

DATA CONVERSION PLAN

**VERSION 1.0**

This template was created to enable departments to more easily develop their project plans. The Department of Technology, Consulting and Planning Division, created this template based on its experiences. The template relies on industry best practices combined with decades of experience on California state information technology projects. The way it was structured is to enable a department to complete the information related to its project without having to write background information related to the discipline. A department may use as much or as little of the template as it wishes.

**Template Instructions:**

* ***Instructions for completing*** this template – written for the author of the project plan - are encased in **[ ]** and the text is ***italicized*** *and* ***bolded.***
* *Examples* are provided as a guideline to the type of sample information presented in each section and the text is *italicized*.
* Boilerplatestandard language for each section is written in the document font and may be used or modified, as necessary.
* A department’s project specific information goes within the brackets ***<< >>***.
* *Informational text is italicized* within square brackets [ ] for informational purposes to the person who has to create the plan and includes background information, explanation, rationale, etc.

APPROVAL SIGNATURES

Example:

| *CONTRACTOR* | | *DATE* |
| --- | --- | --- |
| *<<Deliverable Owner>>*  *<<John Doe, Manager>>* | *<<Signature>>* |  |
| *<<Vendor Project Manager>>* | *<<Signature>>* |  |
| *<<Vendor Managing Director>>* | *<<Signature>>* |  |

| **State project manager** | | | | |
| --- | --- | --- | --- | --- |
| *Date:* | *<<Date>>* | *To:* | ***<<PM NAME>>,*** *Project Manager* | |
|  | *I approve this deliverable and have no further questions or comments.* | | | |
|  | *I approve this deliverable conditionally, contingent on the review and approval of the following corrections (see comments).* | | | |
|  | *I reject this deliverable for the following reasons identified (see comments).* | | | |
|  | | | | |
| ***<<SIGNATURE>>*** | | | | ***<<DATE>>*** |
| *Comments* | | | | |
|  | | | | |

DOCUMENT HISTORY

| **DOCUMENT APPROVAL HISTORY** | |
| --- | --- |
| Prepared By |  |
| Reviewed By |  |
| Approved By | <<Typically the Project Sponsor>> |

| **DOCUMENT REVISION HISTORY** | | | |
| --- | --- | --- | --- |
| Date | Document Version | Revision Description | Author |
| 5/1/2014 | 1.0 | Initial Version | Shawn Nguyen |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

TABLE OF CONTENTS

1. INTRODUCTION 1

1.1. conversion overview 3

2. project definition 7

2.1. Purpose of data conversion Plan 7

2.2. goals and OBJECTIVEs 7

2.3. scope of data conversion 8

2.3.1. Inclusions 9

2.3.2. Exclusions 9

2.4. critical success factors 9

2.5. conversion ACCEPTANCE criteria 11

2.5.1. Define data error severity levels 12

2.5.2. define Acceptance Criteria 12

2.6. Assumptions, constraints, and risks 13

2.6.1. ASSUMPTIONS 13

2.6.2. constraints 15

2.6.3. Risks 15

2.6.4. Risk mitigation strategy 17

2.7. communication strategy 18

3. conversion Requirements 20

3.1. business requirements and expectations 20

3.2. Technology and INFRASTRUCTURE considerations 22

3.3. Data security and privacy requirements 24

3.4. conversion data sources 24

4. current Environment analysis 25

5. data cleansing 29

6. conversion approach 32

6.1. methods of conversion 32

6.2. data mapping 34

6.3. Data extraction and staging process 36

6.4. Data transformation and LOADing process 38

6.5. Synchronization Process 40

6.6. conversion reversal strategy 40

6.7. DATA CONVERSION SCHEDULE 41

6.8. data conversion MILESTONES 41

6.9. conversion STANDARDS and CONVENTIONS 43

6.9.1. Design Standards 43

6.9.2. Coding Standards 43

6.9.3. Naming Conventions 44

6.9.4. CHANGE control 44

6.9.5. version control 44

6.10. data conversion environments 45

7. Conversion project team 47

7.1.1. STAFFING PLAN 47

7.1.2. team organizational chart 47

7.1.3. ROLES and RESPONSIBILITIES 48

8. data quality strategy 49

8.1. Conversion unit testing 50

8.2. Data Usability Testing 51

8.3. Mock Conversions 51

8.4. data validation and reconciliation (DVR) 53

8.5. DATA error resolution PROCESS 56

9. conversion implementation 57

9.1. Approach to implementation 57

9.2. conversion cutover process 59

9.3. implementation planning & considerations 60

9.4. Data certification 62

10. post-conversion support 64

10.1. identify post-Conversion data Issues 64

10.2. Resolve Post-Conversion data Issues 64

10.3. decommissioning and monitoring 65

APPENDIX A. CONSIDERATIONS FOR COTS, MOTS, and CUSTOM A-1

APPENDIX B. LIST OF ARTIFACTS B-1

B.01. Conversion Design Decision template B-1

B.02. Data Conversion Terms and Definitions B-1

B.03. Data Dictionary Template B-1

B.04. Data Mapping Template B-1

B.05. Data Quality Issue Log template B-1

APPENDIX C. DATA CONVERSION TEAM C-1

C.01. Conversion Roles and Responsibilities - Definition C-1

C.02. Conversion Roles and Responsibilities – RACI Chart C-1

APPENDIX D. TABLE OF FIGURES D-1

# INTRODUCTION

[This template is the result of experience and research of best practices concerning data migration/conversion. It is intended to provide a high-level overview for those individuals who are not familiar with data conversion and to serve as a guide or reference for those who are familiar with data conversion and/or who are in the process of developing a data conversion plan for the project at hand. Throughout the document, the terms “convert” and “transform”, as well as, “migration” and “conversion” are used interchangeably.]

[“Virtually everything in business today is an undifferentiated commodity, except how a company manages its information. How you manage information determines whether you win or lose." – Bill Gates.

Data conversions are a leading cause of schedule and quality issues in enterprise application implementations. Research from several sources, including Gartner, confirms that over 80% of data conversion efforts either fail outright or suffer significant cost overruns and delays. As a result, they jeopardize any other IT projects that depend on them. Some of the key challenges for data conversion are:

* Lack of staff with data conversion expertise.
* Lack of clearly established and realistic data conversion requirements, key stakeholder expectations, and data conversion acceptance criteria.
* Lack of documented and/or refined business rules.
* Lack of data governance.
* Lack of relevant business subject matter expert involvement.
* Target system data models that change during the conversion effort.
* Converted data was either validated only at the end and/or tested only with a subset of the data.
* Data conversion planning and scoping were done without a clear understanding of the data architecture and data quality of the legacy systems from which data is to be migrated.]

As depicted in Figure 1‑1, data conversion is the process whereby data from its current sources (e.g., existing legacy systems, hardcopies, document images, etc.) is extracted, transformed, and loaded to a new system. For most state departments, data conversion is often part of a larger legacy system modernization project; it involves multiple databases, file structures, utilities, toolsets, different computer languages, and computer operating systems. Because it is a complex and time consuming process, data conversion is one of the most critical elements associated with a successful system implementation. Furthermore, data is a critical business asset; it is the foundation of valuable intellectual property and the lifeblood of every organization. Thus, unless valid, complete, accurate, and compatible data components are available in the new system, the new system simply cannot be useful regardless of how friendly the user interface is that was designed, how streamlined the processes implemented, or the amount of effort exhausted during the development of the new system.

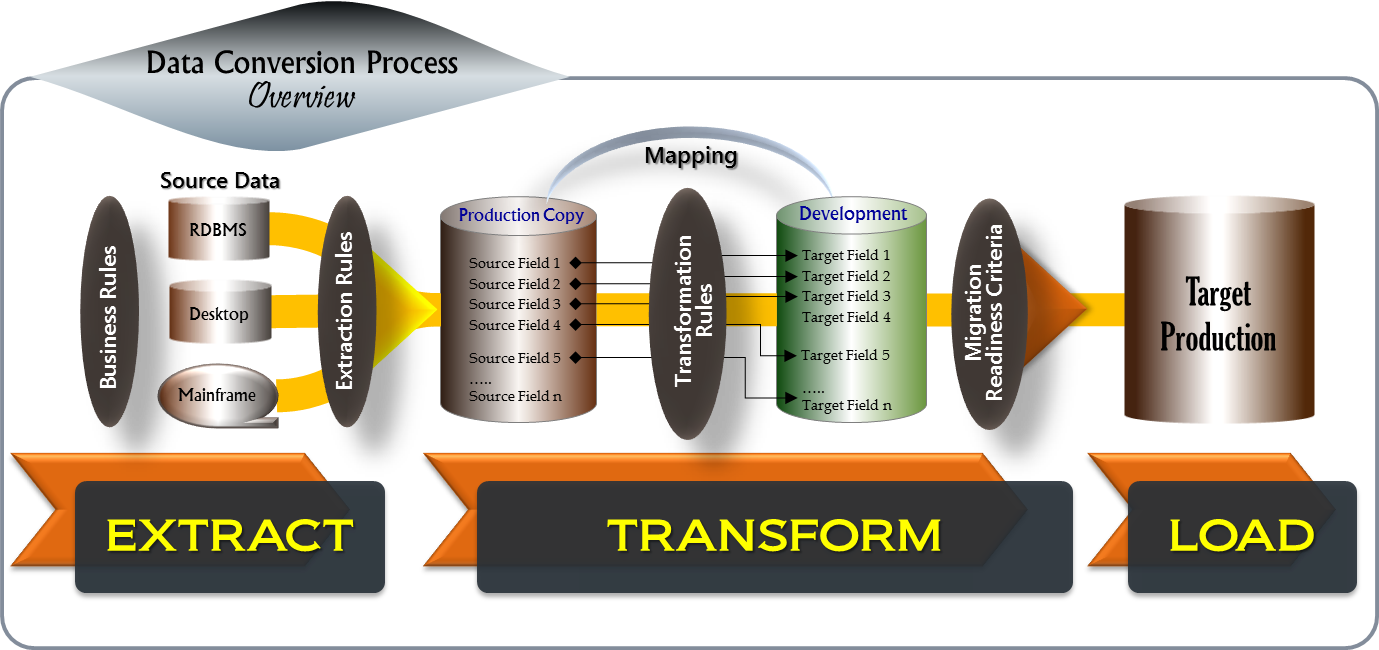


Figure ‑: data conversion process overview

## conversion overview

This section provides an overview of the key aspects of a data conversion effort. The document will discuss each of these aspects in more detail in the subsequent sections.

[Although data conversion is simple in concept, it can be surprisingly complex and time-consuming due to many reasons, including the challenges stated in the previous section. Therefore, in order to increase the chances of success on a data conversion effort, it is necessary to establish and follow a strategy with clearly defined phases, processes, and milestones along the way. Entry and exit criteria should be defined for each phase, process, and milestone..

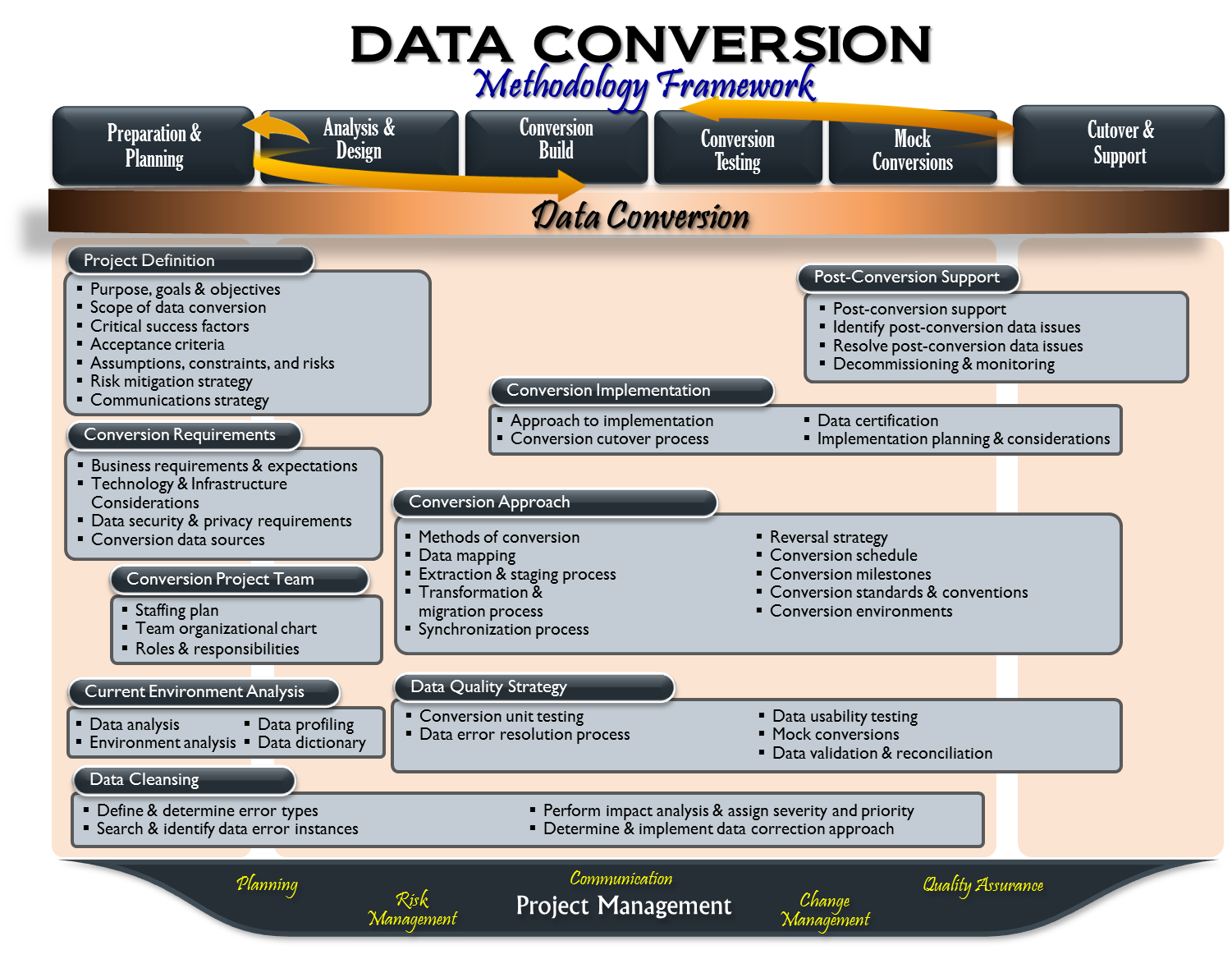


Figure ‑: data conversion methodology framework

The data conversion process follows a typical software development lifecycle with the addition of steps for current environment analysis and mock conversion runs. As shown in Figure 1‑2, a typical data conversion project consists of preparation and planning, analysis and design, conversion build, conversion testing, mock conversions, conversion cutover, and support.

Figure 1‑2 also provides a methodology framework for data conversion. It lays out the logical building blocks that are essential to a data conversion effort. Each of these building blocks consists of a set of activities or processes which span across multiple phases of the data conversion lifecycle. Many of these activities happen concurrently or in an iterative manner but they are all pivotal to the successful implementation of data conversion. The following provides a high-level description for each of these building blocks:

**Current Environment Analysis**

The purpose of current environment analysis is to assess whether the data migration is viable, how much time it will require, what technology it will need, and what potential issues the project team will have to face. In addition, it is to understand where and how the data is stored, backed up, and archived, along with identifying data quality issues and how they will impact the data conversion effort. Also analyzed at this point are the interfaces, network connections between data points, bandwidth and controls, and data security. Furthermore, this is where a detailed data dictionary, business rules, high-level source-to-target mapping specifications, and conceptual and logical data models are developed.

**Conversion Requirements**

Conversion requirements are identified in an effort to understand:

* What data will be migrated?
* How much downtime is acceptable for the current production systems?
* Are new or modified service-level agreements required?
* What are the expectations for the target data storage infrastructure?
* What are the organization’s standards or policies concerning data?
* What historical data is to be migrated?
* Are there any technological aspects of the current environments, conversion environments, and/or target environments that need to be considered?

**Data Conversion Effort Definition**

Information produced by current environment analysis and conversion requirements is essential to solidifying the planning and scoping of conversion activities; it enables the project to further refine and solidify data conversion scope, goals, assumptions, risks, and expectations, timelines, and acceptance criteria. Furthermore, it gives visibility into the work involved, which enables the project to effectively plan and properly allocate necessary staffing resources to support the data conversion strategies outlined in this plan.

In addition, information produced by the previous activities also provides clarity on the current system’s data quality issues which enables the project to analyze what impact these will have on the data conversion effort and, ultimately, on the business if not addressed before the target system is implemented.

To the data cleansing effort, this information provides the data cleansing team understanding of various types of data issues present in the current legacy systems and the data population associated with each data error type. This enables the team to determine proper correction approaches to effectively address data errors.

**Conversion Approach**

This is where strategies and processes for data extract, transformation, and load are analyzed and determined using the information provided by the current environment analysis and conversion requirements. Conversion methods are carefully studied; detailed source-to-target data mapping specifications and conversion design specifications are developed, conversion environments including staging area are modeled, and data quality strategy is examined and planned.

**Data Quality Strategy**

As design specifications are completed and approved, conversion processes and conversion programs are developed accordingly; data is then extracted, transformed, and loaded to conversion data storage, and unit testing occurs. Results produced by these conversion processes and programs are then tested at several levels to identify problems as well as to ensure accuracy and completeness of the conversion process. This is an iterative process, which continues until all necessary conversion processes and conversion programs are developed and successfully tested.

Once all conversion processes and conversion programs are successfully tested, mock conversions or trial runs of the actual conversion process are conducted in a pre-production environment. The purpose of mock conversions is to identify and resolve any conversion program issues and configuration problems ahead of time.

