropriate work stream depending on the effort/skills required.
- Where possible identify any impacts on successful Release 3 conversion as result of BAU data cleansing activities. Raise these issues promptly with Data Owners Forum, and the FAST Data Conversion Team.
- Inform CCSI, CCSB and IT&C of BAU data cleansing work regarding any decisions made by Data Owners Forum relating to their data cleansing work.

2.2.2.2 Work stream 2 – Known R3 data issues - data cleansing work carried out by CCSI, CCSB & IT&C

The Business Data Stream Team holds the following responsibilities in relation to this Work stream:

- Identify data cleansing work that falls into this work stream. Develop solutions and size the issues.
- Access resources within CCSI and CCSB to complete pieces of data cleansing work. Ensure that business impact statements (e.g. time and people requirements) are undertaken by the Capabilities & Outcomes workflow team as required to enable this.
- Access people and resources from within IT and BPS should a technology solution be required to cleanse data in FIRST as required.
- Ensure that legal and policy implications are determined as required.
- Utilising existing automation frameworks such as Business Systems Imperatives and Events (a team within the Capability & Outcomes, Customer & Compliance Services – Individuals Group) processes where applicable for agreed automation opportunities for remediation and data quality issues
- Collaborate with CCSI, CCSB and IT&C regarding work to be completed by their teams
- Plan and monitor progress, reporting back to Data Owners Forum. This includes raising any issues and risks that may require their intervention.
- Communicate deadlines and restrictions on bulk BAU data cleansing work driven by the BT Release 3 Data Conversion Plan to CCSI and CCSB by way of intranet news items and People Leaders updates.
- Receive data remediation work from BT (either from the BTMs/SMEs or from the FAST Data Conversion Team). This will then be channelled into Work streams 1, 2, 3 or 4 depending on the effort/skills required.
- Identify any impacts on successful Release 3 conversion as result of BAU data cleansing activities. Raise these issues promptly with Data Owners Forum, and the FAST Data Conversion Team.

2.2.2.3 Work stream 3 – Converted data cleansing activity

The schedule for this work stream is planned and managed by the Data Conversion Team. See section 2.1 for more details.



The Business Data Stream team is responsible for:

- Administer converted data verification exercises.
- Manage verifiers with support from the FAST Data Conversion Team
- Oversee Customer Data Verification by BST Testers and re-verification of any exceptions logged
- Review verification exceptions with support from FAST Data Conversion Team and CT&SP BTMs and SMEs, these exceptions may lead to identification of new data remediation
- Provide input into changes to the conversion plan and follow on changes to the cleansing plan
- Provide additional training for the verifiers
- Implement changes and corrections to the conversion process.
- START design changes in conjunction with BT SME's
- Manage the Data Quality Register in [Information redacted]
- Identify and raise data remediation work with CCSI, CCSB & IT&C Units that requires manual intervention to ensure that the data is in the best condition possible to enable conversion to START. This will then be channelled into Work streams 1 or 2 depending on the effort/skills required.
- Utilising existing automation frameworks such as Automation Team, Business Systems Imperatives and Events (a team within the Capability & Outcomes, Customer & Compliance Services – Individuals Group) processes where applicable for agreed automation opportunities for remediation and data quality issues
- Access/engage appropriately skilled people and resources should a technology solution be required to cleanse data in FIRST (e.g. IT/BPS or a BSI project team).
- Perform or oversee data remediation as identified as part of Work stream 3.
- Select data to be remediated
- Allocate and manage work given to data remediation officers
- Review/Track remediated data issues
- Log decisions made by the Business Data Stream Team ensuring a robust audit trail is in place
- Coordinate a legal view to any data issues needed to be fixed, including any documentation to gain approval
- Receive data remediation work from BT (either from the BTMs/SMEs or from the FAST Data Conversion Team). This will then be channelled into Work streams 1, 2, 3 or 4 depending on the effort/skills required.
- Plan and monitor progress, reporting back to the START data conversion team and Data Owners Forum. This includes raising any issues and risks that may require their intervention.

2.2.2.4 Work stream 4 – Data Enrichment



The Business Data Stream Team collaborates with the Business Lead, Data Enrichment. See section 4 for more details.

(For detailed descriptions for the known issues at the time this document was finalised see Appendix B)

2.2.3 Business Transformation Managers

Business Transformation Managers (BTMs) are responsible for the management of data elements, data objects and processes, for both content and metadata within their Jellybean.

The 11 Tax and Social Policy Jellybeans are: Customer, Compliance, Financials, Revenue Accounting, Disbursements, Payments, Contact Centre/Voice Channel, Audit/Case, Analytics, Returns & Forms and eServices.

The 2 Tax and Social Policy products are: INC, FAM, and the 2 main Channels are Voice Channel and eServices. These spread across the 11 Jellybeans ensuring a joined up approach.

BTMs take an enterprise view and have a specialist role that incorporates processes, policies, guidelines and responsibilities for administering IR's data in compliance with policy and/or regulatory obligations.

The role of the BTMs in relation to Data Cleansing is to provide specialist knowledge via subject matter experts for thorough remediation analysis, prioritisation and recommendations to the Data Owners Forum.

Refer to Appendix C for details of the BTMs for Release 3.

2.2.4 Data Stewards

Data Stewards are responsible for the management of data elements, data objects and processes, for both content and metadata within CCSI & CCSB.

Data stewards take an enterprise view and have a specialist role that incorporates processes, policies, guidelines and responsibilities for administering IR's data in compliance with policy and/or regulatory obligations.

