Data Integrity in Regulated and Accredited Environments
Part 2: Strategies for Ensuring Comprehensive Data Integrity

Dr David Trew BSc(Hons), PhD, CChem MRSC

1 Introduction

With the current scrutiny, regulated and accredited organisations need to adopt a proactive strategy to provide a high level of assurance that all records and data are both reliable and trustworthy. Organisations not only need reliable and trustworthy data and records, but also need to ensure the integrity of their records and data can withstand scrutiny by sceptical regulators and other stakeholders.

This paper will apply the fundamentals and principles of data integrity, discussed in the previous paper in this series, and recommend approaches to developing and establishing a comprehensive data integrity management system that is designed to ensure the trustworthiness and reliability of all records and data produced by the organisation. A data integrity strategy is essentially a collection of policies and procedures which are used by your staff in their daily work. These policies are designed to provide a high degree of assurance that all of the records and data created by the organisation during the conduct of its operations accurately reflect the events that occur during your organisation's operations. In addition, the data integrity policies and procedures need to provide a high degree of assurance that those records also remain complete and reflect their original content, and have not been altered without retaining their original content. In light of the wide area that data integrity covers it is recommended a multi-disciplinary team be appointed to manage the development and implementation process.

2 Data Integrity Management Master Plan (DIMMP)

It is recommended that a Data Integrity Management Master Plan (DIMMP) be created to serve as a roadmap to control and direct data integrity activities, and which:

- Discusses the organisation's philosophy and strategy to data integrity management
- Establishes a management organisation to oversee data integrity management processes
- Defines roles and responsibilities for members of the management organisation
- Establishes an appropriate data integrity culture within the organisation
- Defines and discusses the risk assessment and management strategy
- Identifies the policies and procedures that need to be established
- Identifies staff training requirements
- Determines how compliance with data integrity policies and procedures will be monitored

- Establishes mechanisms for protecting data and records from being lost or damaged in the event of a disaster and ensuring records will be available throughout their lifetime
- Establishes mechanisms for identifying and investigating incidences that may adversely affect the reliability of records and data
- Establishes mechanisms for correcting and preventing non-compliance with data integrity policies and procedures

The DIMMP is a key quality document that helps a variety of stakeholders who have particular interests in the data integrity management process. In particular, the DIMMP helps senior management estimate how the data integrity program impacts time, people, and money. All members of the data integrity team know their tasks and responsibilities and it helps plan all necessary activities into the schedule, with no 11th hour surprises! In particular, the IT department understands how to support data integrity activities. Finally, clients and auditors understand the firm's approach to assuring the reliability and trustworthiness of its records and data.

3 Policies and Procedures

Policies that need to be established include:

i. Good documentation practices
ii. Prohibiting sharing of computer accounts
iii. Prohibiting use of computer accounts by anyone other than their authorised user, this should include such practices as sharing passwords with other people. As using someone else's computer account can amount to criminal fraud, the policies should include sanctions such as dismissal from employment and reporting to law enforcement
iv. Prohibiting using someone else's electronic signature. As with using someone else's computer account signing a document using someone else's electronic signature can amount to criminal fraud, the policies should include sanctions such as dismissal from employment and reporting to law enforcement
v. Defining the legal status of electronic signatures as legally equivalent to a traditional handwritten signature
vi. Password management policies such as
    - Minimum length
    - Complexity
    - Expiry
    - Re-use
vii. Account management policies such as
    - Username format
    - Account privileges
    - Disabling when no longer required
viii. Audit trails
    - Disabling prohibited
    - What will be captured
    - Review
ix. Data review
x. Monitoring Compliance
xi. Identifying, investigating, tracking, correcting and preventing non-compliance
xii. Backup and archive

In addition to the policies, standard operating procedures (SOPs) need to be established to cover

i. Account management. This should cover the opening, suspension and disabling of accounts
ii. Data review. This should explain who is responsible for carrying out data reviews, and in particular should explain what needs to be reviewed to achieve the necessary confidence in the reliability of the data being reviewed. Some audit trails create many entries, it is important for the reviewer to understand the significance and meaning of these entries, and whether it is necessary to review each entry.
iii. Data backup. This should cover responsibilities, the backup and restore process and the process for confirming the integrity of backed up data. In addition, it should establish a schedule for both the backing up of data and for confirming the ability to restore data
iv. Monitoring compliance with data integrity policies, such as audits. This should define responsibilities, establish procedures for identifying, investigation and appropriate metrics for tracking non-compliance. In addition, procedures for correcting incidents of non-compliance, together with prevention plans should be established.

4 Data Integrity Culture

Many of the practices that undermine the reliability and trustworthiness of data and records appear to be motivated by unwillingness to accept results which did not support some particular preconceived requirement, such as batches of drug product meeting specifications, with the consequences of having to reject out of specification products and the resulting loss of revenue. When an organisation develops a reluctance to accept results which do not conform to some preconceived expectation it undermines the entire purpose of quality control testing.

If the management of an organisation is assuming an out of specification result is due to a laboratory assignable cause, such as analyst error, instrument malfunction or an issue with the validity of the test method, in the absence of evidence to the contrary. This undermines confidence in the entire laboratory testing process, and would lead to questions about the validity of the all the results created by the laboratory, including those that do conform to predetermined specification or expectations. It is fundamental that all scientific work is approached with an open mind and without preconceived conceptions as to what the final results will be.

