



# IMPLEMENTING SAMPLE MANAGEMENT SOFTWARE YOUR 7-STEP GUIDE TO DATA MIGRATION





## Introduction

Your data is one of your most valuable assets. Poor data quality can lead to costly mistakes. It could result in damage to your reputation and have a negative impact on your business and revenue opportunities as well as have legal implications.

When considering the quality of your data the first question you need to ask yourself is – do I know **where all my data is actually held?** You may be surprised (or not) to learn that even corporations and large institutions have rogue spreadsheets and local databases saved on their users' machines. Some institutions have attempted to manage the various spreadsheets in a centralised Content Management system and imposed some access restrictions. However, even these have evolved and expanded over time to become unmanageable and unfathomable data lists.

## The challenges of spreadsheets

Spreadsheets are excellent tools that most people know how to use and understand. When used for the right purpose they are invaluable. However, when used as a data repository or substitute database, there are several restrictions and challenges. These include:

- 1. Difficulties enforcing standards such as formatting or choosing from drop down lists.**  
If your users do not like your spreadsheet they will just create and use their own.
- 2. Lack of version control or data auditing.**  
Unless you create another version of the spreadsheet each time it is updated, it is very difficult to see who changed which critical piece of data (or added it) when and why.
- 3. Prone to accidental data errors.**  
It is very easy to either incorrectly amend numerical data columns. For example, incrementing where you were supposed to copy, or copying when you were supposed to increment. This is often not spotted where the numbers are complex or very similar.
- 4. No automatic duplicate record checking.**  
It is very easy to have the same record within the same spreadsheet or across multiple spreadsheets. This is especially problematic where you have samples that may be transferred from one team to another and you are dependent on manual updates in the spreadsheets to indicate sample movement.
- 5. Challenging to consolidate information for analysis and reporting.**  
It is difficult, and often time-consuming to obtain a consolidated view of information. Trying to monitor progress and productivity or highlight any potential non-compliance is almost impossible.



## Replacing spreadsheets and databases with a consolidated Laboratory Sample Management System

Local databases can often create similar issues to those caused by spreadsheets. Primarily these are the difficulties encountered when trying to enforce standards and common data quality practices.

If you are looking to replace your spreadsheets and legacy databases and move your data into a new consolidated Laboratory Sample Management System, the data quality and integrity improvements will be immediately evident. Many of these systems will enable you to:

- Standardise and simplify data entry using picklists
- Prompt users to enter mandatory information at critical points in the process to eliminate data gaps
- Prevent invalid data formats and characters, such as forcing users to enter numbers only
- Automatically populate data values or records as part of a process
- Cascade updates to related records to maintain data integrity
- Capture audit information to track changes to records or individual data values
- Restrict data access to authorised users only

Before you can use your new system to manage and improve your data quality, you have to migrate your existing data into that new system. Ironically, in order to migrate your data into the application it has to meet all the validation criteria set to ensure its quality. Therefore, you will need to manipulate, cleanse and validate your information into a format that your new system will accept.

This can seem like a daunting task. However, there are some simple steps that you can follow that will help you to break down this task into manageable stages.

There are also some underlying essentials that you should understand before you start. The most important piece of advice that will really help you is to:

***Decide what data you really want to migrate.***

Why spend time, money and resources working on data that you never look at, never use, is not linked to anything that is still relevant or you have no idea what it is? This data would be better archived where your team can still access it as and when required.

## Laying the foundations for data migration

### 1. There are no shortcuts

Unfortunately, when it comes to formatting, validating and cleansing your data to bring into the new system, there will always be some work that you need to do. Although there are tools and companies that can help you, they will be unable to detect context and relevance. They require standards and guidelines to work within. Moreover, you will need to review your data in line with GDPR and other data compliance regulations to ensure you are compliant. You and your team will have to do some groundwork. A further important point to note is that your software provider is highly unlikely to take on this responsibility for you.

