Requirements for Certification of ECRIN Data Centres

with

Explanation and Elaboration of Standards

Version 3.0

October 2015

Requirements for Certification of ECRIN Data Centres, with
Explanation and Elaboration of Standards, Version 3.0

Summary

This document represents version 3 of the 'Requirements for certification of data centres', published by ECRIN, the European Clinical Research Infrastructure Network (the first version was produced in 2011, and the second in 2012). The requirements are the criteria used by ECRIN to identify, and then certify, clinical trials units that can provide high quality, compliant and safe data management, as well as effective management of the underlying systems and IT infrastructure. This latest version results from a review in 2015 by ECRIN auditors, members of ECRIN's data centre Certification Board, and invited experts from a variety of trials units in Europe, and it reflects experience gained auditing trials units in 2014.

Over and above their use for certification, the requirements are intended to describe good practice in data and IT management in clinical research, and in clinical trials in particular. They were developed by senior staff working in non-commercial clinical trials units in Europe, and are intended as a practical guide for staff working in IT and data management in that sector (though the same principles apply to all clinical research environments).

The 129 requirements, or standards, included in the current version are divided into 19 separate lists, some focused on IT, some mainly concerned with data management, and some that deal with both. Each standard has a code, a title, and a single statement summarizing the requirement. This document provides, in addition, explanatory and elaboration material – usually a few paragraphs that attempt to clarify each statement's meaning, and / or give examples of its application, and which also indicate the evidence that would normally be used to assess a unit's compliance. The document also includes a brief introduction to the standards and their development, including a description of the ECRIN audit process, and a glossary of terms.

Contents

Contributors to the document

*Steve Canham**∇, independent consultant

Luca Clivio, Mario Negri Institute, Milano

*Catherine Cornu**, Centre d'Investigation Clinique, Hospices Civils de Lyon, Lyon, France

Will Crocombe, Leeds Institute of Clinical Trials Research

*Nancy De Bremaeker**, Clinical and Epidemiological Investigation Center, Luxembourg Institute of Health

Carlos Domingues, Data Centre, Coimbra Coordinating Centre for Clinical Research, AIBILI, Coimbra

*Michael Faherty**, National University of Ireland, Galway

*François Gueyffier**, Clinical Pharmacology and Clinical Trials Department, University Hospitals, Faculté Laennec, Lyon

Jens Lauritsen¶, Dept. of Biostatistics, Odense University, Odense

Enrico Bjørn Nicolis¶, Mario Negri Institute, Milano

Christian Ohmann¶¶, European Clinical Research Infrastructures Network (ECRIN), Prinz-Georg-Str. 51, 40477 Düsseldorf, Germany

Gilles Palmer, Institut de Santé Publique, d'Epidémiologie et de Développement, Bordeaux

José Miguel Pêgo, Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal; ICVS-3Bs PT Government Associate Laboratory, Braga, Guimarães, Portugal

*Catherine Pham**, PCG Clinical Services AB, Uppsala

*Christian Ruckes**, Inter-disciplinary Centre for Clinical Trials (IZKS) Mainz

*Michael Wittenberg**, Centre for Clinical Trials, Philipps-University Marburg

* ECRIN Auditor

∇ Chair of standard review process, Secretary to the certification board

¶ Certification Board Member

¶¶ Certification Board Chair

SC provided the original draft of proposed revisions and co-ordinated the review process. Others authors provided suggestions, comments and alternative proposals contributing to revision 3.0. All reviewed and approved the final proposals.

1. Introduction and Background

This document describes the systems and functionality that a non-commercial trials unit needs to demonstrate if it is to become certified as an 'ECRIN Data Centre'. It does so by listing a series of standards - some dealing mainly with IT systems, others focused on data management practices, others concerned with more general topics, but all indicative of safe, effective and compliant data storage and data processing.

The 129 standards are divided into 19 different sections, each dealing with a particular topic. Each section is prefaced by a short statement clarifying the scope of the standards within it, or discussing some general issues about those standards.