It is therefore imperative that an appropriate quality and data integrity culture is established within the organisation. This culture should reflect management's philosophy on quality and can be achieved by establishing policies that are aligned to the quality and data integrity culture and develop an environment of trust, where all individuals are responsible and accountable for ensuring patient safety and product quality. The organisation should also establish general ethics and integrity standards which should clearly define the expectation of ethical behaviour, such as honesty. These expectations should be communicated frequently and consistently.

Personnel must be fully aware of the importance of their role in ensuring data integrity and the implication of their activities to assuring product quality and protecting patient safety. This should be communicated to and be well understood by all personnel, which should also

include why the standards were established, and the consequences of failing to fulfil the requirements.

Unacceptable behaviours, such as the deliberate falsification of data, unauthorised changes, destruction of data, or other conduct that compromises data integrity should be addressed promptly. Disciplinary action may be taken, when warranted. It is particularly important that all members of staff understand that data integrity issues, and especially the falsification of data, can have extremely serious, even fatal, consequences for patients. In addition, data reliability issues can have very serious consequences for the business and could even affect its commercial viability. It is also important emphasise that data fabrication and falsification can result in criminal exposure for the individual members of staff. This can include prison time and the inability to secure future employment. Conversely, acceptable behaviour should be appropriately recognised.

Management should not put undue pressures on members of staff that may result in non-compliance with established ethical and integrity standards. Realistic work expectations should be set, considering the availability and allocation of resources.

A confidential mechanism, supported by company policy and procedures, should be established that encourages personnel to bring instances of possible breaches of the ethics and integrity standards to the attention of management, without consequence.

This culture, ethics and integrity standards needs to be initiated by the most senior management within the organisation and should be communicated to all levels. This culture can be facilitated by policies of transparency, openness and approachability.

It is recommended that the installation of the data integrity and quality culture starts during the induction process of new members of staff, and frequent refresher training is carried out thereafter. This could include discussing incidences where data integrity has been questioned, together with the consequences for patients or customers, business and individuals.

5 Risk Assessment and Management

The current regulatory and quality climate is centred around the assessment and management of the risks associated with the failure of quality systems. A number of risk assessment tools can be applied, however, one of the most common is Failure Mode and Effects Analysis (FEMA). This entails:

i. Identifying the potential consequences of an event and assessing the severity of those consequences. This can be done from the perspective of the patient or customer, regulatory consequences and the impact on the business.
ii. Estimating the likelihood of an event occurring
iii. Estimating the likelihood of detecting the event

These three components are then combined to give an overall risk factor.

The severity of a data integrity issue can be assigned to one of three categories[14]:

High severity should be assigned where there is a significant risk of harm to a patient, such as:

- Fraud, misrepresentation or falsification of data.
- Concealment of a product failing to meet specification at release or within shelf life.
- Reporting of a 'desired' result rather than an actual out of specification result when reporting of QC tests, critical product or process parameters.

Medium severity should be assigned when a practice could impact on the product, but with no risk to patient health, or where there is no impact on the product but there is evidence of widespread failure, such as:

- Data being miss-reported, e.g. original 'in specification' results, are altered to give a more favourable trend.
- Reporting of a 'desired' result rather than an actual out of specification result when reporting of data which does not relate to QC tests, critical product or process parameters.
- Failures arising from poorly designed data capture systems (e.g. using scraps of paper to record info for later transcription).
- Bad practices and poorly designed systems which may result in opportunities for data integrity issues or loss of traceability across a number of functional areas (QA, production, QC *etc.*). Each in its own right has no direct impact to product quality.

Low severity should be assigned when there is no impact on the quality of the product or limited evidence of failure, such as:

- Bad practice or poorly designed system which result in opportunities for data integrity issues or loss of traceability in a discrete area.
- Limited failure in an otherwise acceptable system.

Estimating the likelihood of some events occurring presents some challenges. This is particularly so when the initiator of an event requires dishonesty of an individual. This is the reason why installing the data integrity and quality culture, discussed in **Section 5.3**, is such a key activity.

The greatest opportunity to mitigate the risks associated with data reliability is to maximise your ability to detect data integrity events. This can be done by installing audit trails on computer systems and instituting rigorous quality review of all key data and records, with a particular focus on reliability and trustworthiness.

6 Training Strategy

A comprehensive training program is a key element of the data integrity strategy. All new members of staff must be inducted into the organisation's data integrity and quality culture before commencing routine work. In addition, all members of staff should receive regular current awareness training on trends in data reliability and trustworthiness, such as the practices that undermine data integrity. In addition, all staff should receive training in the fundamentals and principles of data integrity discussed in **Chapters 2 and 3** of the first paper in this series.

Training should also be given those charged with reviewing and assuring the reliability and trustworthiness of data and records. This should include instruction on to detect potential integrity issues and attempts to fabricate and falsify data. In particular, training needs to be provided on how to review and interrogate audit trails.