The role of the Data Stewards in relation to Data Cleansing is to provide specialist knowledge for thorough remediation analysis, prioritisation and recommendations to the Data Owners Forum.

2.2.5 Business Stakeholders

For Release 3 a more collaborative approach across all of IR is proposed. Improved transparency of this work is required to enable Data Owners Forum to be able to give advice and guidance on the best use of our limited resources in the time available.

Business stakeholders are the people leaders responsible for directing any data cleansing BAU work within CCSI & CCSB that falls into work streams 1 and 2. The point of contact between the Business stakeholders and Data Owners Forum is the Business Data Stream Team.

2.2.6 Data Owners Forum

The role of the Data Owners Forum is to own the business process (for data) across IR to ensure known data issues are incorporated into the Business Transformation solutions and includes:

- Validate the conversion plan and act as an escalation point for conversion decisions as part of Work stream 3
- Validate all recommendations and prioritisation of work that are to be managed or undertaken via the Business Data Stream Team



- Validate all data work, including management of BAU work, historical data and PICT problems (also known as user system support problems) as part of Work streams 1 and 2
- Act as an escalation point for data quality issues raised by the Business Data Stream Team on all 3 Work streams.
- Validate all proposed Data Enrichment work – to ensure all proposed work is appropriately prioritised and managed effectively to align with the delivery of Business Transformation benefits and customer outcomes.

The responsibilities of the Data Owners Forum are:

- Represent a wide range of Inland Revenue's business functions, act as advocates of the desired IR culture and will be the channel in and out of the business.
- Ensure all relevant and appropriate decisions are escalated for endorsement and/or approval by the Business Transformation Design Authority and/or the Business Transformation Executive Working Committee.

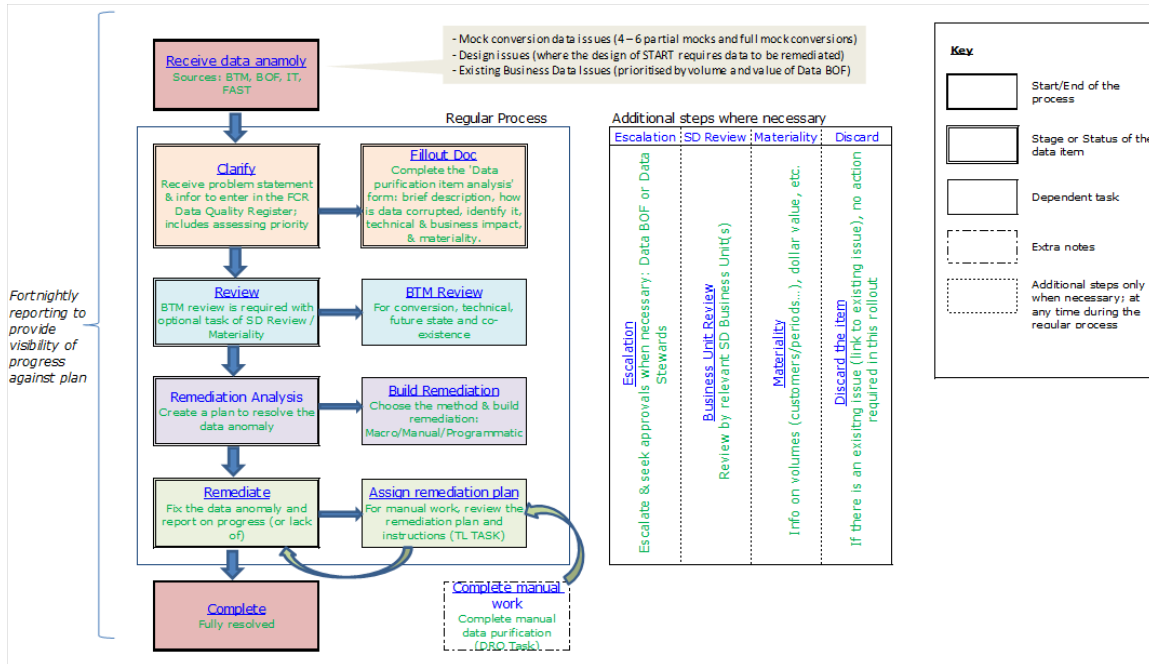
The membership of the Data Owners Forum is:

	Data Owners Forum
Chair	Customer Segment Lead, Small & Medium Enterprises
Members	Programme Director (Deputy Chair) Business Transformation Lead Programme Manager, Release 3 Enterprise Architect - Transformation Lead Group Lead, Individuals Customer Segment Director Corporate Finance START Project Manager Technology Lead, Heritage Change Lead, Organisational Change Management
Guests as required	Tax on Income Lead Families Lead Programme Manager, Intelligence Led Director, Digital Change Programme Manager, Release 4 ESS contact – (if required)
Attendees & presenters	Data Stream Lead, Business Data Stream Data Enrichment Lead FAST Consultant, Project Management
Forum Co-ordinator	Governance co-ordinator, BT PMO

2.3 Data Cleansing Approach

Many of the data issues will require cleansing prior to the Release 3 conversion. However, due to current system limitations in FIRST it might be more feasible to resolve some issues post conversion for Release 3 (e.g. historical or broken data not required in START).