In addition, mock conversions provide the project opportunities for independent data validation of the actual data volumes. Furthermore, mock conversions enable the project to assess data conversion readiness for cutover as well as verifying and ensuring that the conversion process can be finished within the timeframe allocated for data conversion cutover.

**Conversion Implementation**

Conversion implementation involves evaluating different implementation approaches, defining and planning the conversion cutover process, and establishing the data certification process that will be used to facilitate conversion implementation.

**Post Conversion Support**

Once conversion readiness is achieved, the final data conversion is executed. Results are then validated and loaded into the new target production system. Data certification is conducted to ensure data conversion meets the established acceptance criteria.

After data conversion cutover is successfully implemented, inevitably data issues will be uncovered and the number of data issues may be greater than originally anticipated. Post-migration support plan is implemented and post-conversion support staff will begin the process of identifying, recording, and resolving data issues. This is also the time where data conversion decommissioning activities and data quality monitoring activities begin to take place.]

# project definition

This section describes the purpose of the data conversion plan, goals and objectives, scope of data conversion, critical success factors, acceptance criteria, assumptions, constraints, and risks associated with the plan for achieving the goals and objectives.

## Purpose of data conversion Plan

[Describe the purpose of the data conversion plan and provide pertinent information about the systems undergoing conversion; summarize the scope of activities that resulted in its development and state who the intended audience is for this document]

The purpose of this data conversion plan is to describe the strategy, preparation, and specifications for migrating existing <<legacy systems>> data to the <<target system>> data store. This is intended to provide not only a roadmap for performing data conversion, but also a means to discuss assumptions, risks, limitations, constraints, and essential processes that will be used in the process of data conversion.

The intended audience of this data conversion plan is <<key data stakeholders, project leaders who sign off the plan and the data conversion project team>>

## goals and OBJECTIVEs

[“Without goals, and plans to reach them, you are like a ship that has set sail with no destination.”- Fitzhugh Dodson.

There must be specific goals and objectives that the project intends to achieve through the implementation of this data conversion plan. Primarily, the goals and objectives of this effort may be that:

* All data needed to support the following core business processes in the new target system must be converted and migrated from the identified legacy systems to the target system completely and accurately, as compared to the source, and in accordance with department and regulatory policies on information controls and security. Furthermore, converted data must be compatible with the target application. This means there are no dropped or incomplete records, and converted data must work well with the target application.
* Data conversion process must be completed within the allotted timeframe during the conversion cutover.
* The quality of converted data must meet or exceed the established conversion acceptance criteria.]

[Describe the goals and objectives that the project intends to achieve through the implementation of this data conversion plan.]

The objective of this data conversion effort is the migration of data from <<legacy systems>> to <<the target system>> to support core business processes in the new target system.

The underlying goal of this data conversion effort is to populate the new target system with data necessary to support core business processes. The core business processes to be supported by the conversion of data from <<legacy systems>> to the <<new target system>> are:

* <<core business process>>
* <<core business process>>
* <<core business process>>

## scope of data conversion

[In order to increase the accuracy of data conversion scoping, it is important that the data conversion team fully understands the current legacy environment from which data is to be extracted and converted prior to finalizing the scope of data conversion. This is best achieved with a comprehensive analysis of the current legacy environment, which will provide the data conversion team the following benefits:

* Visibility of all the data sources in order to identify and plan to address any potential problems.
* Choice of conversion method being properly determined and optimized based on the state of the source data.
* All data conversion related decisions being made based on proven facts rather than assumptions, theories, or previous experience.
* Scoping, planning, mapping, building, testing, validation, and cutover activities all being determined based on a thorough and accurate assessment of all the source data rather than theories or previous experience.

Scope is a set of boundaries that defines the extent of the data conversion effort. These boundaries determine what falls inside or outside the data conversion effort. Activities that fall inside the boundaries are considered “in scope” and are planned for in the schedule and budget. If an activity falls outside the boundaries, it is considered “out of scope” and is not planned. This can be done by defining what the project will deliver.

As a reminder, scope of data conversion should address all aspects that are related to data conversion, not just the development of conversion programs and processes alone. For example, data validation, data cleansing, and post-conversion support should be considered. Therefore, it is imperative that the scope of data conversion is clearly defined at the outset of data conversion to prevent “scope creep,” which might reduce the project’s chances of success. Since many state department data conversion efforts are part of a larger legacy system modernization project, the scope of data conversion concentrates on migrating the data from multiple legacy source systems to a single target system, and includes interfaces and reports.]

[Define the scope of the data conversion effort. The scope should include all pertinent information to enable the data conversion team to properly plan for the level of effort, the timeline, and the resources needed to accomplish tasks as well as to help identify dependencies and potential risks.]

### Inclusions

[If needed, this section can be used to specifically describe all the major components that are included in the scope of the data conversion effort.]

The following items are included in the scope of the data conversion effort:

| **in-scope items** | **notes** |
| --- | --- |
| << Item 1>> | <<Notes>> |
| << Item 2>> | <<Notes>> |

### Exclusions

[If needed, this section can be used to specifically describe all the major components that are NOT included in the scope of the data conversion effort.]

The following items are excluded from the scope of the data conversion effort:

| **Out-of-scope items** | **notes** |
| --- | --- |
| << Item 1>> | <<Notes>> |
| << Item 2>> | <<Notes>> |

## critical success factors

[Success factors are elements within the structure and context of the project that are conducive to success. Although their presence will not guarantee success, their absence will increase the probability of failure. Successful projects are those that embrace and apply the guiding principles of maximizing opportunities and mitigating risks. The following are some factors that are pivotal to the success of a data conversion effort:

* Key goals and objectives of data conversion are clearly defined.
* Comprehensive analysis of the content, quality, and structure of the legacy source data is performed prior to finalizing the project scope.
* Data conversion is done as an independent project.
* Data conversion plan is well prepared, approved, and followed. It is realistic and can be implemented, specific and can be tracked or measured against, and fit to achieve the objectives; it also clearly identifies relevant assumptions and realistic risks.
* Data conversion acceptance criteria and expectations are clearly established at the outset and managed throughout the process.
* Importance of business data and data quality are determined through the coordination with business owners and stakeholders.
* Validation of converted data is independently performed by a separate team.
* Source and target data, and business requirements, are clearly articulated and understood.
* Individuals, including IT staff, with institutional knowledge and expertise of legacy data are identified and actively involved throughout the project. Documentation concerning legacy system(s) is gathered or created, if not available.
* Roles and responsibilities of the data conversion team are clearly defined, communicated, and understood.]

[Identify all relevant success factors that are pivotal to your data conversion effort.]

The following success factors are considered pivotal to the success of the data conversion effort.

1. <<Success factor 1>>
2. <<Success factor 2>>
3. <<Success factor 3..>>

## conversion ACCEPTANCE criteria

This section defines the conditions that must be met in order for the converted data to be considered ready for cutover.

[“Begin with the end in mind.” - Stephen R. Covey.

It is nearly impossible to have all records migrated from source to target with no functional, reconciliation, or compatibility data errors by the time of system cutover. Therefore, working with the key data stakeholders to formally establish the data conversion acceptance criteria at the outset is crucial to the success of data conversion. Prior to the final data conversion execution, if the data errors found are within the tolerance boundaries of the established acceptance criteria, key data stakeholders can make a decision to cutover operations to the new system. However, if the data errors found are beyond the tolerance boundaries, key data stakeholders can decide to postpone the cutover, continue operations on the source system until the data errors are addressed, or go forward with the cutover as scheduled and address the data errors post conversion.

Therefore, the best place to begin the data conversion effort is with acceptance criteria. Acceptance criteria are a list of conditions that must be met in order for the converted data to be considered ready for cutover. Acceptance criteria have to be clearly defined; ideally they should be SMART: - Specific, Measurable, Agreed Upon, Realistic, and Time bound.

The first step in establishing acceptance criteria is to prioritize the importance of the data. Not all business processes are prioritized equally and neither is the data within each business function. Therefore, data needs to be prioritized by business function and then by data classifications, groups, or types as conceptually illustrated in Figure 2‑1.

Figure ‑: data conversion acceptance criteria

Pay careful attention to the data volume associated with each business function, the severity level, and its corresponding acceptance criteria because the number of records for 1% of 15 million records is very different than that of a half million-record dataset. Acceptance criteria are one of the key factors in determining the feasibility of the data conversion effort. Furthermore, by having the data conversion acceptance criteria clearly defined from the beginning, the data conversion team is better able to measure and monitor the progress of the data conversion effort along the way.]

### Define data error severity levels

[Define the level of each severity to be used as part of acceptance criteria for data conversion. The Figure 2‑2: definition of data error severity levels is an example which can be tailored for the current project.]

| ***definition of data error severity levels*** | | |
| --- | --- | --- |
| ***Severity Level*** | ***Severity Description*** | ***Example*** |
| ***Severity 1*** | * *Defect prevents <<business function>> from rendering critical business services accurately to its employees and Business Partners.* * *No work-around, as acceptable to the <<project leadership team>> ,is available and service cannot be rendered until the defect is resolved.* | *Failure to convert an employee demographic and employment data thereby preventing the employee from receiving correct employment benefits.* |
| ***Severity 2*** | * *Defect prevents <<business function>> from rendering critical business services accurately to its employees and Business Partners.* * *An acceptable workaround was provided and approved by the <<project leadership team>>.* | *Incorrect application of rules to create offset transactions resulting in an inaccurate account balance.*  *Workaround: users should use the Detailed Account screen to manually enter adjustment transactions.* |
| ***Severity 3*** | *Defect does not prevent current services from being rendered accurately by <<Department>> to its employees and Business Partners.* | *Incorrect data types and data defects pertaining to old data used for reporting/inquiry purposes only.* |

Figure ‑: definition of data error severity levels

### define Acceptance Criteria

[Define the acceptance criteria in terms of maximum tolerance as a percentage of converted data volume for each severity level within a given business function. These acceptance criteria, besides being used for determining whether the converted data is ready for cutover, may be used to evaluate and monitor the progress of the data conversion process along the way. The Figure 2‑3 is an example which can be tailored for the current project.]

| ***ACCEPTANCE CRITERIA*** | | |
| --- | --- | --- |
| Business function (ranked) | Severity | MAXIMUM TOLERANCE AS % OF CONVERTED DATA VOLUME |
| ***1st Priority Business Function (Data Volume: ####)*** | *1* | *0.0%* |
| *2* | *0.3%* |
| *3* | *1.0%* |
| ***2nd Priority Business Function (Data Volume: ####)*** | *1* | *0.1%* |
| *2* | *0.4%* |
| *3* | *1.5%* |

Figure ‑: data conversion acceptance criteria

## Assumptions, constraints, and risks

This section describes the relevant assumptions, constraints, and risks concerning the scope, strategies, and goals of the data conversion effort.

[Most projects would never be able to get off the ground if they had to wait until every resource is available to begin. Therefore, assumptions, constraints, and risks form the basis for project planning and filling in the gaps between known facts and guesswork.

* Each **constraint** is a limiting condition, event, or circumstance, setting limits for the project. Constraints limit successful project execution. Constraints must be identified and incorporated into the project plan to ensure that the plan is realistic.
* Each **assumption** is a likely condition, event, or circumstance accepted as true without absolute certainty. Assumptions bring possibilities and allow the project to proceed. Assumptions must be analyzed and monitored to ensure validity and relevancy as the project proceeds.
* **Risk** may be defined as the chance or probability of something (e.g., a wrong assumption) that has the potential to cause the data conversion effort to fail or fail to meet one or more of its planned objectives such as scope, schedule, cost, or quality.

Defining assumptions, constraints, and risks is essential to project planning process. They are used to define and shape tasks, interrelationships, priority, schedules, effort, and resource and budget allocations.]

### ASSUMPTIONS

[Consider the following assumptions as some of them may be relevant to your current project:

* All severity 1 and 2 data exceptions originating from source data will be cleansed by the data cleanup team before the scheduled final data conversion execution.
* At the minimum, one key business subject matter expert per business domain will be available within 24 hours of the request.
* All environments (legacy, staging, and target) are fully documented (conceptual/logical data models, physical data model, business rules, and interfaces), available, and accessible by the data conversion team as scheduled.
* Only client records with “active” status as previously defined and agreed upon will be migrated over to new application system.
* A comprehensive and up-to-date data dictionary of the legacy data is available.
* A final set of relevant business rules will be made available to the data conversion team prior to the currently scheduled start date of the data conversion build phase.
* A draft version of the target data model will be made available to the data conversion team prior to the currently scheduled start date of data analysis and mapping, and the final version of the target data model is due before the currently scheduled date of the first mock conversion.
* Independent data validation is not within the scope of the data conversion team’s responsibilities.
* Only severity 1 and 2 data exceptions are required to be addressed before commencing the next data mock conversion.
* Data obfuscation is not within the scope of the data conversion team’s responsibilities.]

[Describe any relevant assumptions with respect to scope, strategies, and goals of the data conversion effort, particularly level of effort, schedule, resources, budget, dependencies, and quality control.]

The following assumptions are made with regards to the <<project >> and must be taken into consideration prior to the data conversion effort beginning.

| **assumptions** | **level of impact** |
| --- | --- |
| <<First Assumption>> |  |
| <<Second Assumption>> |  |

### constraints

[Constraints are limitations on the project. Typically, these include budget limitations, delivery deadlines, and contractual constraints. Some of the following constraints may be applicable to your project:

* Data requirements and definitions may require clarification by Subject Matter Experts (SMEs).
* Expertise in legacy data may be limited or unavailable due to lack of documentation (e.g., data dictionary) or more pressing priorities such as production support.
* Availability of SMEs may be limited or unavailable due to competing demands on their time.
* Quality control process and security requirements add additional time related overhead]

[***Describe any constraints that could have a significant impact on the data conversion effort, particularly with respect to effort, schedule, resources, budget, dependencies, and quality control.]***

The following constraints have been identified concerning the <<project>> and must be taken into consideration prior to the data conversion effort beginning.

| **constraints** | **severity level** |
| --- | --- |
| <<First Constraint>> |  |
| <<Second Constraint>> |  |

### Risks

[Risk may be defined as the chance or probability of something that has the potential to cause the data conversion effort to fail or fail to meet one or more of its planned objectives such as scope, schedule, cost, or quality. There are many risks inherently associated with moving data between computer systems or storage formats, not to mention the types and number of risks that may be involved in transforming and migrating huge volumes of data with many years of history from multiple sources and platforms, and in various storage formats, to a new computer system storage format. Therefore, it is important that all relevant risks are identified at the outset so that they can be qualitatively and quantitatively analyzed and mitigated. The following are some of the risks that may be relevant to the current project:

* The data conversion plan may not be feasible to achieve the expected goals and objectives because data conversion scoping was based entirely on theory and previous experience.
* Data conversion effort may not be able to meet the planned schedule because the quality of source data is unknown.
* Legacy data architecture artifacts may be unavailable or incomplete.
* The expense of overtime may be required to perform certain steps during non-business hours to reduce impact on the current production system.
* The team may encounter incompatible software, hardware, and/or processes due to: multiple operating systems or vendors, or format incompatibilities (Database Management System (DBMS) to DBMS, DBMS to Operating System, etc.)
* Data conversion effort may not be able to achieve the expected goals and objectives because the existing data conversion team does not have the necessary level of skills and experience to effectively execute the required tasks.
* Integrity and quality of the converted data may be compromised due to lack of data governance.
* Data quality of the target system may not meet the departmental standards due to lack of properly defined business rules.
* Data quality of the target system may not meet the departmental standards because independent data validation was not considered part of the data conversion scope of work.
* Source data may be inaccurately transformed and migrated due to lack of involvement of key business subject matter experts.
* Source data may be inaccurately mapped due to lack of or outdated legacy system data dictionary.
* Data quality of the target system may not meet the departmental standards because only a subset of the converted data was tested.
* Data conversion may not be ready for cutover as scheduled because no data conversion “dress rehearsal” was planned and included as part of the data conversion scope of work.
* Converted data may not be compatible, useable, or processed by the new application system because functional requirements and test were not part of the data conversion scope of work.]

[Describe all relevant risks, their probability, and the level of impact with respect to scope, strategies, and/or goals of the data conversion effort, particularly any risks related to funding, staff expertise and availability, schedule constraints, legacy environment complexity, hardware and software (conversion tools) availability, incompatibility between software and operating system, etc. The following tables may be used for this purpose, if appropriate.]

| **Probability rating** | **description** |
| --- | --- |
| *High* | *Likelihood > 80%* |
| *Medium* | *80% < Likelihood > 50%* |
| *Low* | *50% < Likelihood > 10%* |

| **impact level rating** | **description** |
| --- | --- |
| *Catastrophic* | *Failure of mission-critical services* |
| *Critical* | *Significantly degraded project performance* |
| *Marginal* | *Negligible* |

### Risk mitigation strategy

[It is important to detect and address risks, ideally during the planning phase and design phase of the project as they are less costly now than those discovered during the implementation phase.

For each anticipated risk that may jeopardize the successful completion of the conversion project, the project’s Data Conversion Management Team needs to develop a contingency strategy, which is an alternate method of accomplishing the same objective, or risk mitigation strategy, which is a means of reducing the impact of undesired results.

If the data conversion project is part of the overall legacy system modernization project, this risk mitigation strategy must be in alignment with the overall development effort.