### 2. You should clean, validate and format your data before it goes into the system

You are implementing this new system and it has all the tools to help maintain and improve your data quality and integrity. However, relaxing the rules to allow your data to be migrated so you can clean it once it is in the system, is not a good idea. No customer has ever gone back and tidied the data once it is in the new database system. The reality is that the day-to-day job gets in the way and the result is legacy data that your team cannot use. The poor quality of that data will then be blamed on the new system.

### 3. You must learn to understand your data

Often the people who are responsible for migrating the data are not the data owners or the people using the data daily. Involving the people who understand the data will help you to make informed decisions about the data and understand its context. It is difficult to validate your data, such as identifying incorrect values, formatting issues or duplicates, if you do not understand it and how it is being used. Further, you cannot successfully discuss or introduce data standards such as mandatory fields or picklist options if you do not know how or what data can be obtained.

### 4. The same data validation and formatting rules should apply when adding new data

Whether your users are manually entering data or bulk importing information into the system, the same validation and formatting rules should be enforced. This will ensure consistency and help maintain data quality. In addition, if you decide to integrate with a 3rd party application, such as laboratory instruments, the quality and format of the data being imported must adhere to the standards implemented in your Sample Management system.

### 5. Data quality should be frequently monitored. It is not a one-off exercise

Even once your data has been migrated to your new system it is something that needs to be monitored and managed regularly. Unfortunately, ensuring data quality is not a one-off exercise.

## Your 7-Step Guide to Data Migration

### STEP 1: START NOW!

No matter how much time you think you need to sort out your data – it will take longer. System implementations can be significantly delayed due to data issues. Hence, it is so important to start as early as possible on your data migration preparation work.

### STEP 2: KNOW YOUR DATA SOURCES

You need to understand the size of the task. Identifying all the data sources that are to be included in the new system is essential. However, equally as important is recognising the person/people who understand those data sources. It might not always be the person who you think it is or should be.

To locate all of the data sources, you may find it easier to send an email with a simple template attached to the department or team leaders involved in the project and ask them to work with their team to compile the list. A fundamental part of this step is to highlight any duplicates or overlapping data in the disparate data sources and if so agree which of the duplicate records are the most accurate and relevant to be migrated.

An example template is outlined below.

Team / Dept	System Name	IT / Tech Owner	Business Owner/ Super User	Type of Data Records Held	# Records (All or Subset?)	Integrated / Shares Data with	Skills to Extract Data	Mandatory Outputs

### STEP 3: OBTAIN DATA SOURCE EXAMPLES

The more examples of the data sources you have available to you, the easier it is for you to be able to make informed decisions on validating the data. For example, your users may want to make a selected field mandatory. However, when you look at the existing data it may indicate that that particular data element has not been consistently captured to date. This will allow you to delve further into the actual requirement to assess whether the data can be accurately updated with the mandatory element or if it is not practically possible to obtain.

In addition, this exercise will also prove whether your teams do actually have the skills to accurately extract the data from the relevant data sources.



## STEP 4: CLASSIFY ALL THE RECORD DATA TYPES AND VALUES

This is where a spreadsheet can be an invaluable tool. This job can take a little time. It involves reviewing each of the data sources with their appropriate owners (super user and IT) to go through and identify record types, fields and values. The aim of this exercise is to highlight common data values and entities across all your various data sources. It is easier if you have a data example to work with when going through this exercise as well as any existing database diagrams or schemas for legacy applications. At the end of this process, you should have a list of record types, such as Sample, Storage or Participant, a list of data fields within each type, such as Sample Format and Sample Volume and possibly a list of values for picklists to standardise data entry in the fields.

*Do not include fields where no data has ever been captured.*

It might sound obvious, but customers often include fields with no data where somebody thought “we may capture this one day”. If you have not managed to capture that data in several years, it is highly unlikely you are going to capture it in the future. If you do need it later, then some advanced Laboratory Sample Management Systems can be customised to add new fields as required.

An example template is below. The different data sources are shown as column headings allowing you to see where the common fields are across the different data sources.