2.4 Process for Data Cleansing / Remediation for work stream 3



Data conversion cleansing requirements are identified via the BTM’s and the FAST Data Conversion Team and logged in [Information redacted]

Guided by the Data Conversion Approach (this document Section 1), the Conversion Plan, and the BTM’s and START, the scope of the Data Quality Register items with priority sufficient to be cleansed will only include data required to enable Release 3 conversion and implementation.

The prioritisation of work is done in consultation with START, BTM’s and the SME’s with endorsement or approval (as required) by the Data Owners Forum.

The main purpose of cleansing / remediation is to action data where specific intervention is necessary to provide START with the data it needs to continue operations.

Data Extraction and Verification occurs periodically which identify further data issues for prioritisation.

Where data sets are identified and where action is required manual data remediation will be considered; however manual remediation is not the primary option. Any automation or programmatic options for cleansing will be considered first.

All data deemed as requiring remediation is given a problem statement and if required undertakes data remediation analysis.



The outcome of this analysis can be a recommendation to action or accept.

Should data be accepted in its current form, then that data may convert to START in the current form (with no remediation).

Where the analysis recommends action – then this data is purified by priority. To take some data sets to an outcome will require legal and policy considerations and appropriately delegated decision making. Recommendations are taken to the Data Owners Forum to ensure that appropriate legal, policy and decision making has been applied. Also it is likely that data quality issues with a low priority may not be remediated.

2.5 Data Cleansing Guiding Principles

Approved by PGB April 17 for Release 2, the same principles will be applied for Release 3.

A plan will be developed in conjunction with FAST for how Inland Revenue will manage manual data cleansing and validation that is in scope for Release 3 using existing standards.

We will use existing practices to address data anomalies leveraging current delegations and escalation processes.

A process, including appropriate escalation to manage data issues that cannot be mutually agreed to ensure effective and timely resolution will be developed.

Business data will be managed in a structured way using the correct delegations and ownership.

Standards already agreed across Inland Revenue will be used – however the Data Owners Forum will recommend practical changes if needed.

Ensure that there is an agreed approach that can be consistently applied in Release 3 including identification of long term data issues Inland Revenue needs to address.

Ensure that early consideration of the impacts of Release 3 Income tax changes on Release 3 deployment are made to ensure Social Policy products are not compromised.

Will continue using the Data Business Owner Forum to gather business input and agree decisions when required.

Will consider the Strategic Governance inputs recognising Organisation Design and programme and business requirements, to ensure the right decisions are made at the right time in relation to data.

2.6 Lessons Learned from Release 2

Release 2 data conversion identified many successful approaches for dealing with data and it would be practical to use the same or similar methods for Release 3. The lessons learned during Release 2 and the actions taken against them are shown below:

JIRA Ref	Description	Actions Taken	Status
BTLL-307	Collaboration between the Business Data Stream Team and FAST Conversion Team.	This was a "what went well" lesson so no action required.	Closed
BTLL-308	DataBOF as a governance group.	This was a "what went well" lesson so no action required. Note: the Data Owners Forum is a reporting group rather than a governance group.	Closed
BTL-359	Time consuming and complicated administration tasks for setting up conversion verification.	The Data Stream Lead has met with the FAST Project Manager responsible for the process and organised a fix for this for R3 – they will send through a list of who	In Progress



		the testers are for the following week to our team by COP Friday.	
BTL-360	Earlier identification and allocation of data cleansing of issues as BAU work as part of work stream 2 is required.	The Business Data Stream team are actively considering and managing these during their weekly team meetings.	In Progress
BTL-362	Changing reporting requirements for various governance Group	A new format consolidated view for the Data Owners Forum has been tabled and is now being updated based on their feedback. PMO have confirmed their reporting requirements for Release 3. The Business Data Stream Team will report on work stream 2. The START Conversion Team already reports on work stream 3.	In Progress
BTL-363	Data cleansing work being carried out as BAU - Collaboration between the Business Data Stream Team and the Business on data cleansing work being carried out as BAU as part of work stream 1.	The Data Stream Lead is developing a regular coms plan to enable this.	In Progress
BTL-364	Data Conversion Plan - Conversion rules should be run past Heritage IT as things change. During R2 when the Heritage IT team asked to see the conversion rules they encountered push back.	This has been logged as risk BTBTR-4332. For R3 the recommendation is that the conversion rules be shared with the Heritage IT team so that they can help identify any gaps so that we avoid unnecessary post conversion issues.	Open

3 Data Conversion Methodology and Approach

This section of the document focuses on principals and guidelines that makeup the methodology for converting data into START. Its contents were delivered during stage 1 of BT and are intended to be independent of any one release or stage of the programme.

The accountability for Data Conversion is with the FAST Data Conversion Team.

3.1 Work stream 3 – Converted data cleansing activity

The data cleansing work that is generated by the mock conversion and mock go live exercises is driven by the FAST Data Conversion Team and supported by the Business Data Stream Team.

The FAST Data Conversion Team is responsible for:

- Planning and managing the data conversion schedule.
- Identifying data issues that require remediation in conjunction with the BTM's.
- Identify work required to be remediated as part of Work streams 1, 2 and 3
- Providing technical support for the Data Quality Register in FCR
- Identify and analyse issues and incidents requiring urgent remediation.

For details of the cross-over responsibilities with the Business Data Stream Team on this work stream see section 2.2.2.3

3.2 Inventory Data Resources

The first task in the conversion process is to identify and inventory all of the possible FIRST data sources. Other conversion activities typically cannot be started until this inventory has



been created. The inventory will define the scope of customer data that is available for conversion.