A successful data conversion effort requires the identification of relevant and realistic risks as well as the mitigation of those risks to the business. Identifying these risks (e.g., dependencies on other teams within the overall project, minimal conversion expertise within the organization, or insufficient understanding of source data) and opportunities (e.g., the best-fit and most effective data conversion method) early in the project allows time and opportunity to properly address and manage them and less disruption later. The project team should consider the following options (if applicable) to reduce risks:

* Migrate only the data of “active” members/customers/applications/etc. to target application.
* Delay legacy system decommissioning to a later date.
* Address only severity 1 data quality issues before the final “dress rehearsal.”
* Analyze legacy data before finalizing the data conversion scope.
* Establish a separate team to independently validate and ensure that all data is migrated accurately and completely.
* Ensure that the integrity of the converted data is maintained.
* Ensure that a comprehensive set of business rules is gathered, documented, and signed off by business domain SMEs. These must be centrally located and shared among the data conversion team, data clean-up team, and data validation team.
* Ensure that acceptance criteria are clearly defined, measurable, and signed off by key data stakeholders.
* Ensure that data governance is established and dedicated to address data issues related to the data conversion effort.
* Prepare a detailed inventory of what data and systems’ architecture exist, and identify any data issues relevant to the conversion during the early phases of the project
* Ensure all resource dependencies, such as access to and availability of environments (legacy data, staging, and target system), tools, software licenses, or personnel are thoroughly identified.
* Identify and secure primary and secondary SMEs with knowledge of and experience with the legacy data.
* Schedule recurring risk identification and brainstorming sessions to actively identify potential challenges and opportunities associated with each data conversion deliverable and dependency during the planning phase. This minimizes risks or challenges being introduced at later phases and allows time to proactively address these challenges and opportunities now.]

[Provide a mitigation strategy for each of the anticipated risks identified. The following table may be used for this purpose.]

| **risks** | **Probability** | **impact level** | **Mitigation** |
| --- | --- | --- | --- |
| <<First Risk>> | High/Medium/Low | Catastrophic/Critical/Marginal |  |
| <<Second Risk>> | High/Medium/Low | Catastrophic/Critical/Marginal |  |

## communication strategy

[Communications pertaining to the data conversion effort need to be shared among the data conversion team members, key data stakeholders, and external personnel (when necessary). It is important that all key data stakeholders are consistently kept informed of the state of the data conversion effort. Therefore, it is necessary to develop a communication plan to identify what information to share, who the recipients are, when or how often the information will be available (frequency/schedule), what format it will be in, and how the information will be delivered.]

[List all the recipients, schedule, format, and delivery method of each document prepared and shared as a result of the data conversion effort. The following table may be used for this purpose.]

The following are the documents that the project team will prepare and share with the corresponding recipients according to the schedule, format, and delivery method.

| document | recipient | frequency/ schedule | format | delivery method |
| --- | --- | --- | --- | --- |
| Status Report | Key data stakeholder distribution list | Weekly/ Monthly | PDF | Email |
| Data Conversion Deliverables |  |  |  |  |
| Data Profiling Findings |  |  |  |  |
| Data Cleansing Reports |  |  |  |  |
| Data Validation Reports |  |  |  |  |
| Escalation Procedures |  |  |  |  |
| Conversion Design Decision |  |  |  |  |
| Business Rules |  |  |  |  |

# conversion Requirements

[Accurately defining and clearly understanding each business and technical requirement for data conversion is crucial for by these requirements the project team can determine what data to migrate, whether or not archive data is included as it may require a different conversion strategy all by itself, what acceptable downtime is to the business during cutover, etc. These requirements may take the form of agreements, expectations, and/or objectives of the conversion.]

## business requirements and expectations

[The project team should consult with the business or key data stakeholders and business SMEs to determine if there are any additional requirements that they might impose above and beyond the technical and security requirements. The following questions can be used to help determine business requirements and expectations for the data conversion effort:

* What are the source data (any historical data) to be migrated?
* What is the size of the legacy source dataset?
* How many source systems are involved?
* Are there any specific data availability requirements and/or performance Service Level Agreements (SLAs)?
* Which data elements/datasets are most critical to the target application system?
* Are there any datasets or data categories that are not part of this conversion?
* Are there any production applications that may conflict with the conversion?
* If data conversion is a part of an overall legacy system modernization project, the project team will need to know whether or not converted data is expected to be used by the business during User Acceptance Test (UAT), system readiness verification, reporting, and/or at any other time during conversion. If the answer is “yes”, inherently there will be additional workload that may have not have been accounted for in the original data conversion scope since there are dependencies, checkpoints, risks, assumptions, additional requirements (data volume, types, data condition, readiness verification), timelines, resources, etc. that the data conversion plan must consider.
* Are there objectives for better data quality and/or greater technical flexibility or stability?
* What are the expectations regarding “dirty data” currently residing in the legacy application systems?
* What are the expectations concerning data exchange associated with interfaces?
* How much production downtime is acceptable to the business? This is a window of time for the data conversion team to prepare for cutover. (Depending on the duration allowed for downtime, this could result in additional workload for the data conversion team to ascertain that all final data conversion activities can fit within the cutover window.)
* What are the expectations concerning processes to be used for data conversion testing, data validation of mock conversion along the way, and final data validation at cutover?
* What is the expected timeframe for the overall data conversion effort?
* Are there any future business objectives that may require considerations for growth and scalability?]

[Describe all relevant business requirements and expectations for the data conversion effort. The following table may be used for this purpose. ]

| **Business requirements & expectations** | | **priority** |
| --- | --- | --- |
| 1 | << Requirement/Expectation 1>> |  |
| 2 | << Requirement/Expectation 2>> |  |
| 3 | << Requirement/Expectation 3>> |  |

## Technology and INFRASTRUCTURE considerations

This section describes the hardware and software that are necessary to effectively facilitate data conversion activities.

[In every phase of the data conversion lifecycle, most of the work efforts involved can be automated to some extent and in some cases automation is the only method to meet or satisfy requirements and/or expectations. Therefore, a clear understanding of the entire data conversion process will help determine the best-fit technology to be used at each stage in the data conversion process. Figure 3‑1 attempts to associate each phase of the data conversion process with a set of technology tools that are necessary to achieving the work required. The technology tools outlined in the diagram are specific to data management and most of these are commercially available. The purpose of this diagram is to provide a high-level visibility and understanding of the entire data conversion process in an effort to help the project team properly determine the technology tools that are necessary to effectively facilitate data conversion activities.

Areas to be considered in determining the best-fit conversion technology and infrastructure for your project are:

* Data profiling, data quality, data obfuscation, data modeling
* Metadata management software (e.g., Altova MapForce, Astera **)**
* Data Extract, Transform, and Load (ETL)
* SQL Editor or Developer IDE software (e.g., TOAD, Visual SQL Editor)
* Data modeling and re-engineering tools (e.g., ERWIN, Database Workbench, ER/Studio)
* Performance, stress, and load testing tool
* Certain data conversion software does not support legacy operating systems or may not be compatible with the architecture of the staging environment. Therefore, a clear understanding of the technology infrastructure currently in place and any future requirements of the target environment (e.g., such as scalability, extensibility, and accessibility) will help shape decisions made during data conversion.
* Data structures (i.e., how data is being stored) - are the data structures to be used for target data identical or similar to the current legacy data structures? This is one of the major factors that will affect the time and complexity of the conversion effort.
* Numbers of interfaces and changes to the Electronic Data Interchange (EDI) specifications, network, and communication environment with respect to the target system.
* Network, hardware and software configuration, and network communication protocols.
* If staging environment will be used, what requirements have already been defined and what recovery or failover requirements exist?
* If there are expectations that data integrity will be monitored and reported along the way, the project team will need to consider the appropriate tools to help facilitate this activity.]

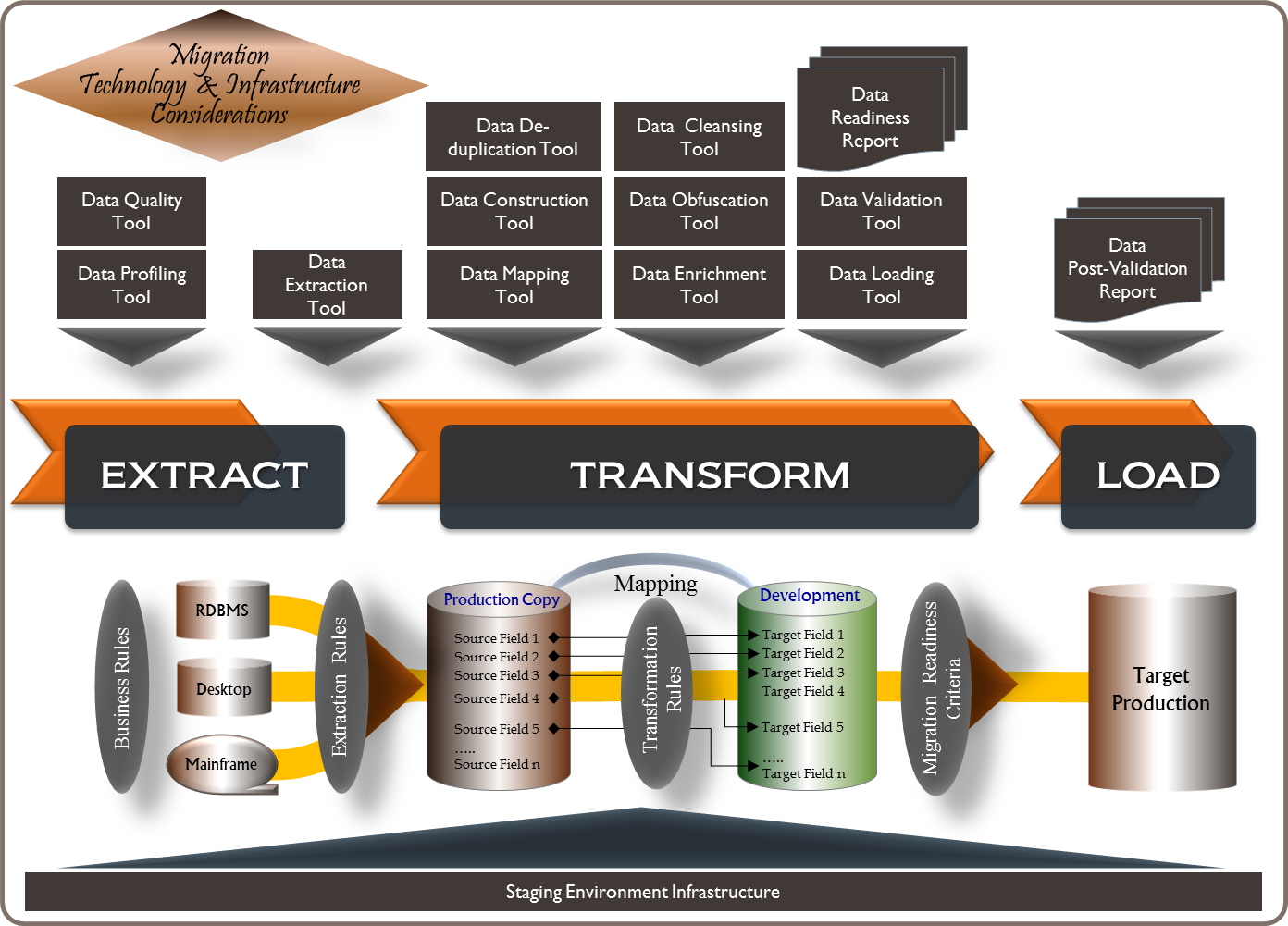


Figure ‑: Technology & infrastructure considerations

[Define all the relevant conditions or capabilities to be satisfied by technology.]

| **technology & infrastructure requirements** | | **priority** |
| --- | --- | --- |
| 1 | << Requirement 1>> |  |
| 2 | << Requirement 2>> |  |
| 3 | << Requirement 3>> |  |

## Data security and privacy requirements

This section describes all relevant data security and privacy requirements to which the data conversion effort must adhere. It also describes the process to be used to ensure that data confidentiality will be protected throughout the data conversion effort.

[What are the organization’s standards, practices, procedures, or policies concerning data security and privacy? Specifically, to which of these must the data conversion effort adhere? Areas to consider are:

* Health Insurance Portability and Accountability Act (HIPAA) Privacy and Security, including Personally Identifiable Information (PII) and Protected Health Information (PHI) related policies, procedures, and/or practices.
* Standards and requirements regarding masking of PHI data.
* If converted data is planned to be used during UAT, does the data need to be obfuscated or masked? Understanding the requirements concerning data security and privacy at the outset will help the data conversion team to prepare, plan, and manage accordingly.]

[List all the relevant data security and privacy requirements that the data conversion effort has to adhere to and describe the process to be used to ensure that the confidentiality of data will be protected and other activities such as validation and testing of converted data are still enabled. The following table may be used for this purpose.]

| **data security & privacy requirements** | | **priority** |
| --- | --- | --- |
| 1 | << Requirement 1>> |  |
| 2 | << Requirement 2>> |  |
| 3 | << Requirement 3>> |  |

## conversion data sources

The data sources identified and listed in the following table are considered to be the only sources from which data will be extracted, transformed, and loaded into the target system:

[Be sure the list of data sources from which the data will be extracted, transformed, and loaded to the target system is clearly defined and approved by signature of relevant stakeholders. Without having a clearly defined and approved list of data sources with detailed description of each data source, the data conversion plan is simply incomplete and inaccurate. It is the responsibility of the project team to ensure that every data source, whether it is an Oracle database or an MS Excel spreadsheet that users have been using for years, is identified, clearly documented, and approved.]

[List all the data sources with pertinent details that are in scope for data conversion. The following table may be used for this purpose.]

| **DATA SOURCE** | **description** |
| --- | --- |
| << Name of Data Source 1 >> or  << Dataset / Record set >> | Platform: Oracle, MS Access, MS Excel, ADABAS, VSAM, ETC |
| Volume/Population: # of tables & rows, # of records, etc… |
| Business Function(s): |
| Key contact: |
| Data Owner: |
| Notes: |
| << Name of Data Source 2>> | <<Description>> |

# current Environment analysis

This section describes the strategy that will be used to gain visibility and have clear understanding of the functional and technical aspects of the current legacy environment.

[“The ultimate security is your understanding of reality.” - H. Stanley Judd.

Understanding both the functional and technical aspects of the current environment is essential to the success of data conversion. This is one of the key reasons that many data conversion projects end up taking far longer than necessary or completely fail to deliver: a lack of essential detailed analysis of the current environments during planning.

The process of gathering information about the current environment and updating or creating documentation where required information is not available can be very time consuming. However, the outcome of this process is crucial to the success of the data conversion effort as it is the key ingredient to scoping, planning, and decision making which directly affects every aspect of the remaining phases of data conversion. Some of the key information essential to the current environment analysis is:

***ENVIRONMENT ANALYSIS***

* Identify and analyze the Logical Data Model (LDM) and Physical Data Model (PDM) of source systems, including the data dictionaries for each data element to be migrated. This information is necessary to determine the current state of the data architecture for each source system. The LDM represents the abstract structure of the business data or information organized in terms of entities and relationships involved in a business function, whereas the PDM presents the physical structure of the actual data structure.
* Review relevant information about every application or system (both internal and external) interfacing with the source systems. This information is very important in coordinating downtime of the source systems, testing, and cutover.
* Identify known technical constraints of the source systems as well as issues and/or concerns regarding data integrity and data quality of the current source system that are expected to be resolved by the data conversion effort.
* Determine data retention policy, data access control, and security requirements.
* *Network connection between data points (such as the source system to production-copy environment; production-copy environment to staging environment; staging environment to test environments; and so on) need to be assessed and clearly understood particularly bandwidth, schedule, and security controls.*

***DATA ANALYSIS***

* Perform data profiling analysis of the legacy source data.
* Perform data quality assessment of the legacy source data based on relevant and current business processes and business rules of the current system and target system including any new or updated changes. This process involves a number of tasks to be accomplished in order to analyze the data, such as:
  + Analyzing the target database structure.
  + Collecting and analyzing samples of the legacy source data for possible data discrepancies and potential problem areas. These issues usually arise from missing fields, incomplete data, duplicate data, incorrect data, or non-standard characters in standard fields.
  + Identifying any specific application business functionality that may cause discrepancies in the conversion.
  + Developing metrics and data audit and reconciliation reports based on the quality of the legacy source data.

***DATA PROFILING***

Data profiling should be the first step of any data conversion project as it is the most effective and practical way to have visibility and understanding of the current data sources before any data conversion planning activity begins. Figure 4‑1 provides an example of a data profiling statistical analysis and assessment report. Data profiling process assesses the data structures and data content of the source systems to understand and determine any challenges regarding data integrity and relationships between data. Typically, there are two types of data profiling:

* **Metadata standard profiling:** analysis of the data structures in place. The outcome of this activity can be used to evaluate compliance with department-wide standards.
* **Content profiling:** analysis of the data content. The outcome of this activity reflects the quality of the content of the data captured and is channeled to the data cleansing effort for resolution.



Figure ‑: data profiling report

**The major processes involved in data profiling include:**

* **Column Analysis**: provides critical metadata which is required in order to perform functional dependency analysis, and as such, must be executed before functional dependency analysis. Column analysis provides important statistical properties for each column. This information includes cardinality, selectivity, null value rates, most common value, unique values, and value frequencies.
* **Functional Dependency Analysis**: compares different columns in a single row to determine if there are data dependencies between different column values. Dependency profiling relates to the normalization of a data source, and addresses whether or not there are non-key attributes that determine or are dependent on other non-key attributes. The existence of transitive dependencies here might be evidence of second-normal form.
* **Redundancy Analysis**: Identifies overlapping values between tables. This is typically used to identify candidate foreign keys within tables, to validate attributes that should be foreign keys (but that might not have constraints to enforce integrity), and to identify other areas of data redundancy.

**The purpose of gathering these statistics is to:**

* Provide metrics on data integrity, including whether the data conforms to department standards.
* Evaluate the risk involved in integrating data for the target system, including the challenges of joins.
* Evaluate whether metadata accurately describes the actual values in the source database.

[Describe the strategy that will be used in order to gain visibility and have clear understanding of the functional and technical aspects of the legacy environment.]

# data cleansing

This section describes the process that will be used to facilitating data cleansing.