Each standard is then presented, along with some 'Explanation and Elaboration' (E&E) material (the term has been borrowed from the Consort initiative: http://www.consort-statement.org/consort-statement/overview0/). This material has been added to clarify what the standard means, for instance by providing examples, and to describe the evidence that would normally be used to show that it had been met.

In a few cases additional material has been added at the end of a section to discuss best practice in that area, over and above the ECRIN requirements.

The ECRIN standards are designed to be used as the basis of an on-site audit by appointed ECRIN auditors, who then report their conclusions to ECRIN's Independent Certification Board (see below for further information on the audit process). They are also designed to be used by units for self-assessment purposes, and as a general guide to what is considered to be good quality practice in clinical research IT and data management. The emphasis is on clinical trials in the non-commercial sector, but the same principles apply to data management in non-interventional studies, and indeed to clinical research in any context.

The focus of the standards, the audit and the certification is the *IT and data management activities* of a research unit, even though that unit will usually be involved in many other aspects of the research process - writing protocols, gaining approvals, analysing results, publishing papers etc. This is why throughout the document the research unit is referred to as a 'data centre', or more often just the 'centre'. It is the IT and data management services that the unit can provide, for itself, for external sponsors, and potentially for other research units, that are under consideration.

Certification as an ECRIN data centre is not just an indicator of good quality systems. It is the intention of ECRIN to maintain and publicise a central list of data centres and, once sufficient units have been certified, to encourage the sponsors of ECRIN supported trials to use those centres to provide the data management infrastructure for their trials.

Origin and development of the Standards

The standards are based upon the principles laid out in the International Conference on Harmonisation's guidelines on Good Clinical Practice (ICH GCP). In many cases, however, these guidelines, as applied to IT systems and data management (DM), are rather vague. Working within the EU FP funded project ECRIN-PPI (2008 - 2011), ECRIN's Working Party 10 therefore developed a set of more detailed IT and DM specific standards for trials units, using the GCP guidelines as a starting point but also considering many other international and national documents and regulations.

The rationale for the standards and the way in which they were developed is described in more detail in a paper published in the journal Trials: Ohmann et al., *Standard requirements for GCP-compliant data management in multinational clinical trials*, Trials, 2011; 12:85; available on the web at: http://www.trialsjournal.com/content/12/1/85

The initial version of the standards was published as a supplementary paper to the Trials paper above, in March 2011. This was the version used for the initial audits within the data centre certification pilot phase, at Düsseldorf and Uppsala, in November 2011.

The experience of the pilot phase was used to drive a review the standards, and as a result they were substantially revised. A new version was created (Version 2.2) in July 2012, with 139 standards divided into 21 distinct lists. The revision process, and a summary of the standards, can be found in: Ohmann et al., *Revising the ECRIN standard requirements for information technology and data management in clinical trials*, Trials, 2013; 14:9, available on the web at http://www.trialsjournal.com/content/14/1/97. This paper includes version 2.2 of the standards as a supplementary file.

This version of the standards was also translated into French. The paper describing and containing the French standards: Cornu et al., *Référentiel ECRIN pour la conformité aux bonnes pratiques de gestion des données des essais cliniques multinationaux,* Thérapie, 2015, can be accessed via its abstract at http://www.journal-therapie.org/articles/therapie/abs/first/therapie150042/therapie150042.html.

Because of delays in establishing the legal structure of the new ECRIN infrastructure consortium (ECRIN-ERIC) no certification activity took place in 2013. It resumed in 2014, using version 2.2 of the standards. A review of the standards had always been planned, to take place after each round of audit activity. In 2015 that review began in June, with input invited from auditors, from certification board members, and from specially invited experts in clinical trials IT systems.

The result is this current version (3.0). The changes have been evolutionary but have slightly simplified the requirements: 4 new standards have been added and 3 dropped, while 11 were incorporated into other closely related standards, giving a net loss of 10. The resulting 129 standards have now been divided into 19 lists, after 2 of those lists were merged into related topics. (The details of the changes that were made, and the reasons for them, are available from ECRIN as a separate document).