7 Strategies for Assuring the Integrity of Electronic records

In the modern business environment much data and many records are created in electronic format. Electronic records in the pharmaceutical industry are regulated by Annex 11 in the European Union and Pharmaceutical Inspection Convention countries, and 21 CFR Part 11 in the United States. Assuring the integrity of electronic records presents some particular challenges, such as, in the absence of appropriate controls and safeguards, it is possible to:

- Alter and manipulate records without leaving any evidence of the change
- Both intentionally and accidentally delete records leaving no evidence the original record ever existed
- Lose all electronic records held by a system in the event of a computer crash

In light of these challenges it is imperative that adequate controls and safeguards are established. These would usually include implementing policies and procedures to.

i. Restrict access to computer systems to authorised personnel. This is usually achieved by allocating each user an account on a computer system which is specific to the individual user and must be for their sole and exclusive use. This will allow the organisation to attribute electronic records unambiguously to a specific individual. Computer accounts must not be shared between different users, such as group accounts, as this will undermine the ability to be able to **attribute** e-records. It is also important to disable user accounts when a member of staff leaves the company or is assigned to a role not requiring access. Leaving redundant user accounts active presents an opportunity for unauthorised access to a computer system.

ii. Establish adequate access and e-signature controls. Access to user accounts are controlled by the use of, which in combination with the account username and passwords which can used as an electronic signature to authenticate records. These e-signatures can be considered legally equivalent to traditional handwritten signatures. Therefore, passwords must be specific to a particular individual, must be kept confidential, and must not be shared under any circumstances. Policies must also be established and staff must be specifically instructed not use each other's' accounts and passwords; as doing so is the electronic equivalent of signing someone else's signature and constitutes the criminal offence of fraud.

iii. Establish controls to prevent passwords being discovered by someone other than their authorised user such as: minimum length, complexity, expiry and re-use. In addition, you must establish procedures to address suspected compromise of login credentials which should include a requirement to change the password and for the event to be reported and investigated by an authorised person. Also to protect against unauthorised access, user accounts should be locked if there are more than a specified number of consecutive unsuccessful attempts to access an account. Again such event should be investigated and tracked.

iv. Create Different types of accounts for different roles. In particular, users should only be assigned privileges necessary to perform the tasks associated with their assigned roles. For example, on a chromatographic data system; laboratory staff performing analysis should be allocated privileges to create sequence, method, raw data, result and report

files; process data *etc.* But will not have the privilege to create, suspend and disable user accounts, reset locked accounts, move or delate files, these functions are assigned to an administrator. Also, supervisors or data reviewers should only be allocated the privileges necessary to review sequence, method, data and audit trail files. But should not be able to create sequence, method, raw data, result and report files or reprocess data.

v. Ensure audit trails are establish and activated on all computer systems that handle and/or store records of activities, associated with legal, regulatory or key business functions should be equipped with activated audit trails. These audit trails should record all activities carried out or records created and any changes made to those records. In addition, audit trails should record
   - what was entered, processed or changed
   - The date and time of activity
   - Who performed the activity
   - The reason for the activity

   Audit trails must not be disabled or altered under any circumstances, doing so will seriously undermine the integrity of your records. Audit trails must be appropriately reviewed during record and data review activities.

vi. Ensure, before allowing access to a computer system, that all users are certified as proficient with the system's use. In addition, a list of all current and past authorised users, together with their associated privileges should be maintained for all computer systems. This will also facilitate compliance with GMP, GLP and ISO requirements.

vii. Ensure all records, created by a computer and maintained by a computer system, are defined together with the purpose of the record. It is particularly important to identify which records are considered raw data that are used to make quality decisions. Changes to the system should be controlled through the usual change management processes

viii. Create back up and achieve strategies

ix. Create policies and procedures to control removal of files from a server.

8 Electronic Data Security on Standalone Computers

The use of computers that are not connected to the corporate net present some particular risks. When data is stored on a standalone computer it is often stored on the computer's hard disk drive along with the files for the operating system and other programs. If an error develops in the operating system, it is possible all the stored data will be lost.

Standalone computers present risks if data needs to be transferred to another computer. Transferring data from standalone computers is often done using Universal Serial Bus (USB) devices. However, this can present significant risks of data loss, due to their small size and ubiquity, USB drives are often not well controlled and are hard to track physically. They are often stored in bags, backpacks, laptop cases, jackets, trouser pockets, or left at unattended workstations. Thus, there is significant risk they could be misplaced and lost. A further significant vulnerability is the introduction of malware that could compromise all of the data stored on a network. If you do need to use USB devices to transfer data from standalone computer who should establish appropriate controls for their use. For example:

- Avoid copying particularly sensitive personal data information on a USB device.
- If you absolutely must put sensitive information on a USB device, encrypt it first. Well-known encryption programs can be downloaded from reliable websites and used to encode information so it can't be viewed without being decoded first.
- Use secure devices. Some other devices have built-in encryption which eliminates the need to use a separate software program to scramble your information.
- Only permit specific designated USB devices which have been individually marked with a specific identifier.
- Establish controls to track their use and location.

Although it is possible to establish all the controls discussed in **Section 5.2** on a standalone computer. It is not possible to remotely monitor activity on a standalone computer in real time. Thus, an attempt to gain unauthorised to a standalone computer, resulting in an account lockout due to consecutive failed login attempts, will not be detected until the authorised user attempts to logon.

A further issue with standalone computers is that backing up data generally cannot be carried out without human intervention especially if the data is being backed up onto a CD with a maximum capacity of 4.7 GB, would require someone to change the disks if large amounts of data needed to be backed up.

Although consumer and industry level Blue-ray discs are typically 25 and 50 GB, respectively, these require compatible readers which are not available on older systems.

9 Data and Record Lifecycle

Understanding activities associated with the collection, processing, reporting and storage of records and data are key requirements in ensuring their reliability and trustworthiness. Once the lifecycle of a data process has been identified you then need to identify the how the integrity of the data or record could be compromised at each stage of that process. The final phase consists of developing and establishing controls designed to mitigate the potential for compromised data or records. This is illustrated in the following example.