[Data cleansing is required to ensure that legacy system data conforms to the rules of data conversion. This process may involve manual and/or automatic updates to legacy system data. Data cleansing should be an ongoing business activity and as long as the legacy systems are active, there is the potential that previously cleansed data issues are reintroduced.

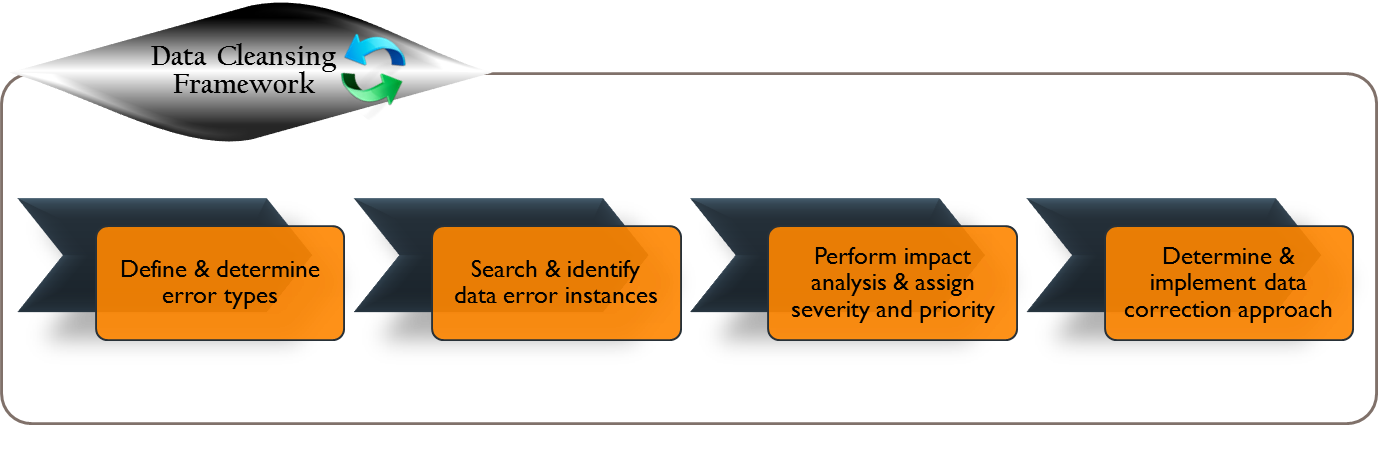


Figure ‑: data cleansing framework

As shown in Figure 5‑1, the data cleansing framework consists of:

* Defining and determining error types.
* Searching and identifying data error instances.
* Performing impact analysis and assigning severity and priority.
* Determining and implementing data correction approach.

It is recommended that data cleansing be performed in the legacy systems prior to data conversion or in the target system post conversion if necessary. Otherwise, it will be very difficult for data reconciliation if data cleansing is performed while the data conversion process is underway.

The primary purpose of data cleansing with respect to data conversion is to address first and foremost legacy data issues that will not work properly in the target application. The work needs to be done up front so that data can be extracted, transformed, and loaded into the target data storage without any intervention required at final conversion time. Data cleansing is one of the most important aspects of a data conversion project since loading “dirty” data into the target system may cause the target application not to function as designed, resulting in incorrect business decisions and greater difficulty to correct later. Figure 5‑2 outlines the steps involved in the data cleansing process.

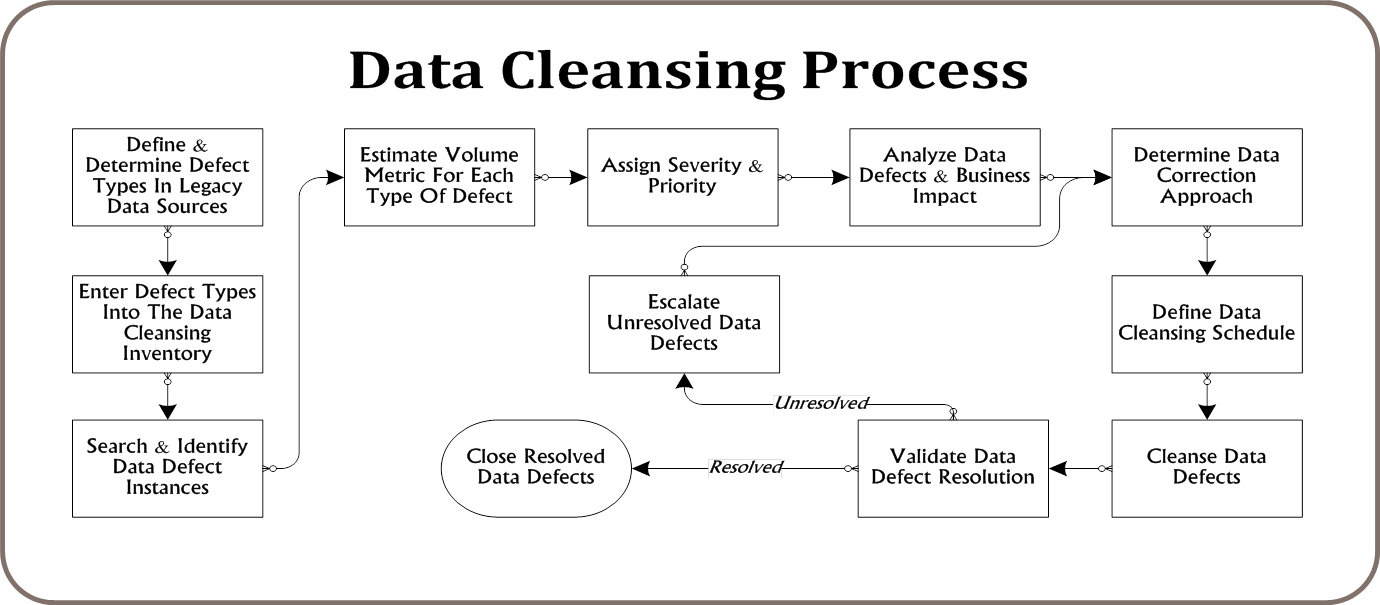


Figure ‑: detailed data cleansing process

There are many different types of data issues requiring cleansing. For example:

* Duplicates - multiple records for the same person, same account, same contract number, or same company.
* Inconsistency in similar data - similar data stored in different formats or different abbreviations across multiple legacy systems.
* Free form text fields – due to limitations within the legacy systems, text fields might be used to store important business information such as reasons for disqualifying an applicant, reasons for increasing the person’s salary, etc.
* Incompatible data values – data values that will fail to load into the target data storage due to data type or format incompatibility, length, lack of an acceptable value in the target system, etc.
* Missing required data values – a data field in the current system is either optional or mandatory but not enforced, hence, intermittent data. However, this field is required in the target system.
* Overloaded data fields – same data fields being used by different divisions, sections, or business functions to store different elements of information.
* Compound data fields – data fields being used to store multiple related data elements (e.g., a data field “contact name” contains both name and phone number).]

[Describe the data cleansing approach that will be used to address the identified data issues. Using the data conversion process shown in Figure 5‑2 and the questions below may help you determine the best-fit strategy for your project:

* Where do the identified data issues originate? Are they caused by missing or incorrect validation rules? Are these data issues part of an “active” dataset? Who is the data owner and who are the domain experts?
* What is the most effective approach to address these identified data issues (while data conversion is underway)? Can the correction be performed in the legacy production environment? If yes, will the correction set off any unwanted production processes to run?
* What are the criteria, timing, priority, and/or constraints (time, budget, resources, etc.) that can be used to appropriately categorize each dataset requiring cleansing? The following table can be used to help classify different conversion categories.]

| **data conversion category** | | | | |
| --- | --- | --- | --- | --- |
| conversion category | Business Impact | timing-priority | Criteria | Environment to Cleanse |
| A | 1 | Pre-Cutover-1  Pre-Cutover-2 | “Fail to load”, “Severely impact business services within first week after cutover”, etc. | Legacy Source |
| B | 1 | Post-Cutover-1 | “Fail to process”, “To be fixed as much as possible by cutover,” “Severely impact business services within second week after cutover; to be fixed manually one record at a time by business staff, ” etc. | Production-copy/ Staging area/  in-flight as part conversion process |
| C | 2 | Post-Cutover-2 | Data issues can be fixed by business staff as part of their regular business process | Target post conversion |

# conversion approach

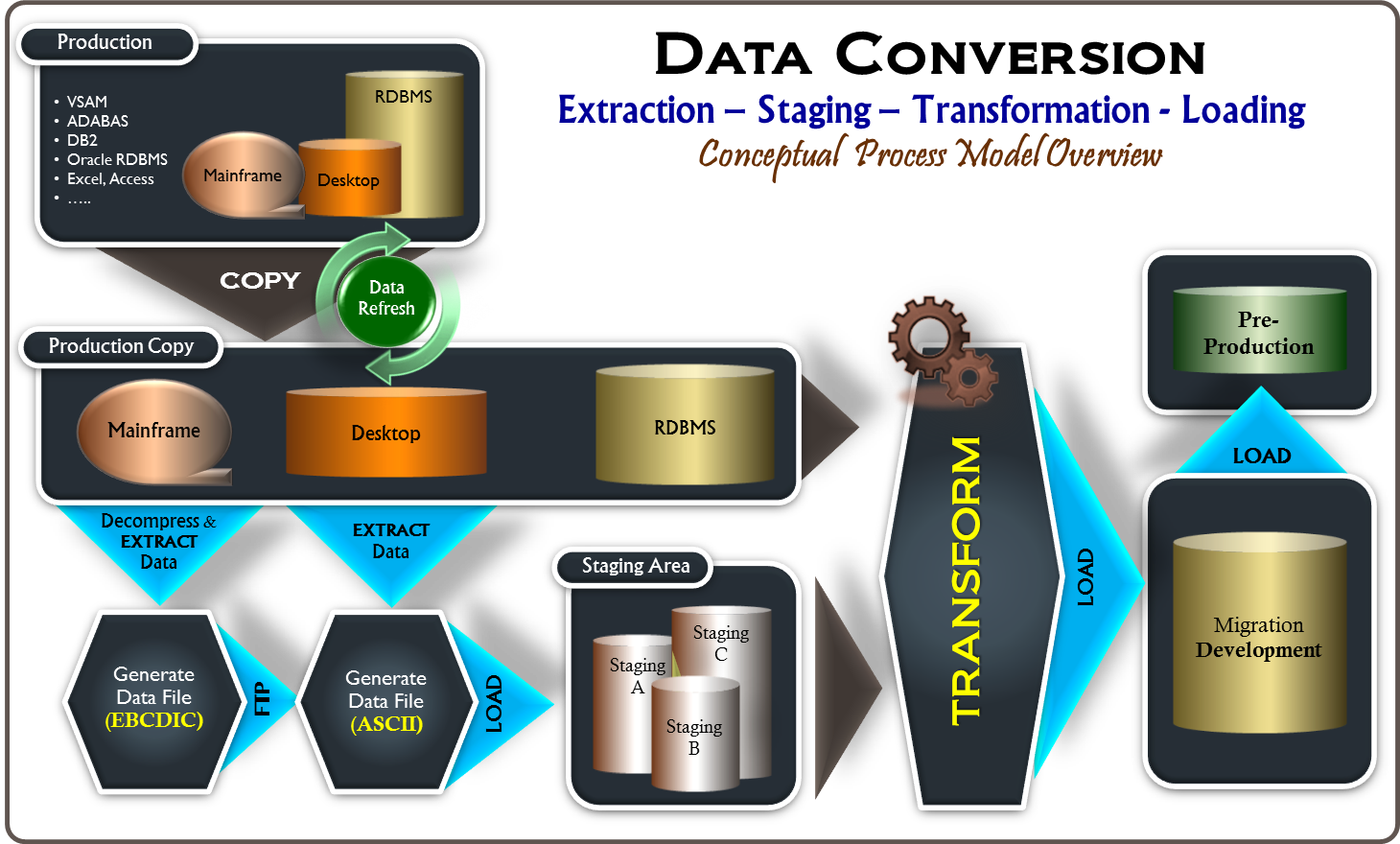
This section describes the approach to data conversion. 

Figure ‑: data conversion process overview

## methods of conversion

This section discusses the conversion method that will be used for migrating legacy data to the new target system.

[Specific data conversion methods and toolsets typically will vary based on different aspects of the project such as:

* Project scope and legacy data sources to be migrated.
* Diversity in the legacy data sources.
* Level of completeness and currency of the legacy data dictionary and business rules.
* Similarity and relationship between the legacy and target data models.
* Target data model complexity.
* Volume of data, if there are performance constraints.

**Manual vs. automated conversion**: most conversion will be accomplished through some form of automated processes. However, there may be some datasets where manual conversion is a more effective approach. The criteria for deciding which datasets warrant manual conversion will be different for each project. For some, it may be because the size of the record set is too small or involves manual research to determine the appropriate transformation and reconciliation rules for each different record type. In either case, automated conversion may not even be possible or may take more time to develop all the necessary automated processes than to convert the data manually. To ensure a smooth conversion cutover, it is recommended that specific plans be developed, verified, and factored into the overall conversion plan for every dataset requiring manual conversion.]

[Describe the overall approach to data conversion. The following should be considered and addressed in this section or the following subsection, if applicable:

* Conversion Category – If applicable, define the conversion categories and for each conversion category, the criteria that will be used to identify what datasets will be converted first, second, third, etc. This will help the project team prioritize the work and ensure that the work is in alignment with business expectations and other project teams, such as the application development team, as they may have specific application functions that require the use of converted data for testing or user acceptance test sooner than the other application functions.
* Manual Conversion - Define the criteria that will be used to identify what datasets will be performed manually and describe the overall approach to manual conversion.
* Automated Conversion - Define the criteria that will be used to identify what datasets will be accomplished by automated conversion and describe the overall approach to automated conversion.]

## data mapping

This section describes the process that will be used to determine what functionality will be implemented in the target system and what legacy data or data sources are necessary to support the functionality in the target system. It also discusses the process that will be used to determine how data will be mapped to the target system and what conversion programs must be developed to extract, transform, and load the target system.

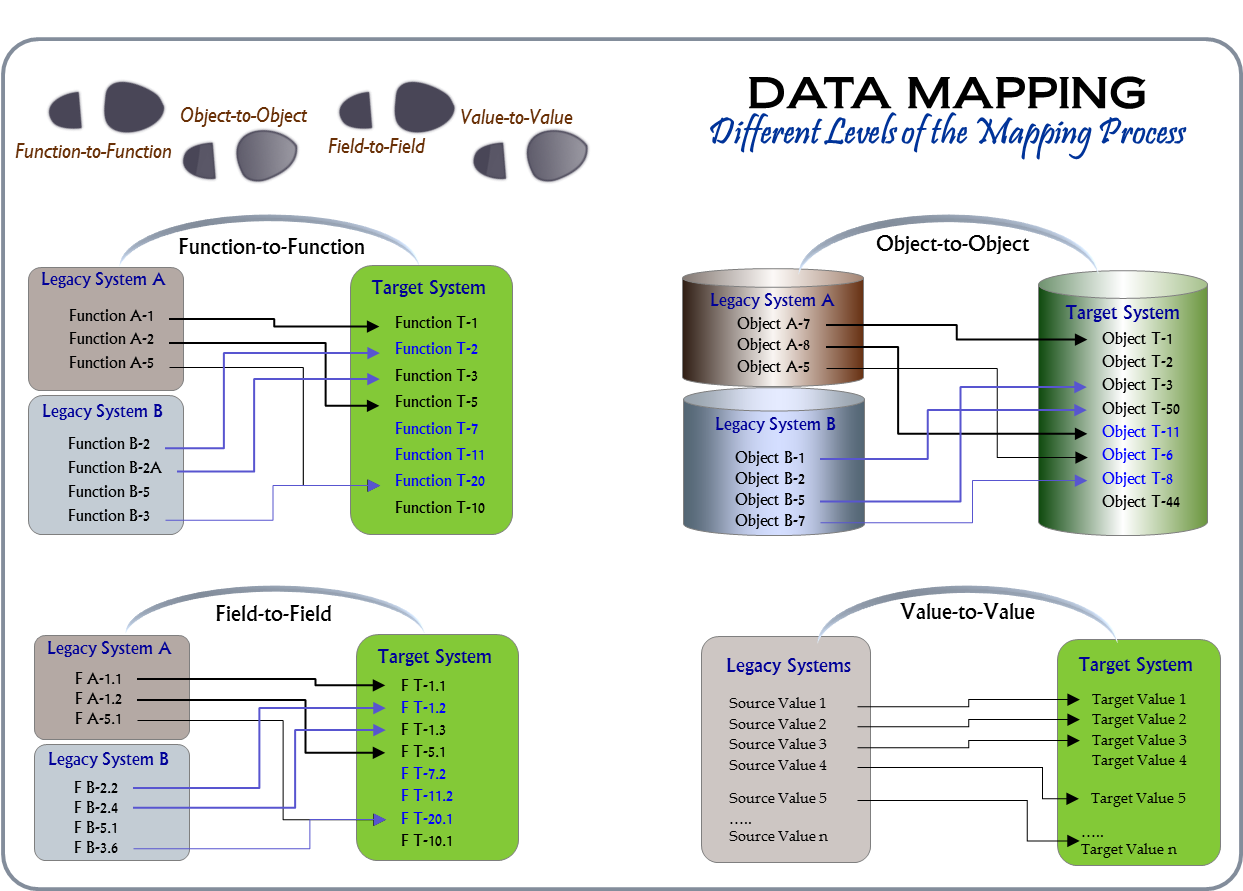


Figure ‑: different levels of the mapping process

[Data mapping is the process of linking data elements between the legacy system data models and the target data model. Data mapping is the fundamental first step in data transformation in that it captures the transformation rules between the legacy and target data models. It also determines the relationship between the data elements of the target and legacy systems and establishes instructions for how the legacy data is transformed before it is loaded into the target system.