3.2.1 Creating the Inventory

In addition to identifying sources of data for potential conversion, information useful for further actions should be collected. This information will include but will not be limited to:

- Types of information such registration or financial information
- Inland Revenue personnel such as subject matter experts with expertise regarding specific data
- Technology and platforms used for data storage.

This task will be done for each rollout, each iteration will produce a separate document.

3.2.2 Review the Inventory for Integrity and Quality

Once the inventory has been created and other conversion activities can continue, each data source should be examined for integrity and the quality of its data. This process should seek to determine if the data in the resource is consistent, reliable, and complete.

This process will require the conversion team to be granted access to the resources inventoried. If granting access requires a protracted process and an extended amount of time then it may be necessary to include others in the conversion team who currently have access.

3.3 Data Purification

FIRST application data will be analysed prior to and during the conversion process, and corrected *within the FIRST application* prior to START conversion. Wherever possible, existing on-line and batch facilities provided by the FIRST applications will be used to make the data corrections in the source environments.

The FAST Data Conversion Team will identify the corrections to be made in order for the data to be compatible with START. The FAST Data Conversion Team is responsible for developing the approach to “filling in the blanks” and setting defaults when START requires data that does not exist in the FIRST data stores. Correction should occur *within the FIRST applications* prior to START conversion. Corrections should not take place in backups, or any other auxiliary replicas of the production data. Nor should data be corrected as it is extracted from the FIRST systems for conversion to START. Ensuring that the data is cleansed in the FIRST application prior to conversion will allow the conversion process to be reconciled and users to verify that the conversion was correct and accurate.

It may not be possible to resolve all issues identified in the FIRST data. For example this could occur if there is not functionality available in the FIRST system to resolve the issue, there are not enough resources available, there is insufficient time to address the issue, or legislation prohibits it. It is recommended that data purification begin as early as possible once issues are identified in order to allow sufficient time for data corrections.

If it is not possible to correct an issue before conversion, an analysis of the impact should be conducted and risks associated with the issue identified. In addition, each issue should include a plan in order to mitigate the impact of unresolved data issues post conversion. Some of the solutions used to address FIRST data issues will include:

- The FIRST selection criteria will be adjusted to exclude the data (where appropriate)
- Use the “Do Not Convert” list to isolate undesirable data. This list is compiled over time; each category is reviewed by the Business Data Stream Team and signed off by a Core Business Transformation Manager (BTM).



- Post-conversion manual cleansing in START, which can be implemented during cutover or provided as work lists for cleansing at a later date
- Indicators can be added at the customer, account, or period level to modify system behaviour post conversion.
- Data can be manually converted

3.4 Detailed Conversion Plan

For each rollout a specific plan of what will be converted will be created and appended to this document. The conversion plan will be stored in the [Information redacted] delivery workbench as definition items. These plans will detail all aspects of conversion including:

- The types of data that will be converted
- The source systems of that data

How that data will be converted

How much of the data will be converted.

The FAST Data Conversion Team Lead is responsible for developing the detailed conversion plan. The decisions made in these plans will be based on the following principles.

3.4.1 Approach to Historic Data

Only current, active information will be converted. For example, the history of a customer's mailing address changes over time will not be converted; the conversion process will provide START with only a customer's most recent address.

3.4.2 Types of Data to be converted

It is helpful to group the types of data to be converted. The conversion plan will include the following groups by default, and additional groups will be added as necessary.

Registration – This will include customer, location, and account demographic information like names, addresses, contacts, etc.

Financial – This will describe how financial information will be converted, typically a system of record will be identified for period level balances and another for the individual transactions inside that period. This section will typically include returns / filings, payments, and disbursements.

Enforcements – This outlines the activities around compliance which usually includes collections, securities, payment plans, etc.

Content – This outlines the approach that will be followed in regard to non-structured data such as images and correspondence. This information is not always converted into START but may be converted to another system or left where it is.

3.4.3 Establish a System of Record

Depending on the FIRST environment, data may be stored in multiple systems. In that situation a system of record needs to be established for each type of data. For example both the registration and collections systems may contain names and addresses, and this begs the question of which system should be used as the master for this information. In this situation the quality and completeness of the both systems along with business direction should determine which system is the system of record.

3.4.4 Live vs. FIRST Reference



Not all FIRST data should be converted to live data in START. Many aspects of FIRST data may not be required for START operations, but may be needed for users to do the research needed and an integral part of their daily tasks. Consequently, data will be converted into one of the two following formats:

Live *(operational): This is the FIRST data START will need on an on-going basis for things such as tax return amendments, collections of unpaid balances, revenue accounting, and issuing correspondence. The conversion plan primarily deals with the conversion of operational data.

Legacy Data: This is FIRST data typically not used in daily START processing but which is available to START and users to view. Legacy data can assist users by providing a more complete picture of the data as it existed in the FIRST systems. Legacy data structure design will typically parallel the structure of the data in the FIRST systems.

3.4.5 Approach to Selection Criteria

It is important to clearly define which data is required to be converted into the START operational database. A realistic scope of the amount of data to convert will be developed, and this scope is referred to as the 'selection criteria'.

Selection criteria will be developed for different categories of data which at a minimum will include periods, accounts, and customers. A separate set of selection criteria will be created for each data category and it is these selection criteria that will determine how much of each category will be converted. The following principles should be observed when developing selection criteria:

3.4.5.1 Impact on Daily Operations

Selection criteria should be developed to support important administrative activities. Converted data will be used primarily for processing amended returns, posting payments to the correct debt, and audit, billing, and collection activities. Consequently selection criteria should identify data in both scope and breadth that will support these (and other important) activities. Additionally, statistical information regarding the use of historical data with respect to these activities should be used to narrowly focus the selection of data.