This version of the standards will be used for ECRIN data centre audits in 2015 and early 2016, after which another review is planned. It is anticipated that changes will become less common in future reviews, unless they are driven by changes in the regulatory and technical environment.

The Audit Process

ECRIN audits are planned to last up to three days, and normally involve a team of three auditors, with the audit results and auditors' recommendations being passed to ECRIN's Independent Certification Board (ICB), who make the final decision about the certification of a unit as an ECRIN data centre.

A centre will be awarded certification if the ICB is confident all criteria have been met. If most of the standards have been achieved, and the auditors estimate that the remainder could be met within a reasonable time, the ICB may request later written evidence, or a

smaller follow up audit, to confirm that changes have been made, after which they will reconsider the certification decision. Otherwise the unit will need to re-apply at a later date.

The audit itself will normally be conducted in English, but ECRIN will try to ensure that the audit team will include at least one individual who can speak, natively, the language of the data centre, so that all evidence can be inspected. ECRIN auditors will be happy to sign confidentiality agreements with audited centres.

Many trials units have experienced radical changes in their processes and procedures in recent years, so that data and IT management now may be radically different than it was even 2 or 3 years ago. ECRIN auditors are interested in the arrangements made for *current* and *future* trials, so will focus their inspection on recent activity and trials that have begun recently, usually within the last 12 to 24 months.

Auditors will expect to see a fully developed quality management system within any candidate unit, with current SOPs and other controlled documents describing most of the areas covered by the standards. Such documents are not sufficient, however - evidence will also be sought of these controlled documents being implemented in practice, by examining trial specific documentation and specific logs, validation records, agreements, meeting minutes, e-mails, etc., as well as interviewing staff. Direct inspection of the centre's systems, especially the clinical data management system (usually only with dummy or test data) will also be required.

The specific evidence that would be expected for each standard is included in this document as part of the Explanation and Elaboration material. This describes only the most common evidence that auditors would expect to see, however, and in any particular case there may be more appropriate evidence available, more relevant to the particular situation of a specific data centre. The references to expected evidence should therefore only be seen as a guide, not as absolute requirements.

Organisational Responsibilities
In some cases, part or all of the functionality covered by a standard may not be the direct responsibility of the trials unit itself, e.g. it may be provided by the parent organisation, or a commercial host, or another collaborating trials unit. Examples are IT services provided by a university or hospital central IT department rather than the trials unit, or a randomisation service provided by an external body.

In such cases the candidate ECRIN Data Centre would be expected satisfy itself, and be able to show any auditors, that the relevant standards were being met. This might involve collecting relevant documents from external organisations, or even carrying out an audit exercise itself on its 'supplier'.

Such activities will often be supplemented by formal written agreements, e.g. in the form of contracts or SLAs, but these on their own are insufficient without evidence they are implemented in practice. Auditors will expect to see the detailed evidence gathered (i.e. by the centre) to support the claims made by third party service providers that they are compliant with the standards.

This point is re-iterated and expanded upon in the Explanation and Elaboration material for individual standards.

2. The Standards

The following pages list the 129 ECRIN standards in their 19 sections. In each case the standard code and title is given in blue, the standard statement in bold black text, and the explanation and elaboration material, usually with notes on expected evidence, are provided below.

IMPORTANT

Vocabulary

Several common terms have specific meanings within these standards and the support material. Please refer to the glossary for definitions of the terms used and explanations of abbreviations.

Phrasing

The standards are expressed in a variety of ways: 'the centre should…', 'the centre can…', 'documents exist…', 'mechanisms are in place…' etc.

For the avoidance of doubt, in each case the standard is actually expressing an **imperative**: the various phrases are all equivalent to '**must**'. It simply sounds a little less arrogant to express them this way, especially when the imperative is repeated many times.

IT01 Management of Servers

The standards in this section are concerned with the servers and related hardware (e.g. network storage) that support the core IT functionality of the data centre. They cover the specification, management and support of that hardware through its life cycle.