Figure 6‑2 illustrates different levels of the mapping process. Once the project has confirmed what functionality will be implemented in the target system, it is recommended that the project begin mapping existing functionality in the current legacy systems to the target system. Next, the project should begin mapping objects of each function in the current legacy systems to the corresponding objects of functionality in the target system; this is followed by mapping the fields of each object in the current legacy systems to the corresponding fields of object in the target system. Finally, map the value(s) of each field to the corresponding field value(s) in the target system, if applicable.

Data mapping is the most critical activity that contributes to the data conversion effort going off-track. This is because data mapping is a manual, labor intensive process and requires multiple testing cycles to get it correct. Furthermore, changes to mappings are difficult to manage among the data conversion team members working in silos. Depending on data architectures of the target and legacy systems and the condition of the legacy data sources, transformation logic can also be very complex. These conditions pose high risk to data conversion efforts in terms of data quality, schedule slippage, and cost overruns. Therefore, it is recommended that the project team research some of the tools that are commercially available to help facilitate data mapping activities.

Moreover, because there are many conflicts among the data in different legacy data sources, it is recommended that a centralized data dictionary be established to facilitate data reconciliation. A data dictionary is a centralized repository of information about data, such as its relationship to other data, related business rules, its format and default values. It is often housed within the staging area for the duration of a data conversion effort. Data mapping is a complex task that requires attention to detail and a comprehensive understanding of source and target data models. The information contained in the data dictionary will be useful for data enrichment, transformation, reconciliation, and data cleansing.]

[Identify what functionality will be needed in the target system and the legacy data sources that are necessary to support the functionality in the target system. Determine how the data will be mapped to the target system and what conversion programs will need to be developed in order to extract, transform, and load the target system. Also, describe the process that will be used to facilitate the data mapping process.   
Since the function mapping will continue to be revised throughout the data conversion lifecycle, rather than providing the information in the table below, it may be more appropriate to have it included as an appendix or an attachment. A data mapping template is also provided as part of this document and a link to it is listed in Appendix B.]

## Data extraction and staging process

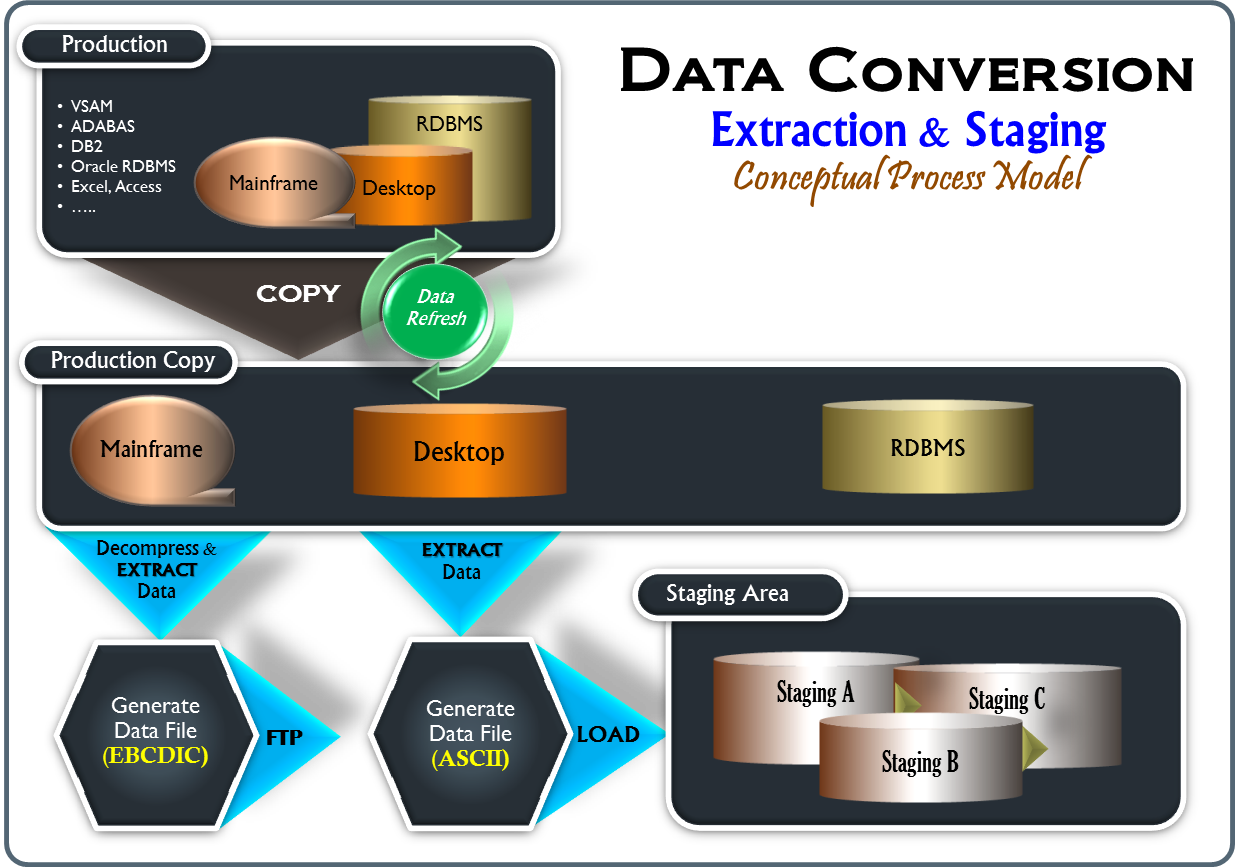
This section describes the approach that will be used to extract and stage legacy data. 

Figure ‑: extraction & staging conceptual process

[Data extraction and staging process describes the key steps of extracting data from the legacy systems and placing them in a universal data store in a homogeneous format for further interdependency analysis and transformation. Figure 6‑3 illustrates a conceptual process of data extraction and staging.

Because most data sources reside in diverse environments and formats (e.g., VSAM, ADABAS, MS Excel, MS Access, Oracle, DB2, etc.), it is recommended that, especially for mainframe data, **all** data from every data source within the data conversion scope first be extracted in its raw, native, stored form and placed in a universal data store in a delimited flat file format. The purpose is twofold:

* Data stored in a delimited flat file format can be easily accessed by most technology available via an import function. Moreover, since the target data structure is still unstable during this time for most projects, the data conversion team can still make substantial progress on the data extract development.
* Extracting every known data element from each of the data sources that are in scope is critical. Having every data field extracted really saves time for the data conversion team as they do not have to go through the process of determining whether or not a data element is needed before writing the extraction code. In addition to that, they do not have to modify the extraction code later for any additional data elements that were initially overlooked.

The followings are the recommended approaches to data extract and staging with respect to the three most common legacy data formats:

***Mainframe Data***

*For mainframe data (e.g., VSAM, ADABAS, etc.), it is recommended that the data be extracted in its raw, native, stored form and placed in a universal data store in a delimited flat file format. These data files will then be converted to ASCII format, which in turn will be loaded into the staging data storage.*

***Desktop Data***

*For desktop data such as Microsoft Excel, Access, and other similar formats, it is recommended that the data be extracted directly into a data file in ASCII format and loaded directly into the staging data storage without any transformation.*

***Relational Data***

*For relational data (e.g., Oracle, DB2, etc.), since the data is already in relational format, it is recommended that the data be copied over to the staging environment but not staged to avoid data redundancy.]*

[Describe the specific approaches that will be used to extract and stage legacy data that reside in diverse environments and formats such as VSAM, ADABAS, MS Excel, MS Access, Oracle, DB2, etc. Describe the process that will be used to specifically extract and stage mainframe data, desktop data, and relational data (if applicable). The following questions may be used to help formulate your data extract and staging approaches:

* Source Data – what are the data source environments and how will the data be extracted from each environment?
* Staging - once the data from each environment is extracted, what happens next? Where and how will it be staged? Are there any other data processes?]

## Data transformation and LOADing process

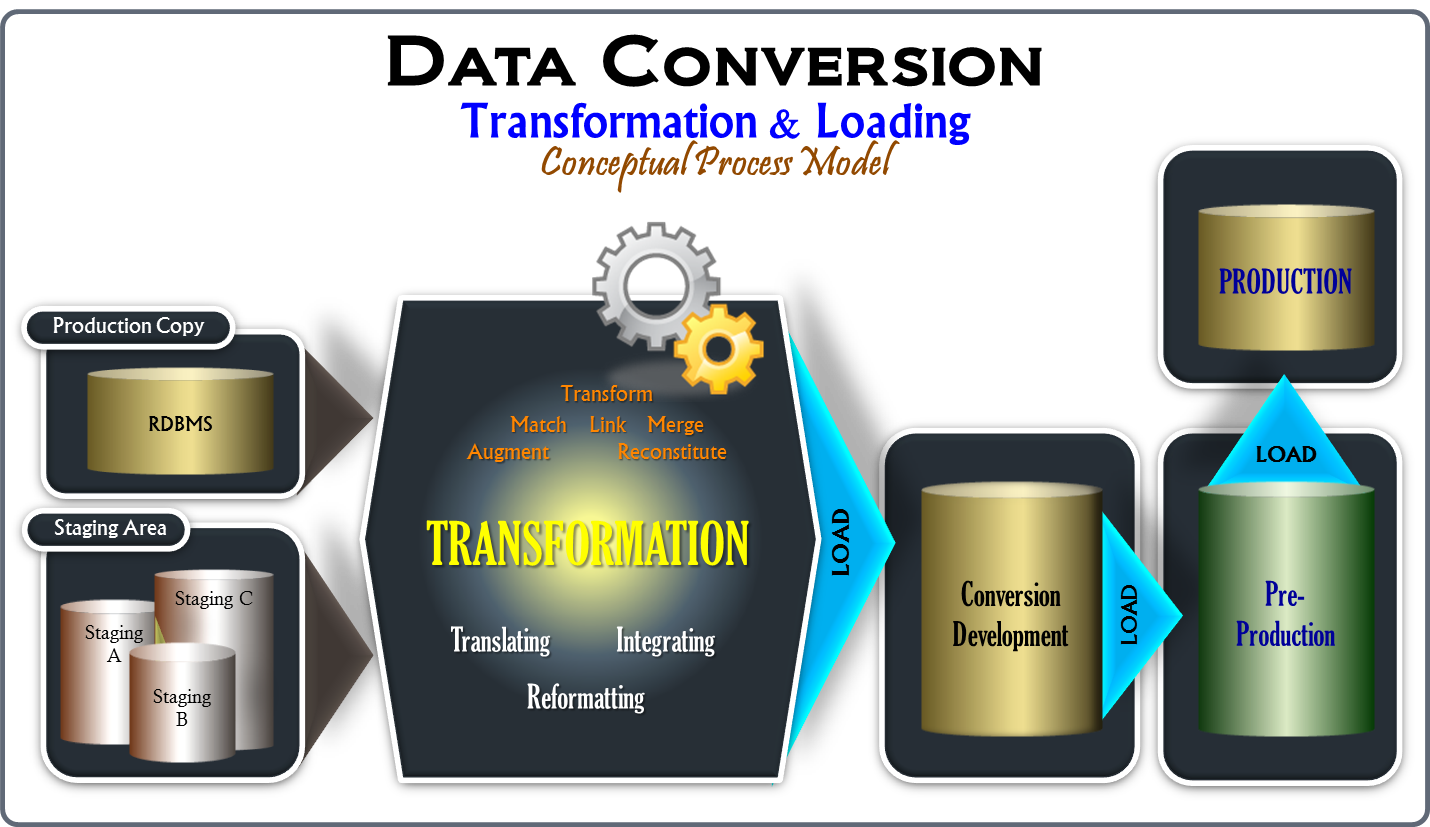
This section describes the approach that will be used to transform and load legacy data to the target data store. 

Figure ‑: transformation & loading process

[The data transformation and loading process incorporates all knowledge about value translation, business logic, and the understanding of both the legacy system and target system. Figure 6‑4 illustrates a conceptual process of data transformation and loading. This process matches and links data of different sources currently housed in the staging environment and transforms the data, as needed, to fulfill the target system functionality. The following are some of the common transformation types and key considerations relating to data transformation:

**Transformation types:**

* **Reformatting:** Format revisions are very common in data conversion as the data must conform to the format required by the target system. Format revisions include changes to the data types, length, and case (e.g., alphanumeric to numeric, integer to decimal, 40 characters to 25 characters, lower case to mixed case, etc.) .
* **Translation:** Due to the difference between the source and target data models, data in the source system may have to be reconstituted, translated, enhanced, or converted during the conversion process to conform to the target data requirements. Take for example gender “male” in the source data might be represented as “1” but “M” in the target data. Therefore, formulating correct mapping and transformation rules to convert the data is one of the key tasks in a data conversion effort.
* **Integration**: When data is being migrated from multiple legacy sources into a single target system, undoubtedly there will be various conflicts among the data and inconsistent representations of the same data (e.g., the same person or entity might be represented differently in different legacy systems – “John Smith Jr.”,” John S. Jr.”, “John Smith”, “LA City”, “Los Angeles City”, “LA”, etc.), and that must be reconciled in order to satisfy the target data consolidation requirements and to have an integrated and reconciled view of data of the department.

**Key Considerations:**

* *Legacy data is often duplicated across multiple legacy data sources and similar data attributes may have conflicting values among the legacy data sources.*
* The target system might have new attributes that never existed in the legacy systems. In addition, data attributes from multiple data sources might be required to be merged under one or more entities.
* *Legacy data might have defects resulting from inaccurate processes; for example, account balances and interest calculations might not match detailed transactions.]*

[Describe the process that will be used to transform legacy data sources currently housed in the staging environment and load converted data to the pre-production environment (if used) and then production environment at conversion cutover. The following questions may be used to help formulate your data transformation and loading approaches:

* Transformation –What does the transformation process entail? Are business rules, mapping and transformation rules available and being used during this process?  
  What data integration strategy will be used to resolve conflicts and duplication among the legacy data sources? Considerations should be given to some of the following common conditions and characteristics of legacy data:
  + Data that are duplicated across multiple legacy data sources
  + Data that have conflicting attribute values across multiple legacy data sources
  + Data with attributes or content that is incompatible with the new system
  + Data with invalid context due to historic changes to operational parameters
  + Data with defects resulting from inaccurate processes
  + Data with invalid record relationships
* Loading – If the strategy includes the use of a staging area as an integration platform to allow validation, cleansing, and/or conversion of the integrated data, what is the approach to load converted data from the staging environment to the pre-production environment (if used) and then production environment?]

## Synchronization Process

This section describes the approach to be used to keep data conversion in sync with other relevant efforts.

[Since a data conversion effort is often part of an overall legacy system modernization project, many project related activities take place concurrently. Particularly for the data conversion effort, unless legacy data environments are frozen and both logical and physical data models of the target system are stable, any structural changes in the models will have a direct impact to the data conversion effort and the larger project as well. ]

[Define the approach to be used to keep data conversion in sync with other relevant efforts. Be sure to clearly define the specific and relevant conditions (e.g., request type=data structure, table update, etc.) that you want to have the synchronization process trigger. Otherwise, you will be inundated with many change requests that are not related to data conversion.]

## conversion reversal strategy

This section describes the process that will be used to roll back the legacy data in the event the data conversion process does not go as planned.

[Conversion reversal is the process by which data conversion can be reversed. Data conversion can fail either because the legacy data could not be converted within the planned conversion cutover window or because the dependencies across legacy processes, systems, and data had not been factored.]

[Describe the process to restore the source data if the need to revert to a previous back-up is identified at any point during the conversion process. Also, determine what precautionary measures can be planned and implemented to minimize the need for conversion reversal. Some suggestions are as follows:

* Define stopping points and check points to terminate data conversion under predefined conditions and rollback conversion to predefined segments, if needed.
* Determine and plan conversion dependencies, schedule, and execution sequence of data conversion programs.
* Tune and optimize data conversion programs continually to ensure the entire data conversion will complete within the time allotted for conversion.
* Determine and plan for hardware and network resources to support data conversion within the cutover window.**]**

## **DATA** CONVERSION SCHEDULE

This section provides a schedule for data conversion activities to be accomplished.

[Unless “big bang” is the decided cutover approach, data will be converted in phases. Therefore, special consideration should be given to what datasets will be converted and rolled out in each phase based on an agreed upon priority and timeline. It is recommended that a high-level conversion timeline or schedule with relevant information be provided in this section or subsection and a specific detailed rollout plan for each phase to follow.

Data conversion activities will be impacted by dependencies across various processes in the legacy systems including data cleansing. In order to lessen the impact, it is preferable to have as many data defects cleansed as possible prior to conversion. Moreover, as long as legacy data remain unfrozen, data conversion will continue to be constrained and/or impacted by the on-going generation of new data in the legacy production systems and especially the potential changes to the legacy data models due to critical business needs.]

[Provide a schedule for data conversion activities to be accomplished in accordance with this Data Conversion Plan. Show the activities in chronological order, with beginning and ending dates of each task, the key person(s) responsible for the task, dependencies and milestones. Make certain that the data conversion schedule is appropriately integrated into the overall project schedule. Since the schedule will continue to be revised throughout the lifecycle, rather than providing the schedule in the table below, it may be added as an appendix or an attachment.]

| data conversion schedule | | | | | |
| --- | --- | --- | --- | --- | --- |
| Activity | Description | Begin Date | End Date | Key Person(s) Responsible | Dependencies |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

## data conversion MILESTONES

This section lists the major milestones pertaining to the data conversion effort.

[There are many milestones associated with data conversion, some of these are:

* Analysis and profiles of legacy source data and reports have been completed.
* Data conversion plan has been written and approved.
* Data conversion acceptance criteria and performance metrics are clearly documented, accepted, and evaluated.
* Audit and reconciliation report requirements are defined.
* Independent data validation effort has been established.
* Data cleansing effort is underway.
* Data conversion requirements (business and technology) are clearly documented.
* Data conversion team is formed and roles and responsibilities and task assignments are clearly defined and accepted.
* Source-to-target data mapping is complete.
* A detailed plan for freezing physical data structures of the source system during cutover is documented and accepted.
* Conversion tools are identified and acquired.
* Conversion programs are developed.
* A conversion strategy is chosen.
* Data conversion staging environment is implemented.
* Target data model is stable.
* Funding for staff, consultants, and tools has been secured.]

