

Eight key steps which help ensure a successful data migration project: A white paper for inspection management professionals

Data migration defined

Data migration is the selection, preparation, extraction, transformation and transfer of data which is of the correct quality.

The challenge

Inspection departments often hold large amounts of information, of variable quality and in various formats. When considering the procurement of a new inspection management solution, inspection professionals are usually keen to understand how a prospective vendor will manage the transfer of their data to the vendors system.

This white paper will outline an approach that Credosoft recommend to ensure a successful data migration project. *Please note that this is not an exhaustive or detailed list..*

1. Ensure the client understand his role in data migration

The client knows best!

Whilst technology will be utilised to manage the physical extraction, transformation and loading of your data, too often data migration is viewed as a purely technical exercise. The reality is that only the owners and consumers of the data understand the meaning and value of the data. This

value must be retained and enhanced during data migration. Emphasising technology over business knowledge should be avoided.

Data migration is a business issue, not a technical issue

The inspection team are the experts in what the department does. They have been running the legacy systems and the systems have been running the inspection activities. Finally the inspection team have the expertise to make judgements about the quality and usefulness of the data. That being said that the client will rightly expect guidance on how to make the data fit the new inspection system.

The data migration analyst cannot know more about the business rules than the inspection team

2. Data profiling

The inspection team know where all the data sources that run the department are located. Early identification of all possible data sources is vital to ensure accurate estimating of timescales.

Departments often have various spreadsheets or mini-databases which support their daily inspection activities. Some may be unacknowledged however could be critical to the departments processes; sometimes addressing gaps in in the 'official' system. Each will be in different formats and varying quality and there will often be difficulties in 'linking' them together.

Once all the legacy data stores have been identified, this is the point where you may agree the scope of work

At this stage we discover and catalog all the data stores and their relationship to one another. We look inside the data stores to understand how they work, what data they hold and what challenges they may hold.

Finally, we need to agree on what data you want to migrate. In an ideal world you would migrate 'everything' however the reality is that there will be data which may not add value or may be covered by a different source.

3. Gap Analysis and mapping

Gap Analysis

This is where we analyse and quantify data gaps. There may be issues in the legacy systems; some issues will be known to the end users (which in part may have caused a proliferation of in-house developed solutions). Some issues will only be discovered when the database is examined for migration. It is important that these issues are 'flushed out' as early as possible so that appropriate action can be taken.

Reality checks	Where domain experts check the data to ensure it matches business reality
Internal gaps	Where the data does not conform to internal rules; for example referential integrity does not exist
Migration model gaps	This occurs when there is differences in the data from different legacy data stores
Target model gaps	This occurs when there are differences in data and data structures between the source and target systems

Mapping

This is the linking of fields in the legacy data stores to fields in the target. Mapping requires close collaboration between the client and the vendor. The use of a spreadsheet is perfectly acceptable for this activity.

4. Define Data Quality Rules

Data quality rules are used to measure the quality of the data and to help fix or mitigate any data quality issues. Every data quality issue should have a data quality rule.

ACTION	DESCRIPTION
Ignore	These are often low level data issues that don't cause a problem. It's likely that the majority of issues will fall into this category. In some cases you may be content to carry the issue into the new system because you have bigger issues to resolve..
Fix in flight	In this case you will perform some kind of transformation on the data to correct it. For example, convert lookup codes to upper

	case. This activity is carried out as the data is loaded into the target
Fix in source	In this case you may fix the data in the source data system. This may be the simplest solution.
Fix in target	In this case you may fix the data after it has been loaded into the target system. However if the data issues breaks validation rules in the target, you may have to temporarily disable the validation rules to all the data to pass into the target. This is perhaps the riskiest option and in most cases would be unnecessary..

5. Create a system retirement plan (de-commissioning)

As switching off the legacy data stores is one of the goals of any data migration project, it is important to have these conversations early on. For example the end users should be made aware that on x date, the legacy system will not be available to them. At this point we need to ensure the user is comfortable with decommissioning. Starting this conversation also helps make real the eventuality of a data migration.

In addition there may be legal requirements governing the storage and retention of the data.

6. Create a migration plan

Although data migration is a key activity in the overall project plan, the main migration activities need to be mapped out and timescales agreed for each.

Agree the implementation approach

- **Big Bang**

All the data is moved in one go. This is the most common approach

- **Phased**

The data is moved in separate parts, perhaps by business or geographic location

- **Parallel**

The data is moved to the target, however the legacy data continues to be used and changes are synchronised to the target.

7. Refine the migration scripts

Refining data migration scripts is an iterative process which can continue until close to go live. Throughout this period the vendor should engage the client in reviews of the scripts and the output, making adjustments where necessary. Ideally the client should be able to view and test the data in the target system.

8. Load

A migration controller, often known as the extract, transform and load tool is used (ETL) to load the data into the target system. The vendor may use as in house or a third party tool. These tools often do more than just these three tasks and the following are key features you should expect:

- Read the data from the legacy data store
- Validate the data
- Reformat the data and merge from multiple sources
- Start and stop the process
- Write the data to the target database
- Manage data errors
- Provide reports on execution
- Provide an audit trail
- Synchronise - in the event that the source data must be used during the migration, changes in the source can be updated with the target