9.1 Data Process Associated with HPLC Analysis

The determination of the amount of drug substance in a dosage form, such as a tablet, employing high performance liquid chromatography (HPLC) is one of the most routine analysis carried out in a pharmaceutical quality control laboratory. A typical HPLC analysis will consist of the following twelve step process, which need to be performed in the following order:

1. Preparation of HPLC mobile phase
2. Preparation of solvents for preparing sample and standard solutions
3. Preparation of standard solutions, these are usually prepared in duplicate, and a standard recovery performed, to provide evidence that the standards have been correctly prepared.
4. Preparation of system suitability solutions (if this is appropriate, as quite often the standard solutions are used to demonstrate the HPLC is operating correctly)
5. HPLC method set up, includes setting the following parameters on the HPLC
    - Flow rate

- Gradient profile (if applicable)
- Run time
- Injection volume
- Sample temperature (if applicable)
- Column temperature (if applicable)
- Detector wavelength
- Analyte retention time(s)
- Integration parameters
- Standard amount

6. System suitability check. This should include injecting: Blank, Standards, and System suitability solution. This will confirm that the HPLC is uncontaminated, the standard solutions have been correctly prepared and the HPLC is functioning satisfactory.
7. Sample preparation. This will normally entail weighing the sample.
8. Injection sequence setup which will include:
    - Entry of sample and standard identity
    - Sample amounts
    - Sequence parameters such as sample dilution factors
    - Raw data and processed data file names
9. HPLC analysis of samples. This will create a raw data file for each injection. All the raw data files in the same sequence should normally have the same base file name, and each individual file is identified by a sequential number
10. Integration of peaks. This will create a processed data file for each injection. These should normally have the same name as the corresponding raw data files, but have different extension.
11. Calibration by comparing the peak area in the sample chromatograms with the peak area in the standard chromatograms
12. Calculation and reporting of analyte amounts. This will create a result file for each injection. As with the result files, these should normally have the same name as the corresponding raw data files, but have different extension.

The next stage is to identify how the data could be compromised at each step in the process. The principle data integrity concerns in the quality control laboratory is the either the falsification of data to conceal test results that show that the is out of specification (OOS) product batches, or using inappropriate manipulation of the raw data processing parameters to cause, an otherwise out of specification result, to yield a result that complies with acceptance criteria. The steps that could be manipulated to falsify results include steps 3 and 7 the preparation of standard and sample solutions, respectively, and steps 5 and 8 which entail entering the standard and sample amounts into the chromatographic data software. In addition, step 10, which is the integration of the peaks, can be manipulated to inappropriately change the peak areas

Initially the reliability and trustworthiness of a set of analytical results can be confirmed by examining the consistency and integrity of each stage of the analytical process that produced the results. This can be done by confirming, from the date time stamps, that all the component events in the analytical process were performed in the expected sequence. Most of the component events create some documented evidence of the time it was performed.

In steps 1 and 2, if the preparation of HPLC mobile phase, and sample & standard solvent entails preparing a buffer of salt solution the balance and pH meter printouts will document the time of preparation. Of cause, it is not possible to measure the pH of a solution until after it has been prepared, therefore the time on the balance printout should precede that of the pH meter. In step 3, the time of standard solution preparation may be a key event in the analytical process, as these solutions may have limited stability. The preparation of the standard solutions will usually entail weighing of the standard reference material and the time the weighing was carried out can be determined from balance printouts.

In step 4, as with the preparation of the standard solutions, the system suitability solutions will usually entail weighing standard reference material and any other substances, which is also used to assess system suitability. Therefore, the date time stamp on the balance printout will establish the time of the weighing's for the preparation of this solution.

In step 5, the saving of the HPLC method file will create a date time stamp; this should precede the collection of any chromatographic data, such as the system suitability check.

Running the system suitability check, in step 6, entails injecting and running sample solutions; this will create a date time stamp each time an injection is made. These should come after the preparation of all solutions required for this step and the time the HPLC method was saved.

The time of sample preparation, step 7, may be a key event in the analytical process, as sample solutions may have limited stability. Sample preparation normally entails weighing the sample, if only to provide evidence of the correct number of dose units were used in the preparation. The balance printout will establish the time the weighing's were made, and sample solutions should be made up as soon as possible after the weighings are made.

Saving the sequence file, in step 8, will create a date time stamp; this should precede the start of the collection of sample chromatographic data. Raw and processed data files should be sequentially numbered starting at 1 and should have no gaps. Any gaps may be indicative of missing files.

During the HPLC analysis, step 9, a date time stamp will be created each time a sample is injected, corresponding to the time the raw data file was created. If an auto-injector is used, and they invariably are these days, the injections should be made at regular time intervals (to within one or two seconds), corresponding to the chromatographic run time plus time to carry out any post run processing and the time taken to inject the next sample.

During the processing of the raw data files to integrate the chromatographic peaks, step 10, a date time stamp is created corresponding to the time the processed data file (which is sometimes called a result file) was created. This may either occur

- during the post run processing prior to the next sample, or
- after all the samples have been injected.

Whichever option is used by the software it should be consistent for every injection within the sequence, and the date time stamps should reflect this.

Following the creation of all the processed data (or result) files, the next step is to turn the raw peak areas into meaningful results which represents the amount of analyte in the sample.