Record Type	Agreed Common Field Name	My Datasource 1	Data Type (and Length)	Picklist	My Datasource 2	Data Type (and Length)	Picklist
Sample	Volume	Amount	Text		Volume	Int	
Sample	Volume Unit	Unit	Text (30)	Yes			
Sample	Format	Type	Text(100)	Yes	Format	Text(100)	Yes
Sample	Date Time Taken	Date Time Taken	DateTime		Taken At	DateTime	
Sample	Reference	Number	Text(50)		ID	Text(80)	

## STEP 5: IDENTIFY THE OUTPUTS TO HIGHLIGHT PRIORITY AND MANDATORY FIELDS

Understanding the way in which the data is used will help identify mandatory fields and those that require standard data values (i.e. needs an associated picklist of options). Some initial questions to ask are:

- What information is required for any management or team reports?
- Do you need to provide customers or suppliers with information? If so, what?
- Which data is typically used to identify samples to include/exclude in processing (such as analysis, experiments, requests)?
- Are labels or documents to be printed? If so, what information is required on these?

Once you have ascertained the critical pieces of data required for you and your team to carry out their daily operations, you can update this on your spreadsheet template of field and table classifications.

As part of this process you will also need to review the example data sets to make sure all of these contain data for fields you have identified as mandatory. If there is missing data, then you will need to assess how you are going to backfill these gaps and if it is possible to do so accurately.

When assessing picklist options for fields to standardise data entry, you should check the values used within your example data sets to see if you can create a meaningful consolidated list. You may need to create a translation matrix for values, for example, in My DataSource1 the Storage Type may be classified as 'Room Temperature' and in My DataSource2 the same Storage Type may be identified as 'Ambient'.

## STEP 6: CREATE A DATA CLEANSING AND FORMATTING PLAN

Once you have identified your data sources, the actual data (including any subsets) and the data values to be migrated, the next step is to create a data cleansing and formatting plan. When creating a plan, some key things to consider include:

- Have you created a template or guidelines for the people that are cleansing and formatting the data to use?
- Do you know who is to is going to cleanse and format the data? What is their availability and how long will it take?
- Can the data cleansing and formatting be completed within the original data sources or does it have to be exported and manipulated?
- Do you have any tools or scripts you can use to cleanse and format the data?
- Are there duplicates across your data sets? If so which are you going to migrate and how are you going to identify the most accurate/up-to-date records to be included?
- What testing/validating are you going to perform on the data to check that is has been formatted/cleansed as required?

## STEP 7: CREATE YOUR DATA MIGRATION PLAN

It is difficult to determine the best approach to data migration until you have answered some of the questions outlined in the previous steps in this document. In addition, your data migration plan will impact the overall implementation plan for your new system and should, therefore, be included as part of your overall software project plan.

If you have multiple data sources and it is going to take some teams longer than others to validate their data, you may want to consider a phased approach. If using a phased approach, consider rolling out training on the new system just prior to each migration. There is little advantage to training your users months in advance of their data being migrated into the system if they cannot use the system until their data has been transferred.

*It will take more than one attempt to successfully migrate your data into the new system.*

Be prepared to carry out several tests before your data is successfully migrated. The larger and more complete the data set examples are that you give to your software supplier, the more accurate the data migration tests will be. In addition, ensure you understand the test reports that will be provided by your software supplier and that the data migration and testing responsibilities for both you and your supplier are clearly defined and understood.

## ABOUT INTERACTIVE SOFTWARE

At Interactive Software, we have over 25 years' experience in helping our customers transform their lab processes and provide greater transparency and insight into their data using our innovative Laboratory Sample Management Software, Achiever Medical.

Migrating existing data into Achiever Medical is an integral component of most of our customers' implementations. Our experienced business analysts and project managers ensure that the data migration plan is discussed and agreed as early as possible in the project process. Our expert team can offer practical guidance and advice to help your team whilst our innovative range of tools enable data to be imported as a one-off exercise or as part of a phased approach.

**For more advice on how you can overcome your data challenges and gain real lab insights, contact us:**

**+ 44 (0)121 380 1010**

**[enquiries@interactivesoftware.co.uk](mailto:enquiries@interactivesoftware.co.uk)**

**[www.interactivesoftware.co.uk](http://www.interactivesoftware.co.uk)**