

Summary

Many Information Communication Technology (ICT) upgrades and/or replacements often involve some element of data migration between systems. Whilst successful data migration is critical to maintaining business as usual, the length of time to plan, prepare and migrate the data are often overlooked with significant business impacts. Of all the issues around data migration discussed in this document, one common theme exists: standardisation. Standardise the use of fields, standardise the entries and standardise the permitted values.

1 Identification

The first step is to identify what data needs to be migrated. A major change in business process may make previous data redundant so there is no need to migrate the data, only a copy of it many need to be kept. Pragmatic business decisions need to be made as to what data are to be migrated and what is not required. The decision is made on real business need compared to cost of migration and cost of loss of data.

2 Standardisation

If there is one theme that comes across in data migration it is standardisation. Standardise the permitted values. Standardise the use of fields and standardise the entries. Achieve this and the process will be a lot smoother.

3 Quality

Migrating an incumbent data set with all of the errors and issues into a new system will only propagate the problems from the incumbent system. A pragmatic business decision has to be made on the level of investment to be made in cleaning up the data.

Resolution of data quality is one of the most time consuming aspects to a project and a data migration exercise. The time to address this and the availability of subject matter experts to qualify the data are the issues.

3.1 Consistency

Even with the best intentions, the value of data deteriorates over time. People use different abbreviations, e.g. Str or St for street, and typing errors are made. All of these cause inconsistencies in the data and raises the issue of data consistency. Whilst

the street example may seem minor, this consistency issue may cause issues in a new system, e.g. searching.

Another element of consistency is the use of the same field for different things. To meet pragmatic business needs, a data field can often be used for different purposes over time, which can cause issues on data migration, e.g. a flag setting can be missed because the records are incorrect. The use of a field needs to be standardised and valid values defined. Where uncertainty exists, the values should be deleted and identified in a different field.

Consistency issues are particularly prominent in addresses and the standardised use of building number, building name, street name and street type. Resolution of this is required prior to attempting a data migration.

3.2 Completeness

Another element of data quality is the completeness. Whether it is simple things like ensuring an address is complete with all of the valid elements through to missing records in a given data set.

Another element to completeness is the addition of information required for the new environment. A common example is the ability of a new system to support a preferred method of communication identified by a flag that needs to be added to the data of the incumbent system.

An assessment of the criteria for completeness required for the different data types and the effort required to resolve the completeness needs to be made based on real business needs.

3.3 Accuracy

Data accuracy is very diverse and varies significantly with data type and needs, e.g. account balances or whether someone still lives at a given address. For the critical business records, an assessment of accuracy is undertaken to determine if the data should be migrated. Full discussion of this is beyond the scope of this document but care should be taken to address it properly.

4 Field Mapping

Matching the records between incumbent and new systems is critical to a smooth and effective data migration. Elements include:

Name matching – Avoid any confusion by establishing a direct mapping of records between the systems based on field name. This should occur at both the database level and at the screen level.

Type matching – This is used to match the data type fields between systems, i.e. characters allowed in one system may be fully numeric in another. Validate the fields and the data to ensure migration can occur.

Value matching – After fields have been matched, ensure the values in the fields can be matched across systems. Issues include use of abbreviations but also ranges of values that can be accepted. Adjustments to data may be required.

Relationship matching – One system may have a one to many relationship between records, whilst others only support a one to one relationship, i.e. records can get lost in migration. After identifying the records and relationship, a decision on what to keep may be required.

5 Conversion

After defining what data are to be migrated, to what quality and to where, comes the actual conversion process. There are various ways to do this and issues to consider:

Timeframe - What can be migrated in a timeframe to meet business continuity needs.

Phased approach – It is often better to stage a migration but completeness of migration for a given customer or groups of customers is required to ensure integrity.

System to System – A system to system transition may be quicker, have fewer steps and potentially reduce the error risk at each step; this approach may not work for complex migrations.

Step by Step – For large complex migrations with many checks and points of failure, a step by step approach is favoured. Extract from old system, re-format to an intermediary step, convert to final format and then up-load.

Error and Exception Handling – Any data migration must properly cope with errors and exceptions and allow for recovery at any stage in the process.

6 Testing

Testing comes in several aspects:

- Testing that the up-load has completed.
- Testing that the right data has been loaded and is in the right place.
- Testing the completeness and accuracy of the data.
- Business as usual testing.

Whilst the specifics and complexity of the testing may differ, the need for robust testing does not.

7 Refresh

An initial data migration may be implemented and used for testing purposes. Subsequently the data may need to be cleared out and a fresh data set loaded for commencement of business as usual. Care should be taken to ensure that the current data are migrated and a process to manage the update established.

8 Partial Migration

Expedient decision making can often result in only the necessary records being migrated with the intention of addressing other records subsequently and/or by keeping the incumbent system going but as a read only source. This approach has been proven to be a path to higher costs and lower levels of service because both systems are kept running when they should not be and it places a higher premium on skilling and maintenance of disparate systems.

Furthermore, this approach causes complications and inefficiency further down the track and can delay critical projects like desktop upgrades.

9 Audit & Compliance

A data migration can be impacted by the needs of audit and compliance. Documenting what was done to the data in migration and how the data were migrated is often required. It may also be necessary to store the data created at different steps of processing during the migration.

10 Archiving

The data of an incumbent system may need to be kept in its original form for archive purposes.

One approach is to keep a simple copy of the data as is, together with a disk image of any applications necessary to reconstruct.

Alternatively flat ASCII files of the data can be kept.

Web: www.imscorp.com.au

© Copyright IMS Corp.. All rights reserved.

All referenced trademarks are those of their respective owners.

11 Data Type Specific

11.1 Documents

The migration of documents and images across systems is perhaps the least complex of data migrations. Issues to consider include formats, naming conventions and managing the links within and between documents.

11.2 Spatial Data

Electronic spatial data poses its own issues in data migration because of the formatting issues. Discussion of this topic is beyond the scope of this document.

11.3 Transactions

Transaction data are mainly stored in databases and are subject to the usual issues as discussed previously. Complexities around transaction data include opening balances, closing balances, migrating the transaction history and the linking of accounts. Relating back to the general ledger is also an issue.

11.4 Database Data

The migration of routine database data has been covered in this discussion, though special care needs to be paid to links within and between databases and real time feeds.

11.5 Real Time Feeds

The migration of real time feeds into and out of a database poses their own issues of coordination. This is often best done by having a dedicated duplicate feed for the period of the transition.

This document is for information purposes only and any advice given is of a general nature only and may not be applicable to an individual or a specific situation. IMS Corp. accepts no responsibility for any consequential loss or damage arising from the use of this document.

IMS Corp. is a specialist provider of services in the diagnosis, prescription and remedy of business problems. With a focus on empowering the client, IMS Corp. has met the needs of exceeded the expectation of clients to return shareholder value across the enterprise; to government, corporations and medium sized business alike.

IMS Corp.

Suite 5 275 Maroubra Road Maroubra NSW 2035 Australia

Email: info@imscorp.com.au