This is carried out during the calibration, calculation and reporting stage of the analytical process, steps 11 and 12. A further date time stamp is created at this point which corresponds to the creation of the report files. This will invariably occur only after all the result files have been created. The date time stamp on the report should become progressively later as the file number increases.

The chromatographic data software records each time a sequence is processed, generally the software will require a reason to be entered every time s sequenced is reprocessed. The number of times the sequence was reprocessed should be checked together with the reason entered each time the sequence was reprocessed. Multiple reprocessing of sequences without satisfactory explanation may be indicative of an attempt the manipulate the final result by manipulating the peak integration, and should be treated with suspicion. The reason for reprocessing a sequence should reflect the actual purpose for reprocessing. Words like update and reprocess are not sufficient.

10 Controlling Chromatographic Integration

High Performance Liquid Chromatography (HPLC) and Gas Chromatography (GC) are two of the most routinely used techniques in the modern analytical laboratory. The specificity, sensitivity and flexibility of the two techniques make them readily applicable to the determination of an enormous range of analytes in a wide range of sample matrices. The raw output from both techniques is a graph of signal strength verses time, usually called a chromatogram. This consists of a series of peaks which represent the component substances of a sample. The area of the peaks is usually proportional to the amount of the respective substance in the sample.

In order to quantitate the analytes in the sample it is necessary to determine the area of the peak; a process known as integration. Modern HPLC and GC instruments are invariably interfaced with computerised chromatographic data systems which are capable of performing the integration process automatically. The role of the analytical chemist is to select appropriate values for the parameters (such as: slope sensitivity, noise threshold, peak width, area threshold, and bunching factor and skim ratio) which are used by the processing software to define the respective chromatographic peaks.

As assigning the most appropriate chromatographic integration parameters requires a certain amount of judgement on the part of the analytical chemist, this presents the following issues from the quality assurance and data integrity perspective:

- Assuring that the integration parameters are used, and therefore the peak integration itself is performed in a consistent way, and
- Assuring that that the integration of peaks, (which when integrated and calibrated against a chromatogram of a standard solution of the respective analyte, may yield an out of specification result), is not manipulated in such a way as to give a result that conforms to specification. In other words, using the integration parameters to manipulate the peak area, in order to obtain a passing result from a peak that would otherwise produce a result that would fail to meet specifications. This is sometimes called integrating, or testing, in to compliance.

The proper integration of chromatographic peaks and the processes used to achieve accurate peak integration attracts significant scrutiny during regulatory audits, and the accreditation assessment process. Therefore, the assignment of integration parameters and the integration of chromatographic peaks will be controlled by a standard operating procedure (SOP) or similar document, and follow a scientifically sound process.

In particular, Changing the integration parameters until the integration looks good is unacceptable[12]. Indeed, multiple re-integrations of chromatograms without explanation or justification are considered indicative of attempts to manipulate the respective peak areas to obtain passing results.

In addition, some authorities believe the same integration parameters should be used for each chromatographic run, once a method has been validated and is being used for the routine quality control operation. Indeed, some regulatory agencies advocate 'securing' chromatographic methods to assure the same integration parameters are always being used[13]. We, however, does not advocate 'locking' methods to prevent changing parameters, as it does not allow atypical situations (such as extra peaks) to be readily addressed. We do, however, recommend including suitable chromatographic integration parameters in the approved written method, together with a picture of a typical acceptable chromatogram, and allowing the analytical chemist to make appropriate adjustments, should this be necessary in unusual circumstances. In these situations, it may be necessary to initiate an investigation, in order to determine the root cause of the unusual event.

It is therefore important that you establish appropriate policies and scientifically sound procedures that provide a high level of confidence that chromatograms are integrated in a consistent manner.

11 Backup and Archive Strategies

The enduring records of your organisations operations are the records and data created during those activities. As these records are the sole testimony of the quality of your products or services it is critical that they are retained in a manner so they are readily available should they be required. During an audit you should be able to produce data and records within thirty minutes, or 24 hours if the records and data are off site. Electronic records and data should be backed up in a manner that preserves their accuracy, completeness and integrity.

It is important to understand the differences between the terms backup and archive. For the purpose of this paper backup is used for short and medium term storage where there is a need for immediate recovery. Archive is concerned with long term storage and immediate recovery is not required.

11.1 Developing Your Backup Strategies

A backup strategy is a set of procedures that you prepare and implement to protect your important digital content from hard drive failures, virus attacks and other events or disasters. The purpose of developing, establishing and maintaining backup processes are to be able to recover:

- From data loss in all circumstances, such as hard drive failure, virus attacks, theft, accidental deletes or data entry errors, sabotage, fire, flood, earth quakes and other natural disasters.

- To an earlier state, if necessary, because of data entry errors or accidental deletes.
- As quickly as possible with minimum effort, cost and data loss.

In addition, an effective backup process should require minimum ongoing human interaction and maintenance after the initial setup. Therefore, it should be able to run automated or semi-automated. You should discuss your back up strategies in your Data Integrity Management Plan.

The first step in planning your backup strategy is to identify what needs to be backed up. In a regulated or accredited environment, you will almost certainly need to seek a comprehensive backup or all of your files associated with your organisation's operations, in order to comply with regulatory requirements. In addition, this entails going through your documents, databases, and files, and identify which files and folders you need to include in your backup plan. As some of this content is irreplaceable, the backup strategy need to protect against all events. Therefore, a good backup strategy should employ a combination of local and offsite backups.