[List all the major milestones pertaining to data conversion.]

The data conversion milestone table below provides a baseline version of the expected schedule of key data conversion milestones.

| **milestone** | **date** |
| --- | --- |
| <<Milestone>> |  |
| <<Milestone>> |  |

## conversion STANDARDS and CONVENTIONS

This section describes the data conversion design standards, coding standards, and naming conventions to be used by the data conversion team to promote consistency, readability, usability, quality, and productivity among all data conversion developers.

### Design Standards

This section describes the process flow that will be used in designing and building data conversion programs as well as what design artifacts will be included as part of design standards.

[Describe the process flow that will be used in designing and building data conversion programs. Also, describe what design artifacts will be developed (e.g., a combination of Visio diagrams, Excel spreadsheets, and data model diagrams will be used to document the conceptual data model, data flow, and process flow diagrams for converting a set of functionality.)]

### Coding Standards

[Define a set of programming conventions to be used by the data conversion development team in developing data conversion programs.]

The data conversion team will follow a set of programming standards and conventions to increase quality, reusability, readability, and consistency across all data conversion programs and artifacts. For each data conversion program, the following standards will be followed:

* Descriptive header – describes the functionality being developed including relevant references. This helps increases readability and enables troubleshooting. Each conversion program will have:
  + Program name
  + List of arguments
  + List of return values
  + Date created
  + Version number and date
  + Modification history and author name.
* Inline comments – contains descriptive notes relevant to the section being developed.
* Indentation - all data conversion programs will be indented to improve readability and consistency of format according to the established coding standards document.

### Naming Conventions

This section provides a set of naming conventions to be used for data files, databases, staging tables, conversion instances, and related conversion artifacts.

[Define a set of naming conventions to be used for data files, databases, staging tables, conversion instances, and related conversion artifacts.]

### CHANGE control

This section describes the process that will be used to control the acceptance or rejection of proposed changes to the scope of data conversion.

[Describe the process that will be used to control the acceptance or rejection of proposed changes to the scope in order to avoid unnecessary risks associated with scope creep.]

Any modification to or deviation from the agreed scope of the data conversion effort will be logged as a change request. It will then be assessed to determine the impact of implementing the change in terms of time, quality, and cost. Change requests will be approved by the <<designated person/board>> according to the delegation of authority from <<the project>>.

### version control

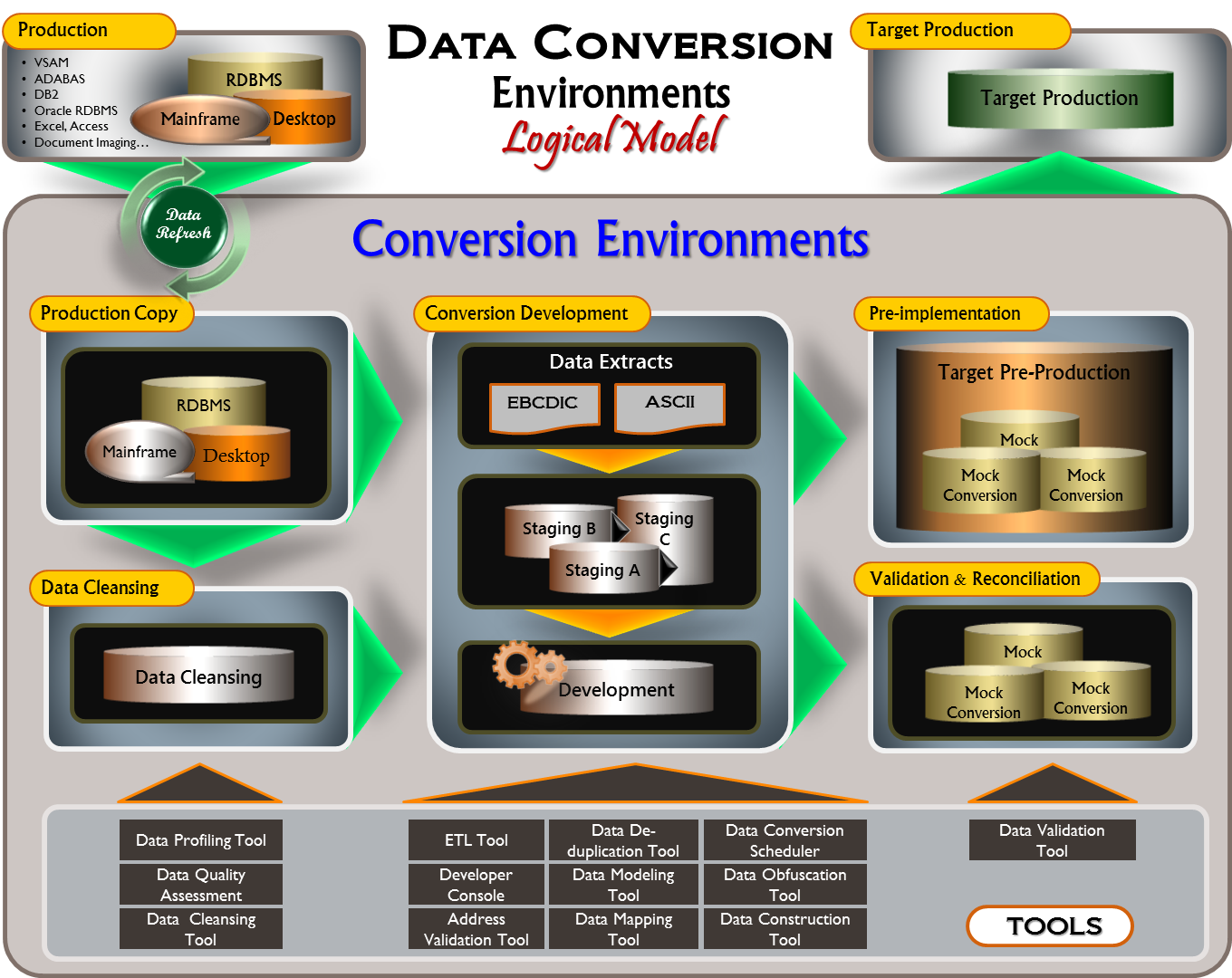
This section describes the process to be used for facilitating changes to conversion deliverables, conversion programming code, and other conversion related artifacts.

[Define the process to facilitate changes to conversion design and programs due to a bug fix, new business requirement, data cleansing process, or functional design, etc. Also, describe the process by which data conversion deliverables, design artifacts, and program source code will be version controlled and managed.]

## data conversion environments

This section describes the conceptual data conversion environments required to facilitate data conversion development, data conversion testing, data validation and reconciliation, and data cleansing.

Figure 6-5: data conversion environments - logical model



[Data conversion is rarely a direct transfer of data from source to target. Therefore it is necessary to have all the source data staged in an interim area to allow additional processing to get the data ready before loading it to the target system. This interim data holding area (a.k.a., staging area) houses the data that was extracted from all the sources, possibly from different platforms (e.g., Oracle, DB2, MS Excel, MS Access, VSAM, ADABAS), and allows for additional processing to be performed in this area. In addition to a staging area, the project team should also consider other environments that are essential to supporting data conversion development, such as data conversion testing, data validation and reconciliation, and data cleansing.]

[Define the data conversion environments required to facilitate data conversion development, data conversion testing, data validation and reconciliation, and data cleansing, if applicable. The project team should begin by:

* Reviewing the mission each data team is specifically tasked to accomplish and what their needs are for the environment.
* Creating a logical layout of the data conversion environments similar to the one shown in Figure 6‑5. Describing the purpose of each environment including the inflow and outflow of data between these environments.
* Specifying the software requirements that are necessary to implement and support the work within these environments.]

This section describes the logical view of the data conversion environments as shown in <<Figure ##>>. It provides a high-level layout of what environments are needed and for what purpose, the inflow and outflow of data between these environments, and the tools required to support data conversion activities.

The data conversion environments are composed of the following individual environments:

* Production-Copy
* Data Cleansing
* Conversion Development
* Validation and Reconciliation
* Pre-Implementation

The following software tools will be used to facilitate data conversion:

* ETL tool – <<Software tool>>
* Data modeling tool – <<Software tool>>
* Address validation tool – <<Software tool>>
* Mainframe data extraction tool – <<Software tool>>
* Data profiling tool – <<Software tool>>
* Other <<Software tool>>

# Conversion project team

This section discusses staff planning, structure of the data conversion team, and roles and responsibilities of each team member.

### STAFFING PLAN

This section provides the data conversion staffing plan. The staffing plan will be regularly updated as changes occur. This staffing plan may be impacted by any change to the project schedule, resource availability, data conversion activities scheduled/prioritized in various areas, and data cleansing schedule. Therefore, the staffing plan will need to be reviewed periodically, and the number of staff required will need to be re-assessed.

[Provide a detailed data conversion staffing plan describing how the project will effectively manage staff resources required at each stage of the conversion effort to support the data conversion strategies outlined in this plan. The following tables may be used for this purpose.]

| DATA CONVERSION STAFFING PLAN | | | | | |
| --- | --- | --- | --- | --- | --- |
| Role | responsibility | skills required | # OF STAFF REQUIRED | sTART DATE | END DATE |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

| DATA CONVERSION RESOURCE LOADING CHART | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ROLE | # OF STAFF REQUIRED | # of Full-time equivalent (FTE) STAFF in PY | | | | | | |
| july | aug | sept | oct | nov | dec | total |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| Total FTE |  |  |  |  |  |  |  |  |

### team organizational chart

This section discusses the structure of the data conversion team.

[A link to the example of a data conversion team organizational chart is provided in Appendix C.]

### ROLES and RESPONSIBILITIES

Data conversion requires collaboration among all the teams involved, particularly the functional teams, technical data conversion team, the data infrastructure team, and the data management support teams. This section defines the team organizational structure and the key roles and responsibilities that are necessary for the data conversion team to effectively achieve the goals and objectives stated in this plan.

[Define all the key data conversion roles and responsibilities for this data conversion effort. It is recommended that roles and responsibilities be specific and clearly defined, and established as early as possible in order to help guide the data conversion team throughout the project. The following table may be used for this purpose. A link to the example of data conversion team roles and responsibilities is provided in Appendix C.]

| DATA CONVErSION ROLES and RESPONSIBILITies | |
| --- | --- |
| ROLE | responsibility |
| <<Role>> | * <<Description>> |
| <<Role>> | * <<Description>> |
| <<Role>> | * <<Description>> |
| <<Role>> | * <<Description>> |

# data quality strategy

This section describes the strategy that will be used to test all custom-developed data conversion programs to ensure that the results produced by the conversion programs meet the established conversion acceptance criteria.

[In order for data conversion to be accepted, converted data needs to be validated against all applicable business rules and transformation rules to confirm that data conversion meets the established acceptance criteria and that all legacy data needed by the target system are converted accurately compared to source and compatible with the target system.

Data quality strategies must be capable of confirming that data loss and accuracy is within the established acceptance criteria during each stage within the data conversion process.

There are different methods or options of testing converted data and content conversions. One of the methods is sampling, where some subset of random data or content is selected and inspected to ensure the conversion is completed as designed. While the sampling method works, depending upon the established data conversion acceptance criteria, this method may not be suitable to provide relevant and complete statistical information to determine whether data conversion satisfies the established acceptance criteria. The followings are some suggestions to help prepare for data conversion test:

* Confirm scope of data source systems. Verification should include data to be included as well as excluded.
* Define the source to target high-level mappings for each category of data or content and verify that the desired type has been defined in the target system.
* Check target system data model stability and verify data requirements such as field names, field types, mandatory fields, valid value lists and other field-level validation checks.
* Using the source-to-target mappings, verify the source data against the data requirements of the target system. For example, if the target system has a mandatory field, verify that the corresponding source data field is not null, or if the target data field has a list of valid values, verify to ensure that the corresponding source data field also contains these valid values.
* Review and formally confirm the data conversion specifications, which should include the following elements:
  + Source system definitions
  + Number of records in the source system and growth rate
  + Data cleansing requirements
  + Performance requirements
  + Testing requirements
  + Source-to-target data mapping documents
  + Conversion design requirements
  + Referential integrity constraints of the target database
  + Business/validation rules
* Two aspects of data conversion testing:
  + Functional testing - verifies that converted data support the functionality of the target application system.
  + Non-functional testing - confirms that all data in scope was successfully migrated to the target system in terms of accuracy and completeness per the established data conversion acceptance criteria.]

[Define the strategy to be used to test all custom-developed data conversion programs to ensure that the results produced by the conversion programs meet the established acceptance criteria. Particularly, describe the approaches to conversion functional testing and conversion non-functional testing with respect to each of the following tests (if applicable):

* **Conversion Unit Testing** - initial unit testing is conducted by the developer to verify that the conversion program performed according to specifications and that the appropriate tables and fields were populated.
* **Data Usability Testing** - primarily focuses on verifying that converted data is functionally compatible with the target application system.
* **Mock Conversions** - controlled “dress rehearsal” of all the conversion execution activities required to transfer legacy data to the target system.
* **Data Validation and Reconciliation (DVR)** – leverages relevant information from the conversion design specifications to validate the entire converted data volumes to collect and provide statistical information that is necessary to determine whether data conversion satisfies the established acceptance criteria. Since this has significant downstream impact to the organization and the data owners in terms of data quality, integrity, completeness, and usability, it is highly recommended that the department own this important responsibility to ensure that integrity and completeness of the lifeblood of the organization (data) are still preserved through the conversion process**.]**

## Conversion unit testing

The main purpose of conversion unit testing is to verify whether conversion programs conform to detailed design specifications and also to ensure that the converted data produced by these conversion programs are accurate compared to legacy data. This section describes the process that will be used by conversion developers to conduct unit testing.

[Describe the process that will be used by conversion developers to conduct unit testing. Also what reports will be produced as evidence of due diligence that unit testing is properly executed.]

## Data Usability Testing

The primary focus of data usability testing is to determine how compatible the converted data is with the target application system. In other words, data usability testing verifies to ensure that there is successful interaction between the target application and the newly converted data. Data usability testing is accomplished through the following mechanisms:

* System test and User Acceptance Test (UAT): testing mock converted data through the application user interface using existing system test and UAT test cases with a selective set of converted data. The purpose is to discover discrepancies between the application and the converted data.
* Independent data validation: testing mock converted data through a series of batch processes and functional queries developed based on data validation rules and business rules. This test helps identify inconsistencies and incompatibilities between the converted data and the target application.

[Describe the approach to be used to verify that converted data is functional and compatible with the target application system.]

## Mock Conversions

A mock conversion is a controlled “dress rehearsal” that includes all steps that will occur during the actual live conversion to migrate data from the legacy systems to the target system. Each mock conversion simulates the real conversion cutover process with actual data volumes. The purpose of mock conversions is to identify and resolve any conversion program issues and configuration problems ahead of time. Also, it provides opportunities for independent data validation of the actual data volumes, assessment of data conversion readiness, and ensures that the entire data conversion process can be finished within the timeframe allocated for data conversion cutover.

[Since the aim of mock conversions is to identify and resolve any conversion program issues and configuration problems as early as possible, each mock conversion should have a well-defined set of specific strategic targets of what it intends to achieve. The following is a suggested format and outlines the important elements that should be included in the planning of a mock conversion:]

**Title:** Mock Conversion <<##>>

**Purpose:** <<what does this mock conversion intend to achieve? >>

**Duration:** <<planned start and end dates>>

**Key Participants**:

<<Consultant key staff or teams>>

<<State key staff or teams>>

**Other Participants**: <<staff or teams that need to be informed>>

<<Consultant key staff or teams>>

<<State key staff or teams>>

**Effort:** <<number of hours estimated>

**Basis for measuring progress**: <<what will be used to measure progress? >>

**Dependencies:** <<list all the items that this mock conversion depends upon>>

**Exit Criteria**: <<list the criteria or requirements that must be met in order for this mock conversion to be considered as complete. >>

[Furthermore, the following are some best practices to consider:

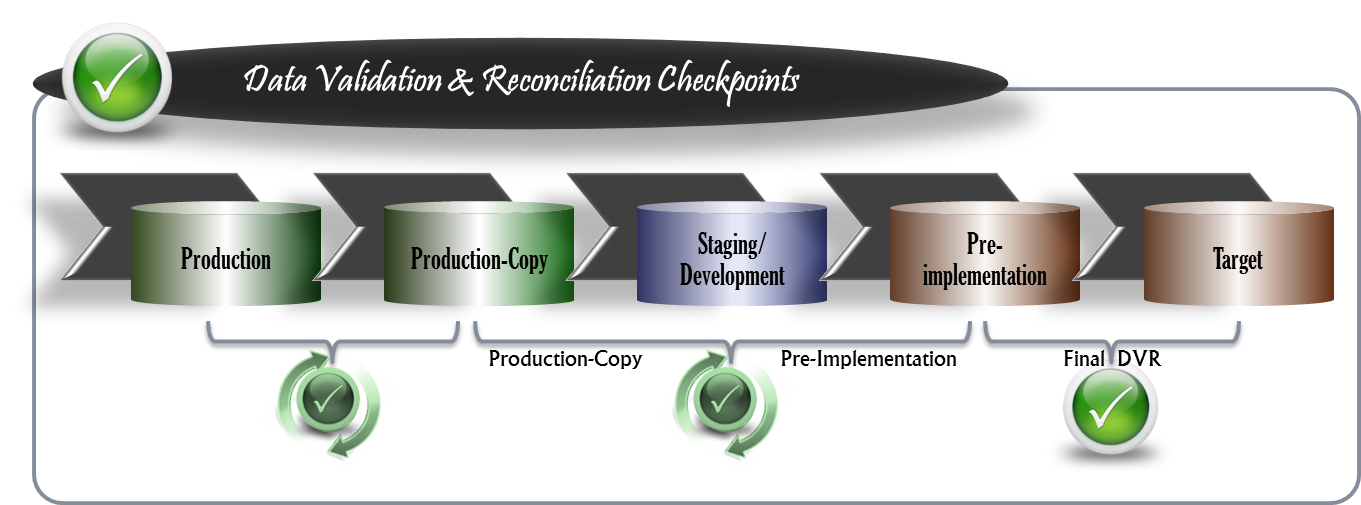
* A data conversion run book should evolve out of the mock conversions to detail each step of the conversion process, when each conversion step occurs, dependencies, who is responsible, and so on.
* Each mock conversion should be started with new data extracts, staging (if applicable), transformation, and then load processes.
* All error resolution and data validation processes should be conducted as part of the overall mock conversion process to identify and resolve errors, determine new error resolution, verify data validation requirements, and continue to refine the process.
* Mock conversions should be conducted in an environment that closely resembles that of the target environment. Configuration and customizations in this environment should be frozen including the database instance where the mock conversion takes place. If changes are made to the configuration as a result of necessary adjustments made to the mock conversion process, they should be documented.]