3.4.5.2 Impact on Planned System Processes

Converted data must be sufficient to support planned systematic processes. For example, if the process for calculating penalty requires payment information over the last 24 months, then this must be taken into consideration when defining the scope of period information for conversion.

3.4.5.3 Impact on Data Purification

Because data quality typically declines with age, age should be a consideration when defining selection criteria. The decision to convert older data will increase the effort required to achieve minimum quality standards.

3.4.5.4 Impact on Cutover Timeline

Converting more data will take longer to achieve. Because conversion is performed in a "Big Bang" fashion where all data is cutover to the new system during a shutdown period the decision to convert more data will impact the amount of time required to perform overall conversion during that shutdown period. Often the difficulty will be that the amount of time required will not be determined until the mock run process has progressed in include to all data, full mock which will occur much after selection criteria needs to be decided.

3.4.6 Automatic vs. Manual Processes

Where appropriate, the conversion plan will call for the use of START on-line functionality to manually convert small amounts of data. These decisions will be driven by data volumes, the complexity of the manual task, and the resources of developing, testing, and maintaining automatic conversion processes. In some cases a manual conversion process will be more appropriate, less error-prone, and more efficient, than developing and testing a one-time-use conversion program to convert a small number of items.

3.4.7 Approach to Work in Progress

Work that is still in progress at the time of conversion should not be brought into START. This therefore will be a key task that should be completed prior to cutover.

3.5 Conversion Process

The standard START conversion process entails

Obtaining data extracts from the FIRST systems.

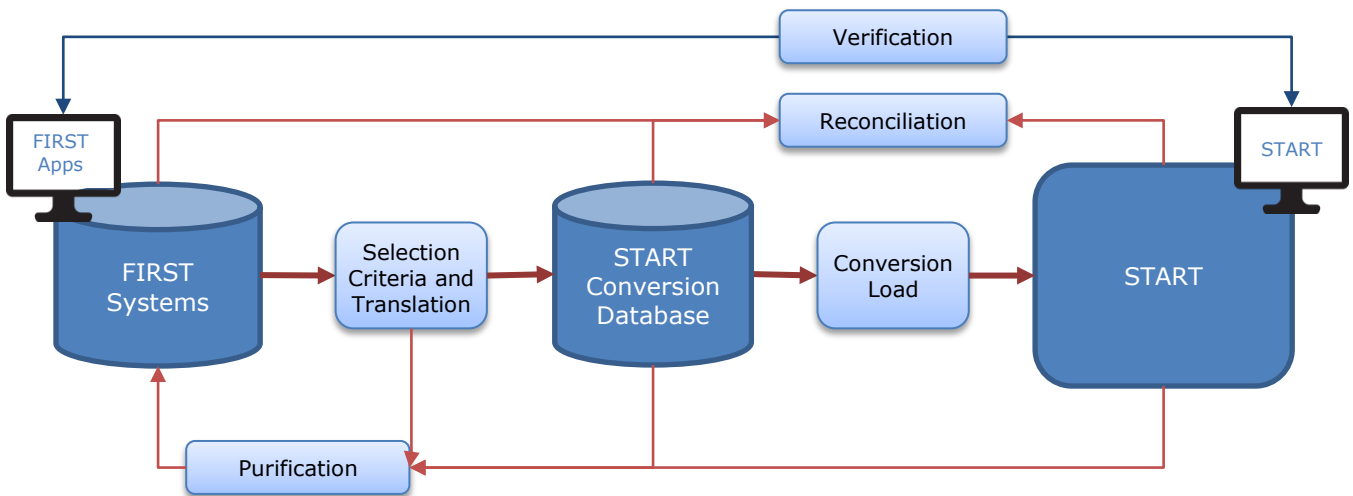
Translation of data extracted from FIRST systems to be loaded into START.

Data loads into START via business objects tailored for Inland Revenue business needs.

Reconciliation of data converted to START.

User verification and testing of data when it arrives in START.

Data purification in the source system.



3.5.1 Extract Data from FIRST Applications

Necessary data from the applicable FIRST applications will need to be made available to the START conversion environment. This data must come from the production system and must fully represent the FIRST application data. This data is then translated to tables used to load the data into START. Inland Revenue (IR) is responsible for facilitating the conversion team with access of this data.

How data will be accessed from the primary FIRST system, FIRST on the mainframe, has already been determined. IR has contracted with the mainframe support vendor, to use a technology named DataBridge to create SQL Server clone of all data in the OLTP and ARCHIVE databases on the FIRST production mainframe. This data is made available to the START conversion environment through SQL Server log shipping allowing the conversion team to control how frequency updates are applied to the START conversion environment copy of the FIRST data.



Access to other FIRST system will be made by SQL connections as well as flat file extracts.

3.5.2 Translation of Extracted Data from FIRST Applications

Data will be translated from data extracts of FIRST systems into various tables used to load the data into START. The conversion team is responsible for creating and running the translation process. START jobs and SQL files will be heavily used by the conversion team in the translation process.

Separate formats will be developed for most record types, such as Account and Financial records. The extraction process will be a 'big bang' that will extract data in one relatively short, well-defined procedure. The following key principles will be observed.