The standards *do not* apply to SaaS (software as a service) systems, where the underlying IT infrastructure is the responsibility of an external system supplier, and where the user may be sharing infrastructure with other users or system 'tenants', with little control over the hardware being used. They *do* apply, however, to data centres that purchase IaaS or PaaS systems from external hosts (infrastructure or platform as a service), as well as the traditional situation where the data centre manages its own hardware directly.

Whatever the details of server deployment and management, it will be the responsibility of the data centre to have all relevant evidence available during an audit. The centre may therefore need to gather material from its service providers beforehand and / or arrange that staff and facilities from those service providers are available during an audit.

Contractual and Service Level Agreements (SLAs) between the centre and service suppliers may form part of the evidence for these standards, but the centre should be able to show that such agreements are actually being met - i.e. there is an expectation that a centre will monitor and document the performance of its service providers.

Note that smaller items such as desktop PCs, laptops, and printers are seen as more straightforward to obtain and configure and are *outside* the scope of this section.

IT01.01 Server specification

The centre can demonstrate that they are able to freely specify the servers (and related equipment) they use, with their specifications determined by the software and functions being supported.

Whatever the methods used for obtaining servers and related equipment (for example through the traditional procurement of physical machines, or by purchasing virtual servers from the local host organisation or from a remote hosting facility) the functionality and configuration required needs to be specified beforehand. That specification should come from the data centre and its requirements and not simply be dictated by the options on offer from system suppliers.

Specifications may include operating system requirements (e.g. specific versions to support particular products), or hardware specifications (e.g. particular amounts of memory, storage, connectivity etc.), or functionality requirements (e.g. needs to support a number of databases of a certain size, or a specified number of concurrent connections, with acceptable performance). In other

words, the specifications should match the centre's needs and those of the software it intends to run.

The data centre should demonstrate that it has, or can call upon, the expertise to identify requirements and specify servers and related equipment accordingly. Such servers might be real machines, but are more likely now to be virtual machines with access to defined amounts of processing power, memory and storage.

Quite often one of the 'standard packages' available from an IT supplier, be it a hardware vendor, university IT department or an external host, will be close enough to the specified requirements to be a valid option for use. In others it may be necessary to purchase additional features and / or services, and the data centre should have the freedom to do so.

The most important evidence that this standard is met will come from

- controlled documents that provide an overview of the process;
- the detailed records, including e-mails, of previous and ongoing server acquisitions and / or service purchases, as well as supporting interviews with relevant staff.

IT01.02 Server configuration records

Detailed records of server configurations must be available, allowing accurate rebuild.

The current configuration (operating system version and settings, applications, users, utilities etc.) of each server directly supporting clinical trials activity should be stored. This allows a machine to be accurately rebuilt to the same state if necessary, and also permits further work on a server to be carried out safely, based on full knowledge of the machine's existing state.

For centres using external IT infrastructure (including the institution's central IT department) this requirement will normally be taken care of by processes internal to the host infrastructure, and be transparent to the data centre itself. Taking and storing machine snapshots, nightly and / or before application of patches, is a common mechanism for doing this. In such cases the centre would not normally see the day-to-day records of configuration management, but it still needs to satisfy itself that effective processes are in place to provide such management, and be able to justify that judgement.

For centres directly controlling their own machine configurations, server monitoring systems may allow configuration information to be updated automatically. In others 'snapshots' of server configurations may be taken only at critical time points, e.g. at initial build and before and after major changes.

Using periodic snapshots is acceptable as long as there are accurate records of any updates and patches that are applied *between* those snapshots. All updates should therefore be logged and, along with the configuration snapshot information, the log should always be available (the update log that Windows

automatically maintains on a server is not sufficient, because the times that a server becomes inaccessible is exactly when the details are most likely to be needed).

The evidence required to show this standard has been met would normally be:

- controlled documents or documents from a hosting organisation detailing how server configuration information is maintained and by whom;
- for locally controlled servers, up to date configuration records and patch logs for the servers concerned.

IT01.03 Server support

Hardware support arrangements should be in place to allow equipment to be replaced or repaired in accordance with the centre's own planned times for disaster recovery.