Local backups allow you to back up a huge amount of data at low cost, and are also useful for their fast restore speeds, allowing you to get back online in minimal time. Whereas offsite backups are needed for a wider scope of protection from major disasters or catastrophes not covered by local backups.

Another major consideration, when planning your backup policy, is how often you back up your data. Some folders are fairly static and do not require frequent backed up. Other folders are frequently updated and therefore should be backed up more frequently, such as once a day or more often. Your decision regarding the frequency of backup should be based on a worst case scenario. For example, if tragedy struck just before the next backup was scheduled to run, how much data would you lose since the last backup. How long would it take and how much would it cost to recover that lost data?

You would typically want to run your backups when there's minimal usage on the computers, as backups may consume some computer resources that may affect performance. In addition, files that are open or in use may not get backed up. Scheduling backups to run after business hours is a good practice providing the computer is left on overnight. Backups will not normally run when the computer is in "sleep" or "hibernate mode". Some backup software will run immediately upon boot up if it missed a scheduled backup the previous night. Since servers are usually left running 24 hours, overnight backups for servers are a good choice.

Another major consideration when developing your back up strategy is the type of backup. There a number of different backup types, and each has its own advantages and disadvantages. These will be discussed in more detail in **Section 5.10.2**.

A further major consideration is the type of storage media you wish to store your data on. There are several different types of available storage media which are discussed **Section 5.10.3**.

As part of your backup plan, you also need to decide if you want to apply any to your backups. For example, when backing up to an online service, you may want to apply compression to save on storage cost and upload bandwidth. You may also want to apply compression when backing up to storage devices with limited space like USB thumb drives.

If you are backing up particularly confidential data to an offsite service, you might wish to consider data encryption. Encryption is a good way to protect your content should it fall into

malicious hands. When applying encryption, always maintain records of your encryption key in secure locations, such as in sealed envelopes in fireproof document safes. You will not be able to restore it without your encryption key or phrase.

A backup is only worth doing if it can be restored when you need it most. It is therefore critical that you periodically test your backup by attempting to restore it. Some backup utilities offer a validation option for your backups. While this is a welcome feature, it is still a good idea to periodically test your backup with an actual restore.

Simply copying and pasting files and folders to another drive would be considered a backup. However, the aim of a good backup plan is to set it up once and leave it to run on its own. You would check up on it occasionally but the backup strategy should not depend on your ongoing interaction for it to continue backing up. A good backup plan would incorporate the use of good quality, proven backup software utilities and backup services.

11.2 Types of Backup

There are a number of different backup types and this is a compilation of the most common types of backup with a brief explanation of their meaning, common examples, advantages and disadvantages of each backup type.

1. **A Full Backup** is where all the selected files and folders will be backed up. When subsequent backups are run, all the files and will be backed up again. The advantage of this backup is restores are fast and easy as the complete list of files are stored each time. Full backups are useful for projects, databases or small websites where many different files(text, pictures, videos etc) are needed to make up the entire project and you may want to keep different versions of the project.

   The disadvantage is that each backup run is time consuming as the entire list of files is copied again. Also, full backups take up a lot more storage space when compared to other types of backups.

2. **Incremental Backup** is a backup of all changes made since the last backup. With incremental backups, one full backup is done first and subsequent backup runs are just the changes made since the last backup. The result is a much faster backup then a full backup for each backup run. Storage space used is much less than a full backup and less then with differential backups. However, recoveries can be slower than with a full backup and a differential backup.

3. **Differential Backup** is a backup of all changes made since the last full backup. With differential backups, one full backup is done first and subsequent backup runs are the changes made since the last full backup. The result is a much faster backup then a full backup for each backup run. Storage space used is much less than a full backup but more then with Incremental backups. However, recoveries are slower than with a full backup but usually faster than Incremental backups.

4. **Synthetic Full Backup** is similar to an incremental backup where the incremental backups are combined with the existing full backup. The end result is a full backup that is indistinguishable from a full backup that has been created in the traditional way. This results in greatly reduced recovery times. As it doesn't require the backup operator to

restore multiple tape sets as an incremental backup does. Synthetic full backups provide all of the advantages of a true full backup, but offer the decreased backup times and decrease bandwidth usage of an incremental backup.

5. **Full PC Backup or Full Computer Backup**, in this type of backup an image of the hard drives of the computer is backed up rather than the individual files. With this type of backup, you can restore the computer hard drives to its exact state when the backup was done. Not only can the work documents, picture, videos and audio files be restored but the operating system, hardware drivers, system files, registry, programs, emails etc. are also be restored.

6. **A Local Backup** is any kind of backup where the storage medium is kept close at hand or in the same building as the source. It could be a backup done on a second internal hard drive, an attached external hard drive, CD/ DVD –ROM or Network Attached Storage (NAS). Local backups protect digital content from hard drive failures and virus attacks. They also provide protection from accidental mistakes or deletes. Since the backups are always close at hand they are fast and convenient to restore.

7. **An Offsite Backup** is when the backup storage media is kept at a different geographic location from the source, this is known as an offsite backup. The backup may be done locally at first but once the storage medium is brought to another location, it becomes an offsite backup. Examples of offsite backup include taking the backup media or hard drive home, to another office building or to a bank safe deposit box. In addition to the protection offered by local backups, offsite backups provide additional protection from theft, fire, floods and other natural disasters.