3.5.2.1 Mapping and Reformatting vs. Transforming Data

While it will be necessary to perform mapping and reformatting in the extraction process it should be done in a way that does not meaningfully alter the data. Meaningfully altering data during the extraction process will lead to reconciliation issues and impede the ability to verify that the process is comprehensive and accurate.

3.5.2.2 Avoid Production Copies

Data should be extracted from production systems, rather than backups or other copies such as a Quality system.

3.5.2.3 Reduce Risk by Reducing the Toolset

Only the tools necessary to execute a standard conversion should be used. START includes out of the box conversion utilities, and extensions to these utilities will also be developed on site as necessary. This tool set is necessary and sufficient for running and managing the entire conversion process from end to end, from extraction to load.

3.5.2.4 Leverage IR Expertise

Knowledgeable and experienced IR staff, experienced with the technical and business aspects of FIRST data, will be leveraged to create conversion extract and translation processes.

3.5.3 Loading Data into START

3.5.3.1 Validate Transformed Data

Before data is loaded into START a series of checks will be performed on that data. While some of these checks could also be done during extraction and translation they should be performed directly before data is converted into START. These checks will include:

A systematic search for orphaned records and missing details

Initial reconciliation to selection criteria tables and FIRST report

Verification of metadata such as Business Industry Codes match those maintained in START

3.5.3.2 Customer is the logical Unit of Work

Data will be converted into START on a per-customer basis, and all data for a single customer will be converted in a single transaction. This means that a customer will not be converted unless all of the data associated with that customer can be loaded without error (an error will result in a rollback). This prevents the introduction of incomplete data into START, enables detailed reporting on conversion exceptions, and enables multiple customers to be loaded concurrently without interference from exceptions thrown by other customer records.

3.5.3.3 Use Standard Business Objects

Extracted data will be converted into START using START business objects; including several fit for purpose extension built specifically for IR. These are typically the same business objects that START runs during normal activity. This ensures that START will be able to process non-native data without exception and also provides consistency between converted and native START data.

3.5.3.4 *Suspending Some Core Functionality*

As stated above, converted data will be created primarily using the same business objects and configuration employed for processing native data.

Automatic system notes can be suppressed since all ids, names, and addresses will be added by the conversion process.

These changes will be effective only during the conversion process and will not affect any other START processing.

3.5.4 Reconciliation

Reconciliation is a critical part of the conversion process. It ensures the relevant data was extracted from the FIRST systems and converted into START. The major steps include:

Reconciliation of extracted data to the data loaded into START through the conversion process.

Reconciliation of the data converted into START with *independent* FIRST reports of the applicable data in FIRST systems.

3.5.4.1 *FIRST Report*

A FIRST report will be used in the reconciliation process to ensure that the correct data was converted. This report should be generated independently of the conversion team and should be as simple and direct as possible. Removing complexity from the report will reduce the risk of not reconciling because of invalid logic built into the report. Where available an existing well established report should be used. The FIRST report should include, at minimum, non-nil balance periods by individual period. This granularity is necessary to help identify and reconcile any possible discrepancies.

3.5.5 User Verification

While computer-based reconciliation of data extracted and loaded into START is necessary, it is insufficient to verify an accurate conversion. During user verification of converted data, knowledgeable users will view converted data in START and compare it to how it appears and is used in the FIRST applications. This stage will greatly increase the confidence in the conversion process and can identify, for instance, incidents of data elements being converted into the wrong field on a return.

As users identify issues with the converted data those issues will be logged and triaged and may lead to changes in the conversion plan. It is typical to find issues that will lead to changes in:

- Data purification
- Training
- START Development.

3.5.6 Mock Conversion Runs

NOTE: This section addresses mock conversion runs not Mock Go-Lives. Mock conversion runs occur much more frequently and occur much earlier than Mock Go-Lives.



Because the conversion process will be a one-time use, it must be tested and trusted. To provide that high degree of trust a series of iterative mock conversions will be completed. A mock conversion, sometimes referred to as a mock run or dry run, will test the conversion process by performing the same steps as will occur during the final live conversion. Mock conversions will include extraction, translation, conversion, reconciliation, and user verification.

Mock conversion runs will be executed iteratively and start as partial mock conversion runs and slowly ramp up to full mock conversion runs. A partial mock conversion run lacks either some type of data to be converted or uses limited selection criteria. A full mock conversion run contains all types of data and all data available.

Successive mock conversion runs will increase the breadth and/or depth of FIRST data in "step" fashion as START configuration for each step becomes more advanced. For example some steps will include:

- Customer level data
- Customer and Account level data
- Customer, Accounts, and Period level data
- Customer, Accounts, Period, and Collection data.
- It is important that mock conversion runs use, as close as possible, the actual steps, data, infrastructure, and users that will be involved in the final conversion. This will confirm that the conversion process is accurate, complete, and efficient. Mock conversion runs will verify the following:
 - The steps in the conversion process are correct and in the proper order
 - Data is being converted correctly
 - Assumptions regarding data volumes
 - The timing of conversion events including but not limited to:
 - Length of time required for the final full conversion is run
 - Pre-conversion and post conversion batch jobs in START
 - Sequencing of reconciliation tasks.
 - The adequate allocation of hardware resources for conversion.

3.5.7 Business Function Definitions

During conversion process another key document is used that records key Business Function Definitions.

In summary these definitions provide configuration specifications of how data is selected, designed and converted for START. These are assigned and defined by relevant Core Subject Matter Experts, designed by FAST Developers and approved by relevant BT Manager. It also provides foundation of various Business System Testing scenarios. The log of approved definitions is regularly provided to Data Owners Forum for review and confirmation.

