BUTTERWORTH LABORATORIES
science on demand

Data Integrity in GMP Laboratories

Electronic Data Retention: Is Virtualisation our Saviour?

Data integrity is probably the most discussed topic currently affecting GMP laboratories. The drive by regulatory authorities to ensure that the data we produce conforms to the principles of ALCOA (Attributable, Legible, Contemporaneous, Original and Accurate) has presented a raft of challenges to overcome. For Electronic data one of the toughest challenges is presented by the second attribute "Legible".

Can you read your Data?

Most people would consider the output of an electronic data system to be legible; after all it isn't subject to the variations seen in handwritten records. Therefore to say that the "Legible" component of ALCOA presents a challenge when applied to electronic data might seem contradictory. However, one must consider that the actual data is only legible once translated by a compatible computer program and this translation must be validated.

One must also consider that the accepted definition of legible includes a requirement of permanence. To remain legible, the data must be accessible and readable throughout its lifecycle, including any defined archiving period.

Data, including any associated metadata, belonging to currently operational systems should be easily accessible by authorised personnel. The key phrase in that statement is 'currently operational'. What happens once a system has been retired; what can be done to ensure that archiving commitments can be met and that data can be reliably retrieved in its entirety?

The problem is not so much to do with the durability of the data itself but the lifecycle of the analytical data system used to produce and read it and, perhaps more so, the Operating System that it runs on.

Nearly all modern analytical data systems rely on commercially available Operating Systems such as Windows and these are important products for the likes of Microsoft and Apple who are keen on moving consumers away from outdated versions to create fresh revenue for their latest product. The same applies to Instrument Vendors and the software they supply. It is a rarity to find a product near to the end of its life-cycle receiving updates to make it compatible with the latest Operating System. You can try upgrading to the latest Operating System, but there are clear risks in doing so if the Instrument Vendor hasn't claimed (or demonstrated) compatibility.

The subject of cyber security and external threats to analytical data isn't addressed in detail in the data integrity guidance. However, it is a risky policy to not pay attention to this often weak area and as older Operating Systems no longer receive security updates, they are increasingly vulnerable to cyber threats.

Ironically though, there is anecdotal evidence suggesting that very old Operating Systems have become less vulnerable to some of the latest threats due to the incompatibility of their code with the latest viruses.

Mothballing hardware and/or software isn't the solution either as licensed copies of outdated Operating Systems become harder to find and when they are, they may not support current hardware due to a lack of adequate drivers for the newer components. Relying on installing old data systems onto new hardware as part of a data integrity plan is another high risk strategy. Similarly, maintaining old hardware cannot provide any assurances either, as it has a habit of failing just when you need it the most.

Within their guidance, the MHRA allows the migration of data off an original system so that it can be stored in a more enduring and accessible format. The proviso is that all the data is Legible, Original (or a true copy) and Accurate and that all components of the data "story" are either kept together or are retained in an easily accessible and readable format for the same duration. Often the complexity and structure of system audit logs does not permit the extraction of specific metadata as a distinct package. Where this is the case, the original system is necessary to read the metadata and transferring the analytical data to a more enduring file format, such as pdf, may not reduce this reliance.

At worst, migration may result in a physical and logical separation of the necessary components of the complete data "story".

So what is virtualisation and how can it help?

Virtual systems offer a flexibility that is simply not available to more traditional hardware systems and this can be enormously useful in helping to ensure better compliance to data integrity requirements. However, virtualisation can be expensive to install and maintain and also requires a higher level of IT competence to manage. So do the benefits to a smaller laboratory with limited resources outweigh the inevitable costs?

Virtualisation or the use of virtual computers moves away from the traditional idea of a computer being a combination of software running on a dedicated piece of hardware. Instead, we have powerful server hardware sharing resources to a number of installed computers (virtual machines), each of which exists only as software stored in memory. These virtual machines are not reliant on the hardware on which they run so they may be transferred if necessary to other hardware within the virtual environment, sometimes without the user even noticing. This provides the ability to use redundancy and load balancing to both reduce the risk of downtime due to failed components and improve performance.