8. **Online Backup** are ongoing or done continuously or frequently to a storage medium that is always connected to the source being backed up. Typically, the storage medium is located offsite and connected to the backup source by a network or Internet connection. It does not involve human intervention to plug in drives and storage media for backups to run. The storage data centres are located away from the source being backed up and the data is sent from the source to the storage data centre securely over the Internet. Although many commercial data centres now offer this as a subscription service to consumers, the customers may not have control over the management of the facilities within the data centre, such as configuration changes or upgrades which could have implications for those operating in a regulated or accredited environment.

9. **Remote Backups** are a type of offsite backup where you can access, restore or administer the backups while located at another location. You do not need to be physically present at the backup storage facility to access the backups, but you do need to be able to electronically access the software to administer the backup. Online backups are usually considered remote backups as well.

10. **Cloud Backup** is where data is backed up to a service or storage facility connected over the Internet. With the proper login credentials, that backup can then be accessed or restored from any other computer with Internet Access. However, as you do not have management control over the associated infrastructure this is probably not an option for those working in a regulated or accredited environment.

11. **FTP Backup** is a kind of backup where the backup is done via FTP (File Transfer Protocol) over the Internet to an FTP Server. Typically, the FTP Server is located in a commercial data centre away from the source data being backed up. When the FTP server is located at a different location, this is another form of offsite backup. However, as with cloud backup unless you have management control over the associated infrastructure this is probably not an option for those working in a regulated or accredited environment.

11.2.1 Backup Media

Regardless of the repository model that is used, the data has to be stored on some data storage medium.

Magnetic tape is probably the oldest medium still commonly used for bulk storage, backup, archiving of data. However, as tape is a sequential access medium, access times may be poor, and the tape is vulnerable to damage if handled incorrectly.

Magnetic hard disks are a commonly used medium for bulk storage, backup, archiving of data. The main advantages of hard disk storage are low access times, availability, capacity and ease of use.[7] External disks can be connected via a range local interfaces or via longer distance technologies. Some disk-based backup systems support data deduplication which can dramatically reduce the amount of disk storage capacity consumed by daily and weekly backup data. However, hard disks are easily damaged, and their stability over long periods of time is relatively unknown.

Recordable CDs, DVDs, and Blu-ray Discs are commonly used with personal computers and generally have low media unit costs. However, the capacities and speeds of these and other optical discs are typically an order of magnitude lower than hard disk or tape. Many optical disk formats can only be written to once and are then protected from alteration, which makes them useful for archival purposes. The use of an auto-changer or jukebox can make optical discs a feasible option for larger-scale backup systems. Some optical storage systems allow for catalogued data backups without human contact with the discs, allowing for longer data integrity.

Solid state storage devices are convenient for short term backup up of relatively low data volumes. As a solid-state drive does not contain any movable parts they are less susceptible to physical damage that their magnetic drive counterparts. In addition, they can have huge throughput in the order of 500Mbit/s to 6Gbit/s. However, as discussed in **Section 5.7** careful controls must be established, especially for physically small devices, to prevent loss.

Remote backup services are gaining in popularity as broadband Internet access becomes more widespread. Backing up *via* the Internet to a remote location can protect against some scenarios such as, fires, floods, and earthquakes, which would destroy any backups in the immediate vicinity along with everything else. There are, however, a number of drawbacks to remote backup services. First, Internet connections are usually slower than local data storage devices. Residential broadband is especially problematic as routine backups must use an upstream link that's usually much slower than the downstream link used only occasionally to retrieve a file from backup. This tends to limit the use of such services to relatively small

amounts of high value data. Secondly, if a third party service provider is used, users must trust it to maintain the privacy and integrity of their data. In addition, the user normally has no control over infrastructure management, such as changes to the configuration and upgrades. This may preclude the use of a third party service provider in a regulated or accredited environment. Ultimately the backup service must itself use one of the above methods so this could be seen as a more complex way of doing traditional backups.

12 Strategies for Managing and Assuring the Reliability and Trustworthiness of non-Electronic Data

Assuring the integrity of non-electronic data is just as important as assuring the reliability of electronic data. Assuring the reliability of paper data is also an exercise in developing systems, policies and procedures by applying the fundamentals and principles of data integrity discussed in the previous paper. Traditionally, records and data have been documented using one of two systems:

i. Notebooks and logbooks, and
ii. Worksheets

When recording data in notebooks or logbooks, or on worksheets, those notebooks, logbooks and worksheets should of pre-paginated. Worksheets should be paginated in the format page x of total pages, to assure there are no missing pages. In addition, each book and worksheet should have a unique identifier. Books and worksheets should be issued in a controlled manner by a specific individual. A log should be maintained for each book and worksheet issued, to whom it was issued and when it was issued. The log should also record when the book was completed and archived. If books are later copied for archiving, appropriate controls should be established to prevent subsequent data entry.

When using books, the data should be entered directly on to page in a consecutive manner. Pages should not be left blank for later data entry.

If information needs to be attached to a page of a book or a worksheet, such as balance or pH meter printouts, or photographs. These should be attached with acid free glue and sticky tape. The printout or photograph should be signed and dated such that the signature and date cover both the printout or photograph and the page. In addition, the book and page number or worksheet number should be written on the printout or photograph.