3.6 System Cutover

3.6.1 Cut-off for Processing Items in FIRST Systems

To allow functionality to be cutover from the FIRST applications to START, a number of activities will be stopped, started, or altered. The System Cutover plan will detail the required activities and associated schedules and process owners and responsible staff. These activities are managed by the deployment team. The Conversion Team will be responsible for the



automatic conversion set of activities and the Data Cleansing Team will be responsible for manual conversion and data cleansing activities.

A point in time will be identified after which no new data should be processed in the FIRST applications. Common examples of these activities include:

- Payment batch uploads
- Return batch uploads
- Disbursement / refund processing.

- Other FIRST processes may have to be run one final time before that functionality can be performed in START, such as:
- Closure of revenue accounting time period
- Various month-end, year-end, period-end processing.

This may include running annual and semi-annual reports where data from FIRST and START will be required to generate a summary report.

3.6.2 FIRST Enters a Read-Only State

Once the conversion extract process has been initiated, no new activity or data should be entered in the FIRST systems. Access to FIRST applications during this time will be limited to read-only and might be unavailable due to impacts from the cutover process. Once data is being extracted for conversion to START the FIRST system should be set to "read only" for the products affected by that rollout in order to prevent users from performing work, such as new registrations or financial updates, that could have a detrimental effect on data conversion. Putting FIRST into a read-only state may require FIRST system programming changes. Such changes include, but are not limited to:

Modifying FIRST on-line functionality to prevent users from modifying converted accounts

Modifying FIRST off-line or batch functionality to prevent the FIRST system from automatically processing converted accounts.

3.6.3 Final Conversion Process

This will be the culmination of all efforts in the conversion process. The final conversion run will be managed and monitored using a detailed conversion checklist. Due to repeated full mocks acting as rehearsals leading up to conversion, the exact sequencing and time requirements for every item on the checklist will be known. In addition, the final reconciliation reports will be similar to those produced during the final full mock conversion.

3.6.3.1 Suspend Maintenance

Due to conversion activity, non-critical hardware or software updates should not occur in either FIRST or START environments during the final conversion. Any critical updates that should have occurred during the conversion process will be addressed after go live.

3.6.4 Going live

The primary goal of the project's conversion effort is to provide the agency with a base set of data to support its business needs when the new system comes online. At this point in the cutover plan that base data has been provided. The final steps are to accept the new system and enable access for all authorized users.

4 Data Enrichment Approach



Data enrichment covers any improvements to data with the aim of contributing to benefits realisation of the Business Transformation Programme. This is work stream 4 as mentioned in section 2.2.2.4. Many of the data issues that fall within this work

The Business Lead, Data Enrichment is responsible for:

- Develop selection criteria – including identifying relationship to Benefits and programme sequencing
- Develop prioritisation commitment
- Determine effort, cost, risk, value proposition
- Recruit external team
- Work closely with range of stakeholders including :
 - BT programme team – particularly BTMs and Product Owners
 - Segment Management Leads – understand existing work underway or planned in the business
 - FAST
 - Heritage
 - Data Lead, Business Data Stream Team – ensure alignment with existing work
- Regular reporting, measurement and monitoring to DATA OWNERS FORUM and potentially Executive Working Committee.
- Full scope to be drafted, project plan, milestones, deliverables
- Weekly meeting with Sponsor