It is no surprise that the use of virtualisation has become common in organisations that are keen to protect their data and maximise service levels.

So how does this help with the "legibility" problem? Other than the undoubted operational and business continuity benefits, virtualisation also offers an attractive solution for the archiving of legacy data belonging to retired systems. Once a data system has been retired a physical to virtual (PtoV) transfer can be carried out, effectively mapping the physical system onto a newly created Virtual Machine (VM). Adding the virtual platform as a layer between the Operating System and the hardware can enable the running of old data systems on hardware for which there would otherwise be no compatibility.

A general consensus is that virtual platforms should be able to support Operating Systems at least one or two generations old. For example, it is still possible to run Windows XP virtual machines on VMWare, one of the market leaders in virtualisation. This means that one can extend the life of old systems and improve the retention (and recovery) periods for data belonging to retired systems.

As the hardware is current and operational, virtualisation provides a more robust solution than keeping old and outdated hardware running with the increasing risk of failure and lack of spares/support over time.

Don't Wait!

A lot of the obstacles to virtualising a retired data system arise because it is not carried out until the system is retired. At that point the system is old, the Operating System is almost certainly obsolete and compatibility with current generation virtualisation platforms will always be an issue.

So why wait?

Why not virtualise all analytical data systems from the outset, using boxed copies of software and Operating Systems to ensure that whatever occurs, the system remains licensed. This includes any future transfers between VM's if necessary. By doing so, any issues with running the system in a virtual environment can be tested and the setup can be validated there and then, rather than finding out problems later in the day when it may already be too late. This vastly reduces the burden and cost to the lab when retiring systems as effectively nothing changes to the system in the virtual environment. There are other benefits too, such as the ability to remove controlling PCs totally out of the lab environment where even mildly corrosive atmospheres can take their toll on vulnerable hardware.

Controlling instruments in the lab from a VM presents some technical challenges including the connection of non-networked instruments and peripheral devices such as auto-samplers and external detectors. There are straightforward solutions to these issues however, such as utilising Serial to Ethernet adapters and other commonly available standard equipment.

The reliance on the company's internal network is also an important factor and adequate performance and resilience is needed along with adequate security controls. However, from our experience the main hurdle to overcome is often persuading the installing engineer that this is a valid approach as the concept of virtualisation is not always fully understood.

Proceed, But With Caution

Many organisations are only recently starting to appreciate the challenges involved with keeping electronic data "Legible" throughout its entire lifecycle and meeting regulatory requirements for record retention. Computerised systems are becoming ever more complicated and the rapid rate of change of technology means that this year's cutting edge analytical system might be completely obsolete long before it is retired.

Virtualisation is a powerful tool to help extend the lifespan of a system, either in operation or as a means of retaining access to electronic data after the system has been taken out of service. However, using virtualisation without consideration of its own requirements and limitations has risks that should be clearly understood before that path is taken.

Ultimately, virtualisation is itself a computerised system, subject to the same pressures of change and obsolescence that affect analytical systems. There is no way to know what the technology landscape will look like 5, 10 or 15 years from now and policies and procedures need to reflect that uncertainty. Care must be taken before promising clients unrealistic data retention periods and Labs must ensure that they fully understand the limitations of their chosen approach.

Author Biographies

David Hawkins BSc CBiol MBiol MRQA
Managing Director

David started at Butterworth in 1990 as a Junior Analytical Chemist and has had various roles including Senior Microbiologist and Head of Quality Assurance and IT before becoming Chief Operating Officer in 2013 and Managing Director in 2017

David Hawkins BSc CBiol MBiol MRQA
Managing Director

After beginning his career as a chemist in 1992, Colin switched careers to IT in 2002. Pursuing higher education, he gained a degree in Computer Science before studying for a PhD in Artificial Intelligence. He joined Butterworth Laboratories in 2012.