[Describe the approach that will be used to identify and resolve any conversion program issues and configuration problems ahead of time. Also, describe what approach is to be used to assess data conversion readiness, and to ensure that the entire data conversion process can be finished within the timeframe allocated for data conversion cutover.]

## data validation and reconciliation (DVR)

The purpose of data validation and reconciliation is to ensure that all required legacy data has been accurately converted to the target system. It is a comprehensive process that rigorously tests the data as it is migrated through different stages to determine whether the converted data is a true representation of the data that exists in the legacy data sources. Figure 8‑1 illustrates the different checkpoints that are essential to data validation and reconciliation.

Figure ‑: data validation & reconciliation CHECKPOINTS



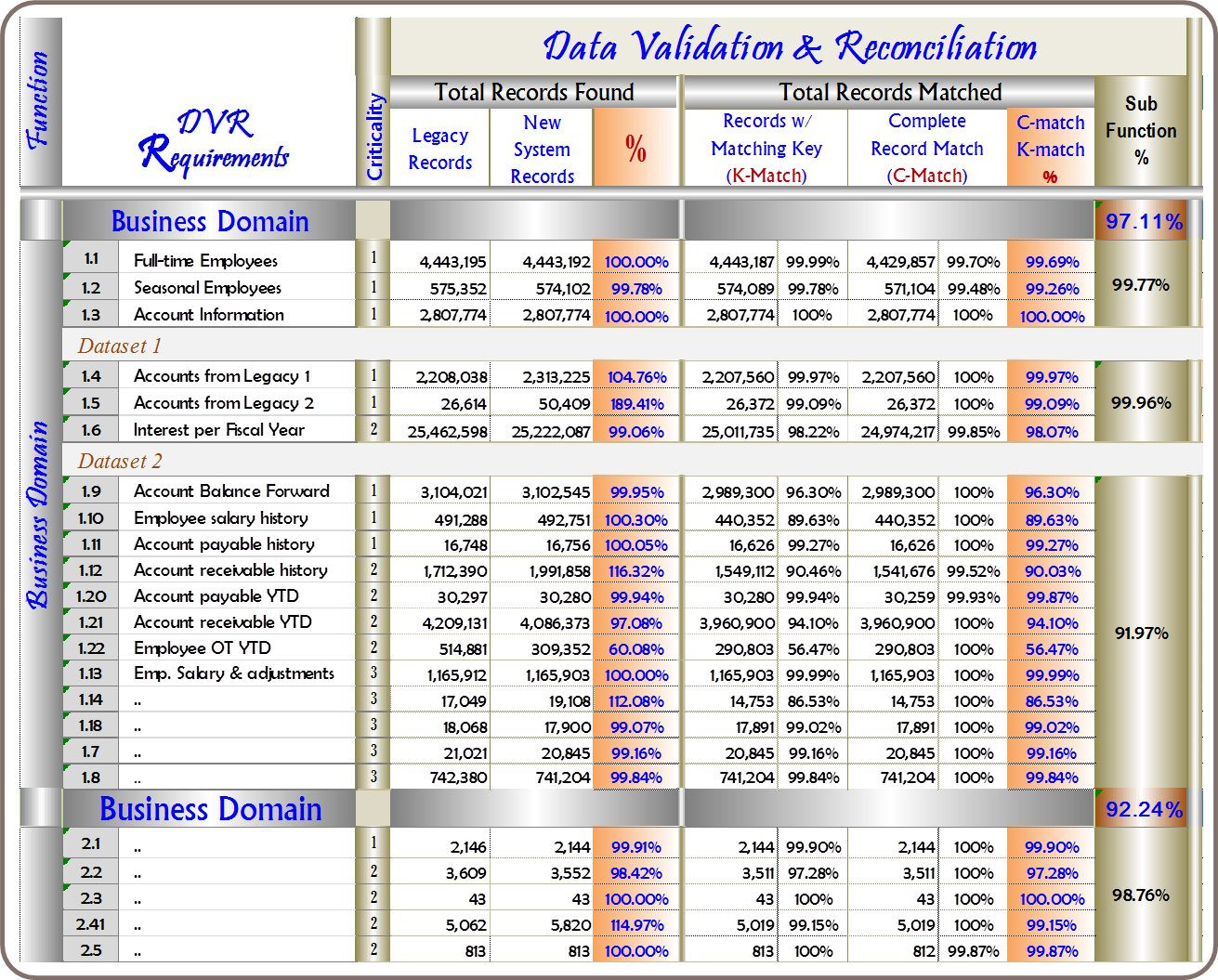
The data validation and reconciliation process verifies the validity of the converted data and the conversion metrics. These metrics are the counts of records converted and the summations of critical numeric values such as annual deposit totals, quarterly account balances, etc.

Furthermore, the data validation and reconciliation process ensures that all records to be converted are accounted for and that the critical numeric amounts are either the same in the target system as they are in the legacy systems or any variance is a result of adjustment and transformation rules approved by the business via documented decisions. This includes accounting for records that are not converted, either intentionally or unintentionally, and the reasons for the failures. Figure 8‑2 is an example of a data validation and reconciliation report.

[The objectives of data validation and reconciliation should be to:

* Ensure that critical conversion errors are discovered and corrected early.
* Ensure that the data needed to support the target application functionality is converted and validated for correctness prior to system cutover and during the application user acceptance testing cycle.
* Ensure that all pertinent records and other information is correctly transformed and converted including all of the relationships among the various data elements.
* Ensure that all pertinent records and other information is completely transformed and converted and provides project leadership with the confidence that no information is lost in the process.
* Provide input to the formal and final data certification of the converted data which is to confirm whether data conversion meets the established acceptance criteria.

Figure ‑: dATA VALIDATION & RECONCILIATION REPORT



Since data validation and reconciliation has significant downstream impact to the organization and the organization data owners in terms of data quality, integrity, completeness, and usability, it is highly recommended that the department own this important responsibility to ensure that integrity and completeness of the lifeblood of the organization (data) are l preserved through the conversion process. Furthermore, data validation and reconciliation should be an independent effort and all of its validation and reconciliation processes should be 100% automated (if at all possible) in order to support multiple iterations and ultimately to complete validation and reconciliation of the entire converted data volumes within the timeframe allotted for both data conversion and validation at cutover.

Since accurately converted data is essential to successful system implementation, DVR should be performed thoroughly and iteratively. Ideally, data validation and reconciliation efforts should be initiated early in the process to derive and document validation and reconciliation requirements via data conversion scope, business requirements and expectations for data, acceptance criteria, business domains, business process priority, conversion design decisions, business rules, and transformation rules. It is recommended that a separate data validation and reconciliation plan be developed to ensure that all essential aspects of DVR are covered. Like any software development project, the data validation and reconciliation effort should be carefully planned and followed. Furthermore, data validation and reconciliation requirements should be the driver of all validation and reconciliation activities.]

[Describe the approach that will be used to ensure that all data to be migrated are accurately converted compared to source and compatible with the target application system. The following suggested questions can help guide you in creating a DVR plan for your project:

* What is your overall strategy concerning ongoing data validation and reconciliation during the development and testing phases of data conversion?
* What is the strategy you will use to ensure critical conversion errors are discovered and corrected early?
* What is your approach to data anomaly resolution?
* What is your approach to meeting expectations, requirements and/or commitments of the business, project leadership, data cleansing, system user acceptance test, and data conversion teams such as schedules, data validation and reconciliation performance, reporting, etc.?
* What is your strategy for validating and reconciling full mock conversions? How much time are you allowed to complete the data validation and reconciliation of a full mock conversion?
* What information (i.e., volume metrics) are you required to capture? What resources (i.e., DVR tools, staffing level, expertise, environments, system access, etc.) do you need in order to meet these expectations or requirements?]

## DATA error resolution PROCESS

This section describes the process that will be used to identify, escalate, and resolve data errors during the data conversion process.

[Each conversion dataset or data domain will have its own set of unique data errors that will require resolution during conversion testing, mock conversions, and even during final go-live conversions. Therefore, it is recommended that a data error resolution process be developed for each business function in order to effectively address its own unique data errors. The following are the two common types of errors specific to conversion data loads.

* **Critical** data errors are those that prevent a record from being loaded into the target data storage and/or cause data integrity errors. These types of data errors should be identified and addressed as soon as possible. If possible, these types of data errors should be corrected in the legacy system prior to subsequent extracts and loads. Critical data errors will more than likely prevent continuing with other conversion loads that are dependent on the failed records and must be resolved quickly or the records have to be skipped or removed from subsequent conversions until fixed.
* **Non-Critical** data errors are those that have invalid values or missing configuration data which will not prevent a record from being loaded. These types of errors should be identified and reported for resolution.

[Describe the process that will be used to identify, escalate, and resolve data errors during the data conversion process and the roles and responsibilities required to facilitate the process including but not limited to, the following items:

* Define process to report any new data defect/issue.
* Define mechanism to track issues and resolution.
* Define process to identify data defects and exceptions during the ETL process.
* Define process to isolate defective records.
* Define process to escalate defective records for resolution either in the legacy source or in a separate defect resolution environment.
* Define process to identify functionality impacted by data defects.
* Define process to incorporate corrected records into the ETL process.**]**

# conversion implementation

This section discusses the implementation approach, conversion cutover process, implementation planning and considerations, and data certification process that will be used to facilitate conversion implementation.

## Approach to implementation

This section discusses the conversion implementation approach that will be used for this data conversion effort. It also discusses the pros and cons of this approach, the risks associated with the selected implementation approach, as well as the strategies required to mitigate those risks.

[Essentially there are two approaches for transferring the source data, which has been transformed, cleansed, tested, and validated, to the production environment of the target system. Depending on business needs, one of the two approaches may be a better fit for the project than the other. However, each approach does come with pros and cons. For this reason, special consideration should be given to each approach before the final decision is made. Please note, parallel option requires both the legacy systems and the newly implemented system to run concurrently for the duration of the agreed time in order to allow business the opportunity to test drive and validate the new system before signing off on its acceptability.

1. **The “big bang” cutover** (with or without parallel option)

This approach, in concept, requires all the source data to be extracted, transformed, and migrated all in one process during the time span of the cutover window.   
Pros:

* + - No two systems running simultaneously
    - No synchronization between systems to deal with
    - With parallel option
      * Allows business time to fully validate and sign-off the new system

Cons:

* + - Risks associated with having a limited time-frame
    - Rollback strategies may be challenging
    - Business downtime
    - With parallel option
      * Risks and costs associated with keeping data current in both systems
      * Dual-keying for system users
      * Requires more resources

1. **The Incremental cutover** (with or without parallel option)

This approach reduces the risk of a big bang approach by extracting, transforming, and migrating discrete parts of the business and associated data. This approach, in concept, sounds great but it can be very challenging to execute effectively due to the complexity of the underlying data structure, application architecture, data, and the interwoven business processes.

Pros:

* + - Fewer risks associated with having a limited time-frame
    - Rollback strategies may be less challenging
    - Less business downtime
    - With parallel option
      * Allows business time to fully validate and sign-off on the new system

Cons:

* + - Extremely difficult to manage all the risks associated with intricate relationships between business processes and the underlying data associated with those business processes
    - With parallel option
      * Risks and costs associated with keeping data current in both systems
      * Dual-keying for system users
      * Requires more resources

Business needs must be the primary driver in determining the best-fit data conversion implementation strategy for the project. Different business needs require different implementation approaches and it pays to fully understand each approach as well as its associated pros and cons so the right decisions can be made and work can be planned at the outset of the project.]

***[***Describe the data conversion implementation approach that will be used for this data conversion project and discuss the pros, cons, and risks associated with the selected implementation approach from the business perspective***. Consider the following items when developing the detailed conversion cutover plan*** for your project. Depending on the selected implementation approach, some items may be less applicable than others.

* Timing and span of the conversion cutover window.
* The need for a plan for freezing the legacy physical data structures.
* Data retention requirements for legacy data file extracts and staging databases.
* Steps for migrating data to the target environment at conversion cutover.**]**

## conversion cutover process

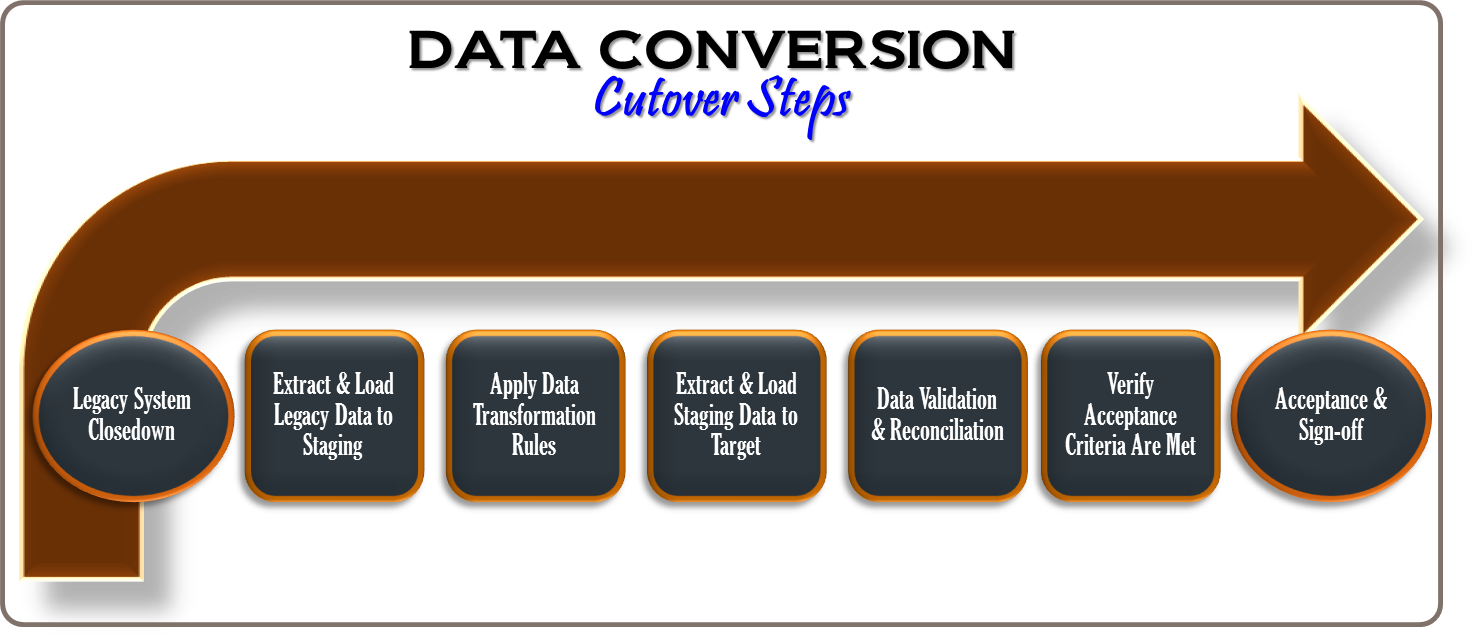
This section discusses the approach to extracting, transforming, and loading legacy data to the target production environment. Also, it discusses the data retention requirements for certain data conversion artifacts concerning legacy system data and the conversion cutover window.

Figure ‑: data conversion cutover steps

The following describes the activities shown in Figure 9‑1 which must be completed before the legacy data can be loaded to the target system:

1. Legacy System Closedown

**Freeze all input** - At this point in time, no data input should be allowed into any of the legacy systems, from which source data will be extracted, either by direct entry or through the various interfaces.

**Complete all data processing** - At this stage, verify that all background data processing jobs such end-of-day financial reconciliation or posting processes are complete.

**Freeze the data** - Once all background data processing jobs are complete, legacy data will be “frozen” and no data changes will be allowed until data conversion process is complete and the target system is operational.

**Start data conversion process** - During this stage, data conversion process will follow the conversion run book developed during the mock conversions.

1. Extract and Load Legacy Data to Staging
2. Apply Data Transformation Rules
3. Extract and Load Staging Data to Target
4. Execute Data Validation and Reconciliation Process
5. Verify Acceptance Criteria Are Met
6. Obtain Business Acceptance and Sign-off

## implementation planning & considerations

This section discusses the implementation planning and considerations that will be used specifically for <<*static data, archive data, document images, and dynamic data>>.*

[The main consideration in a data conversion schedule is the time allocated for the cutover window on the date of the conversion. The cutover window is the time allotment in which legacy production data is extracted, converted, and migrated to the target production environment. Typically, the project team will need to consult with the business to carefully determine the timing and span of the conversion cutover window such that downtime and disruptions to business services are minimized. Therefore, this should be defined up front and targeted throughout the data conversion process to ensure that it is achievable. In order to ensure that the entire data conversion process (extract, transform, and load) can be completed within the timeframe allotted for data conversion cutover, it is important for the project team to actively look for ways to reduce the data amount that must be migrated at the time of conversion cutover and also to minimize the time it takes to complete the entire data conversion process. Some areas to consider:

* **Mock conversions** – one or more may be needed depending on the criticality of the project. A mock conversion is a controlled “dress rehearsal” of all the execution activities required to migrate data from the source system to the target system. Each mock conversion simulates the real conversion cutover process with actual data volumes.  
    