5 Reporting

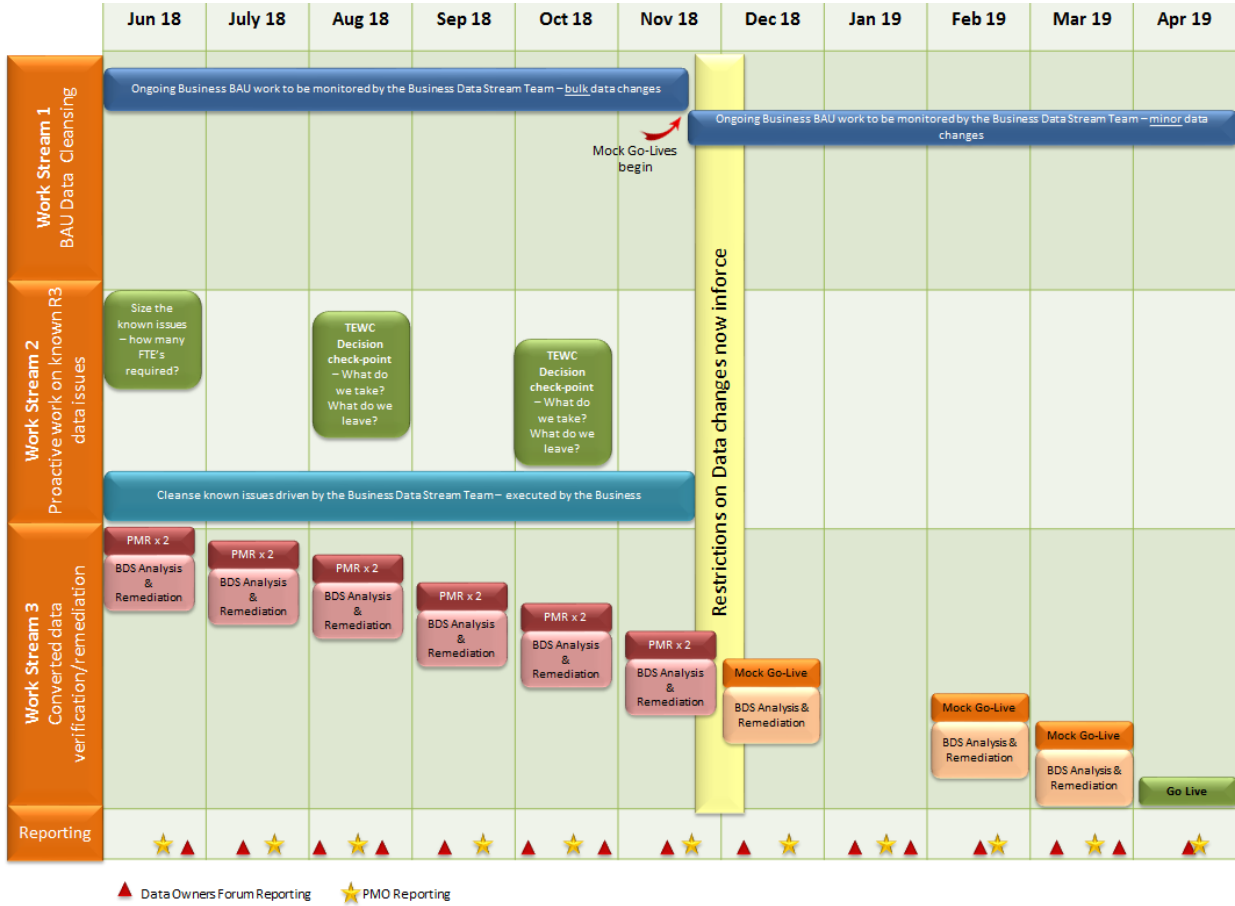
Data Owners Forum

Regular status reports for data cleansing, conversion and enrichment will be produced by each team (FAST Conversion Team, Business Data Stream Team and Data Enrichment Team) and presented to the Data Owners Forum for inclusion in the relevant Release 3 Programme Management reports as required.

Appendix A – Business Data Stream Team – The 4 Work streams



Appendix B – High Level Data Cleansing Plan/Timeline





Appendix G – Definitions

Data Cleansing refers to the correction of data that is required to enable the successful conversion of data to START and includes any automated or manual fixes in either the FIRST or START systems to provide START with a base set of data for IR to continue its operations.

Data enrichment refers to improving data in order to ensure that the intended benefits of the Business Transformation programme are realised both for the customer and the business experience. Often this work does not impact on data conversion while it may still impact on the business or the customer.

Data purification and/or remediation these terms are sometimes used interchangeably and over the top of Data Cleansing and Data Enrichment and depending on the context could refer to either. The qualifying factor will be whether the change is required to enable successful conversion of data from FIRST to START or to improve the Customer/Business Experience. Often there is a cross-over between the two outcomes.

Data Stewards Data Stewards are responsible for the management of data elements, data objects and processes, for both content and metadata within CCSI & CCSB. Data stewards take an enterprise view and have a specialist role that incorporates processes, policies, guidelines and responsibilities for administering IR's data in compliance with policy and/or regulatory obligations. The role of the Data Stewards in relation to Data Cleansing is to provide specialist knowledge for thorough remediation analysis, prioritisation and recommendations to the Data Owners Forum.

Business Stakeholders are the people leaders and decision makers within the Business that initiate and drive data cleansing work as part of BAU. This work falls into work stream 1. A close relationship needs to be maintained between the Business Data Stream Team and CCSI, CCSB & IT&C. Stakeholders to give Data Owners Forum transparency of all data cleansing activities being carried out across IR and to enable the communication of decisions and advice to the Business.

Core Subject Matter Experts (SMEs) are responsible for the management of data elements, data objects and processes for both content and metadata. Core SMEs take an enterprise view and have a specialist role that incorporates processes, policies, guidelines and responsibilities for administering IR's data in compliance with policy and/or regulatory obligations. There are a set of SME's for both FIRST and START systems that work closely with the Core Business Transformation Managers (BTMs). The Business Data Stream Team Data Remediation Analysts provide input and advice during design discussions held by the SMEs relating to impacts on conversion.

Core Business Transformation Managers (BTMs) are responsible for the management of data elements, data objects and processes, for both content and metadata within their Jellybean. The Business Data Stream Team Data Remediation Analysts provide input and advice into design discussions with the BTMs relating to impacts on conversion.

Converted Data Verification is an activity undertaken as part of testing the Release 3 conversion solution, and verifies that the data that we are moving from FIRST systems to START is moved/transposed accurately, in the correct location and as intended by the conversion plan.

The accountability and responsibility for the conversion plan (including Extraction, Transformation and Load) sits with the FAST Data Conversion Team. (See section 2) The Business Data Stream Team support and administer converted data verification by means of Work stream 3 of the Data Cleansing Plan.

Work stream 1 – refers to everyday BAU data cleansing carried out by CCSI, CCSB and IT&C rather than directly by the Business Data Stream Team (See 3.2.1.1)

Work stream 2 – refers to data cleansing work carried out on known R3 data issues by CCSI, CCSB and IT&C, driven by the Business Data Stream Team (See 3.2.1.2)

Work stream 3 – refers to converted data verification cleansing activity, driven by the FAST Data Conversion Team. (See 3.2.1.3)

Work stream 4 – refers to data enrichment to enable Transformation benefits realisation, driven by the Business Lead, Data Enrichment (see 3.2.1.4)