It is important to establish a standard format for recording dates. These must be unambiguous, acceptable formats include: 4 Feb 2016, Feb 4, 2016 or 4 February 2016. Formats that represent the month with a number are unacceptable as these are ambiguous.

13 Strategies for Detecting and Handling Non-Compliance with Data Integrity Policies and Procedures

Non-compliance with data integrity policies and procedures has the potential to result in some extremely serious consequences for both individuals and the organisation.

In the health products sector the most significant consequence is the potential of substandard products being delivered to patients, whose systems have been compromised by disease,

which could cause further and even fatal complications. In addition, there are significant regulatory consequences that can include fines, injunctions, seizure of products and the total failure of the business. There are also significant potential consequences for individual employees who fail to comply with data integrity policies and procedures which can include dismissal from employment, disqualification from employment in the health products sector and criminal exposure.

In some of the sectors where organisations have been accredited under ISO 17025 although there are no potentially fatal consequences for the organisation's clients. Non-compliance with data integrity policies can still result in substandard data being delivered. This can result in erroneous decisions being made, which could adversely affect the client's commercial interests.

In light of this it is imperative that your organisation establishes mechanisms to monitor compliance, to detect and investigate incidences of non-compliance. The first step in this process is to identify and understand the vulnerabilities which could result in non-compliance. These could include:

- Unauthorised access to computer systems
- Inappropriate user privileges
- Corruption and loss during data transfer activities
- Introduction of computer malware
- Loss of data due to some disaster
- Accidental or deliberate deletion of data, records and documents
- Misplacement of data files, records and documents
- Human error
- Selective reporting

The next step in the process is to implement mechanisms which are designed to monitor the vulnerabilities and detect incidences of non-compliance. These could include:

- Audits. Audits should be carried out on a regular basis and include
  - Instrument audit trails to ensure there are no unaccounted entries for which no corresponding reports exist. The presence of unaccounted entries would suggest that 'trial or demo analysis' are be carried out, coupled with selected reporting.
  - Instrument usage logbooks for unexplained entries. This, again, would suggest that 'trial or demo analysis' are be carried out, coupled with selected reporting.
  - Laboratory notebooks for uncompleted work or unreported results.
  - Reconciliation of issued laboratory notebooks, worksheets and batch records
- Review of all records compiled during operations, together with associated audit trails. It is important that reviewers should be trained in the meanings of all audit trail entries, and in particular, should understand which are the most significant.
- Investigation of client complaints and review of client feedback.
- Investigation of staff reports of concerns regarding colleague work patterns or conducts. Employees should be encouraged to confidentially report any concerns they may have regarding the integrity of their colleague's work.
- Monitoring employee's work output. An excessively high sample throughput by a laboratory analyst may be indicative of so called 'dry labing', that is reporting results but not carrying out the tests.

The final part of this strategy is once an incident has been identified, that could affect the reliability and trustworthiness of data or records, is to carry out an investigation. Investigations of potential data integrity issues should follow the same pattern an any other irregularity. In particular, the investigation should focus on determining the root cause of the event. As once this has been identified more effective corrective and preventative actions can be implemented. For an investigation to be meaningful it must be:

- **Thorough**, investigations should consider all aspects associated with the event.
- **Timely**, investigations should be completed in a timely manner and usually within thirty business days of being initiated. In addition, any corrective and preventative actions should be implemented within defined time lines
- **Unbiased**, investigations should be carried out without any preconceived perceptions, for example a conclusion of 'analyst error' must only be reached if there is significant convincing evidence to support such a conclusion
- **Well-documented**, all the findings of an investigation should be well documented, the use of appropriately designed forms can assist this
- **Scientifically defensible**, sound scientific principles should be applied to all investigations. A good scientific principle to apply is that of Occam's razor which can be formulated as "hypothesis should not be made more complicated than necessary to explain the known facts".

If the normal investigation process is going to be used to investigate data integrity issues. The DIMMP can simply make reference to those procedures. However, any specific differences should be described.

14 Corrective and Preventative Strategies

Once you have identified the root cause of the data integrity issue the then need to take steps to correct the situation and to prevent a reoccurrence. It is important to understand the differences between corrective action and preventative actions:

- Corrective actions address issues that have already occurred, this requires that you understand what has happened and use root cause analysis to identify fundamental reasons why the event happened.
- Preventative actions proactively address potential issues before they occur. This requires you to conduct trend analysis to identify potential events and address them before they become issues.

Corrective actions will usually be the result of an adverse finding. This could be as a result of an internal or external audit, or an issue being identified during the routine checking of data and records. Irrespective of how the issue was identified, it must first be fully investigated to identify the root cause and remediation implemented to correct the issue. The next step is to identify and implement corrective actions to prevent a reoccurrence. The investigation and the process of identifying and implementing remediation and corrective actions must be fully documented. The investigation, remediation and identification of corrective actions will usually follow the same procedures used to handle any other quality deviation.

Preventative actions will usually be the result of a customer recommendation, of an adverse observation or warning letter issued to another organisation or of emerging regulatory expectations. A preventative action is a proactive measure taken to prevent an issue arising.

Although preventative actions can usually be handled through change control procedures, a specific procedure for identifying preventative actions is usually necessary.

15 Conclusions

In the current quality and regulatory environment organisations that are subject to regulation or accreditation should have a comprehensive plan to assure the reliability and trustworthiness of their records and data. However, as this paper has shown establishing a comprehensive data reliability management system is a significant undertaking and it is often helpful to seek expertise from outside of the organisation.