  The purpose of mock conversions is to identify and resolve any conversion program issues and configuration problems ahead of time. Also, it provides opportunities for independent data validation of the actual data volumes, assessment of data conversion readiness, and ensures that the entire data conversion process can be finished within the timeframe allocated for data conversion cutover.
* **Conversion programs optimization** – This consists of keeping track of all the details associated with each mock conversion, actively fine tuning and optimizing data conversion programs, and making sure that a data conversion run book is built, kept current, and the order of execution for each conversion program is continuously monitored and optimized.
* **Data categorization and prioritization** – special consideration should be given to the following data categories with respect to conversion cutover:
* **Static data**- data that will remain unaltered such as prior fiscal year data.
* **Archive data** – data that is no longer actively used which is stored on a separate data storage device for long-term retention.
* **Document images** – paper documents that were scanned and converted to digital images.
* Dynamic data – data that is actively being used, updated, or newly generated.
* Open data transaction – a business transaction that has not completed its process cycle (e.g., a workflow item or a service ticket that remains open with additional activities required prior to being closed).
* **Closed data transaction** **–** a business transaction that has completed its business cycle and is subsequently used for information purposes only, for example a service ticket with all related activities completed and a ticket status of “closed.”
* Datasets that require manual conversion.
* Datasets that can be extracted, transformed, and loaded to the target system ahead of time before the final cutover.
* Datasets that can be migrated incrementally.
* Datasets that can only be migrated at the time of cutover.]

[Describe the implementation planning and considerations that will be used for conversion cutover with respect to archive data, dynamic data, etc.]

## Data certification

This section describes the approach that will be used for determining whether data conversion meets the established acceptance criteria.

[Data certification is the process by which the state department verifies that the converted data is of sufficient quality to enter data conversion cutover and enable initial system implementation. In other words, data certification is the data conversion readiness verification process.

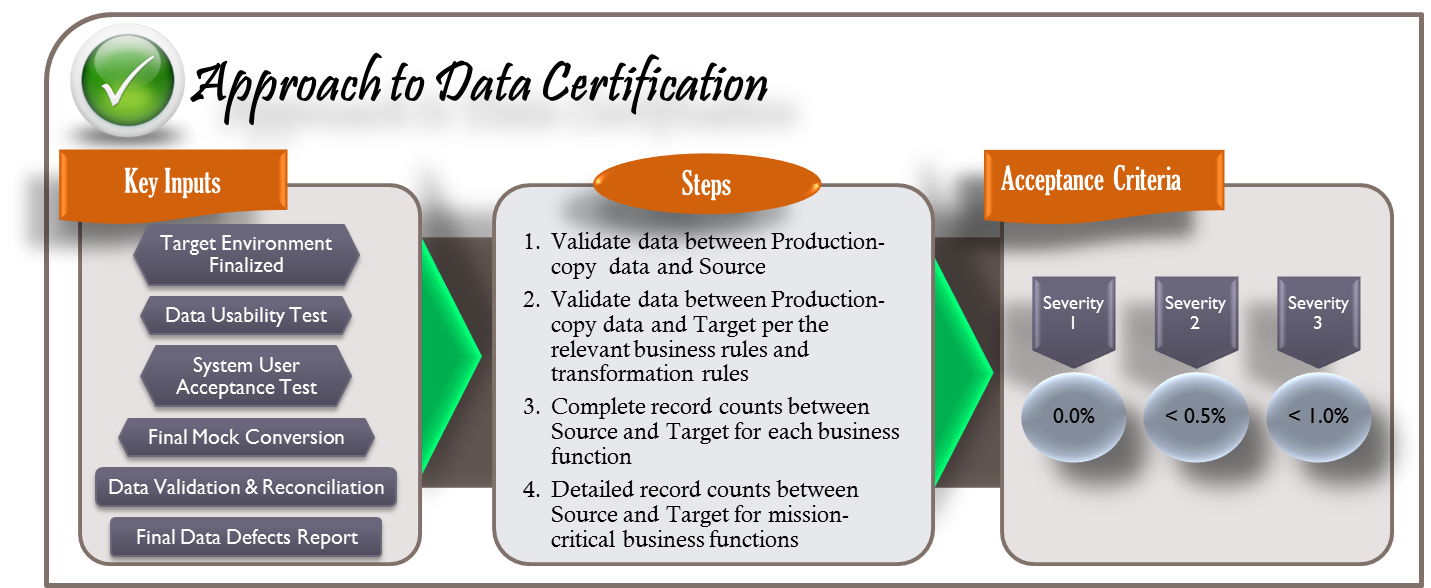
The data stored within the current legacy systems is the lifeblood of the state department business and therefore plays a significant role in meeting business functionality in the new system. Without accurately converted data, the new solution, regardless of its new look and feel and technology uplift, will be of limited use to the department business and potential public relations incidents may result.

Figure ‑: certification of data conversion acceptance criteria

Figure 9‑2 illustrates a conceptual approach to data certification. In that, it lists the key inputs or potential artifacts and information that are essential to determining whether data conversion meets the established acceptance criteria. It also outlines the key steps that will be used during the data certification process. Furthermore, Figure 9‑3 provides an example of a data validation and reconciliation report with relevant detailed information to support the data certification process.

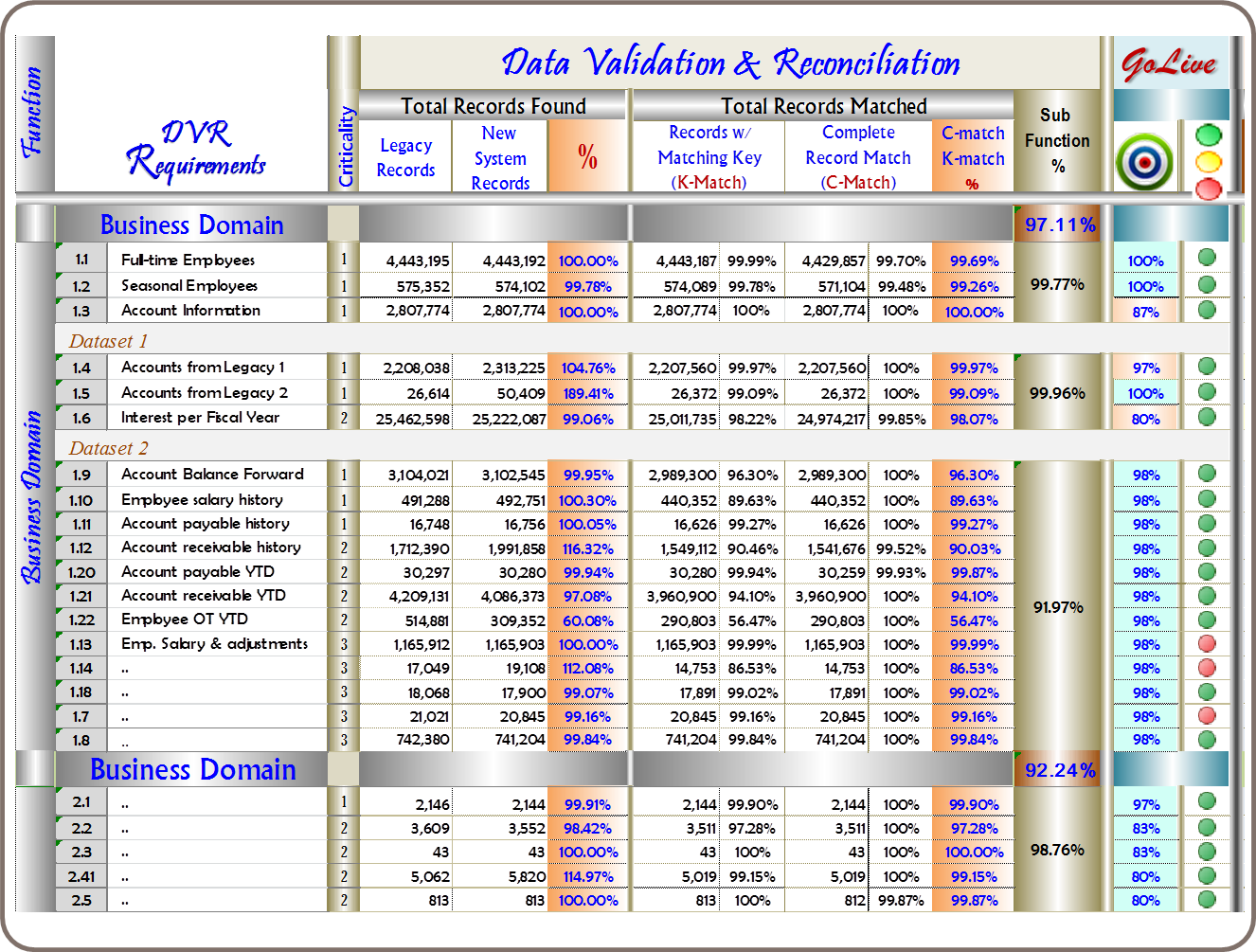


Figure ‑3: DVR as key input to data certification

[Describe the approach to be used for data certification. Specifically describe what key inputs or information will be needed in certifying the converted data, what steps will be carried out in the data certification process, and finally what acceptance criteria the converted data will be certified against.]

# post-conversion support

This section describes the approach to be used for identifying and addressing data issues discovered in the target system after conversion implementation.

[Inevitably, after the data conversion cutover, data issues will be uncovered and maybe greater than originally anticipated. Furthermore, there is a steep learning curve and tremendous pressure to overcome in order for anyone to become effective in addressing data issues especially during the post-conversion period, as it requires knowledge and understanding of the source data and data structure, the target data and data structure, and all the transformation rules that were applied to the source data during conversion. Therefore, the project/department ought to plan to retain the data conversion teams, who have been intimately involved with data conversion, to continue providing production support during post-conversion period. The retention period will be determined by many factors, some of which may be:

* Number of data errors, types of data errors, severity level, and priority
* Growth rate of data errors vs. number of data errors resolved by the team within a given period
* Knowledge transfer and training]

## identify post-Conversion data Issues

This section describes the approach to be used for identifying and classifying data issues.

[Describe the approach to be used for identifying and classifying poste-conversion data issues. The following are some common types of data conversion related issues:

* Data issues not cleansed in the source data at the time of cutover.
* Data conversion bugs not fixed at the time of cutover.
* New data issues discovered in the target system after cutover.]

## Resolve Post-Conversion data Issues

This section describes the approach to be used for prioritizing and resolving data issues.

[Describe the approach to be used for prioritizing and resolving data issues, particularly the types of data issues as outlined below:

* Data issues not cleansed in the source data at the time of cutover:  
  *These are known data issues in the source data that were not resolved at cutover.*
* Data issues caused by conversion bugs not fixed at the time of cutover:  
  These are data issues resulting from bugs in the data conversion programming code discovered during various stages of testing and validation.
* New data issues discovered in the target system after cutover:  
  These are data issues which are unknown at the time of cutover, and are discovered after the new system goes live.]

## decommissioning and monitoring

This section discusses the strategy to be used for decommissioning staging data after conversion implementation is complete.

[If staging area and pre-production data storage will be created as part of the data conversion process, the project team needs to work with the key data stakeholders to formulate strategies to be used for managing data in the staging area once data conversion cutover is complete.]

[Describe the strategy to be used for managing the data in the staging area (if used) once data conversion cutover is complete. The strategy should address the following:

* What data to retain.
* Where and how to retain the data.
* What pre-conditions are required to be met before the system and its data can be decommissioned? Be sure that these conditions are fully documented and agreed upon early on so the project can begin confirming that the data conversion has met these conditions.
* How the ownership of the data environment and the monitoring of the data environment will be handed over.]

1. CONSIDERATIONS FOR COTS, MOTS, and CUSTOM

| **CONSIDERATIONS FOR COTS, MOTS, and CUSTOM IMPLEMENTATION** | |
| --- | --- |
| **COTS** | * Data conversion/conversion may result in the loss of information if the Commercial Off The Shelf (COTS) target data structure has or requires fewer features or data elements than the source format. * Sufficiency of skilled staff in both business and IT functions affected by the COTS Product implementation. *(general)* * “As-is” requirements must be derived from the current business processes |
| **MOTS** | * “As-is” requirements must be derived from the department’s current business processes * “To-be” requirements are defined by a set of business processes on which the COTS production was designed PLUS the gap identified between the business processes supported by existing systems and those supported by the COTS product. This is the basis for the Modified Off The Shelf (MOTS) selection. * Most of the COTS products require extensive customization in order to meet the needs and requirements of current business processes. Since no COTS product was specifically designed to meet the department's unique requirements, there will be a gap between the business processes supported by the department’s existing systems and those supported by the COTS product. Thus, the department should understand and clearly define this gap well before the MOTS implementation commences. In addition, the department should clearly define the level of customization required for the COTS product baseline to adequately address current needs and requirements. |
| **CUSTOM** | * Comprehensive analysis of the current environment should be performed prior to finalizing the project planning and scope. * Data conversion should be done as a separate and independent project. * Data conversion acceptance criteria and expectations should be clearly articulated and documented at the outset and managed throughout the data conversion process. * Validation of converted data should be independently performed by a separate team. * Individuals, including IT staff, with institutional knowledge and expertise of legacy data should be identified and actively involved throughout the project. * It is recommended that the department have a dedicated team to address legacy data issues that will not work properly in the target application as soon as possible. |

1. LIST OF ARTIFACTS
   1. Conversion Design Decision template
   2. Data Conversion Terms and Definitions
   3. Data Dictionary Template
   4. Data Mapping Template
   5. Data Quality Issue Log template
2. DATA CONVERSION TEAM
   1. Conversion Roles and Responsibilities - Definition
   2. Conversion Roles and Responsibilities – RACI Chart
3. TABLE OF FIGURES

[FIGURE 1‑1: DATA CONVERSION PROCESS OVERVIEW 2](#_Toc409767185)

[FIGURE 1‑2: DATA CONVERSION METHODOLOGY FRAMEWORK 3](#_Toc409767186)

[FIGURE 2‑1: DATA CONVERSION ACCEPTANCE CRITERIA 11](file:///N:\Consulting\Methodology%20Ready%20for%20Review\Data%20Management\Data%20Conversion%20Plan%20Template%20dc%202015%200122%20SN.docx#_Toc409767187)

[FIGURE 2‑2: DEFINITION OF DATA ERROR SEVERITY LEVELS 12](#_Toc409767188)

[FIGURE 2‑3: DATA CONVERSION ACCEPTANCE CRITERIA 13](#_Toc409767189)

[FIGURE 3‑1: TECHNOLOGY & INFRASTRUCTURE CONSIDERATIONS 23](#_Toc409767190)

[FIGURE 4‑1: DATA PROFILING REPORT 27](file:///N:\Consulting\Methodology%20Ready%20for%20Review\Data%20Management\Data%20Conversion%20Plan%20Template%20dc%202015%200122%20SN.docx#_Toc409767191)

[FIGURE 5‑1: DATA CLEANSING FRAMEWORK 29](file:///N:\Consulting\Methodology%20Ready%20for%20Review\Data%20Management\Data%20Conversion%20Plan%20Template%20dc%202015%200122%20SN.docx#_Toc409767192)

[FIGURE 5‑2: DETAILED DATA CLEANSING PROCESS 30](#_Toc409767193)

[FIGURE 6‑1: DATA CONVERSION PROCESS OVERVIEW 32](file:///N:\Consulting\Methodology%20Ready%20for%20Review\Data%20Management\Data%20Conversion%20Plan%20Template%20dc%202015%200122%20SN.docx#_Toc409767194)

[FIGURE 6‑2: DIFFERENT LEVELS OF THE MAPPING PROCESS 34](#_Toc409767195)

[FIGURE 6‑3: EXTRACTION & STAGING CONCEPTUAL PROCESS 36](file:///N:\Consulting\Methodology%20Ready%20for%20Review\Data%20Management\Data%20Conversion%20Plan%20Template%20dc%202015%200122%20SN.docx#_Toc409767196)

[FIGURE 6‑4: TRANSFORMATION & LOADING PROCESS 38](file:///N:\Consulting\Methodology%20Ready%20for%20Review\Data%20Management\Data%20Conversion%20Plan%20Template%20dc%202015%200122%20SN.docx#_Toc409767197)

[FIGURE 8‑1: DATA VALIDATION & RECONCILIATION CHECKPOINTS 53](file:///N:\Consulting\Methodology%20Ready%20for%20Review\Data%20Management\Data%20Conversion%20Plan%20Template%20dc%202015%200122%20SN.docx#_Toc409767198)

[FIGURE 8‑2: DATA VALIDATION & RECONCILIATION REPORT 54](file:///N:\Consulting\Methodology%20Ready%20for%20Review\Data%20Management\Data%20Conversion%20Plan%20Template%20dc%202015%200122%20SN.docx#_Toc409767199)

[FIGURE 9‑1: DATA CONVERSION CUTOVER STEPS 59](#_Toc409767200)

[FIGURE 9‑2: CERTIFICATION OF DATA CONVERSION ACCEPTANCE CRITERIA 62](file:///N:\Consulting\Methodology%20Ready%20for%20Review\Data%20Management\Data%20Conversion%20Plan%20Template%20dc%202015%200122%20SN.docx#_Toc409767201)