Centres or their host IT organisation should have a maintenance agreement in place, usually with the original hardware suppliers, to allow for the prompt repair or replacement of critical equipment like servers.

For centres using external IT infrastructure this requirement will normally be taken care of by the hosting organisation, and will normally be transparent to the data centre itself. Even though the centre would not be directly involved in managing support arrangements, it still needs to satisfy itself that those arrangements are in place and that they meet the centre's requirements for service continuity, and it should be able to justify that judgement.

Centres with direct control over their own infrastructure will need to develop their own arrangements, with the details of those arrangements, e.g. the response times, often varying with the type of hardware provision (e.g. leased versus purchased, virtual versus physical servers).

In either case, the support arrangements required are also likely to vary with the type of functionality being supported (e.g. an on-line randomisation service versus a CDMS for paper based trials, development systems versus production) as well as the degree of redundancy built into the system (e.g. using clustered servers, mirroring, or log shipping).

Evidence that the standard has been reached include the documents and / or commercial agreements (e.g. SLAs) that detail how repairs and replacements are managed so that specified response and recovery times can be achieved for the various systems used by the centre.

IT01.04 Server retirement

There should be a retirement policy for servers (and related equipment) that takes into account usage, expected hardware lifetimes and support arrangements.

Support arrangements for servers and related equipment (e.g. SANs) are usually only available for a finite time, reflecting the increasing risk of failure as a system ages - even though individual machines may continue to operate without problems well beyond that point.

Older machines are also more likely to suffer from a reduction in performance when the load upon them increases, because of more users or more demanding software, and may therefore need replacement. A managed replacement programme to keep machines properly supported and fit for purpose should be in place.

'Fit for purpose' is an important factor - a server may be too slow in a production environment with a few hundred users, but perfectly adequate in a test environment with just one or two. It is therefore not uncommon for a machine to be 'retired' from production use but still continue in use for many years as a test or training machine.

Leasing and hosting arrangements often have such lifecycle management processes built in, and virtual environments often make the details of the replacement process transparent to the data centre, being managed instead by the host IT organisation. Whoever is responsible for actually carrying out the process, however, the centre should have a policy and / or agreement that ensures the machines it is using are subject to appropriate life cycle management, with retirement from critical functions as necessary, or be confident that such a policy is being carried out on its behalf.

The evidence that this is the case includes:

- controlled documents that outline the life-cycle management policies for servers and related equipment and how they operate in practice (these may be from the IT host organisation);

- for any physical machines whose deployment is controlled by the centre, details of how the servers and related equipment currently used adhere to that policy.

IT01.05 Server maintenance

Necessary patches and updates should be identified and applied in a timely but safe manner to server operating systems, utilities and applications.

This standard requires that there is active management of server patching and upgrades, i.e. a set of procedures that determine how this is done, when, and by whom. Though there can be a risk in *not* applying patches, because they often close security loopholes, there is also an inherent risk in adding a patch or update

to a functioning system. Patch management should include safeguards to try and minimise these risks.

In the standard 'utilities' mean things like programs to support anti-malware systems, remote access and backups, whilst 'applications' include (but are certainly not limited to) databases and clinical data management systems. The standard effectively applies to *all* software installed on servers directly supporting clinical trial activities.

Responsibilities for patching can be complex but must be understood by all parties or there is a danger that some updates will 'fall through the gaps' and not occur at all. In many cases patches to the underlying operating systems and utilities will be managed by the organisation hosting the IT infrastructure, while updating applications will be the responsibility of the data centre, but the situation will vary considerably between centres.

Patch testing for operating systems and common applications may be carried out by specialist commercial patch testing services. Using such a service reduces risk but does not eliminate it, so patch management should still include defensive mechanisms (e.g. taking data backups and configuration snapshots) so that the patch can be rolled back and the system restored quickly to its former state if necessary. Patches and upgrades to less generic programs, like a CDMS, or a statistics package, will often need additional management, e.g. application to a test server and evaluation or re-validation by staff before application to a production server. Like all change management practices, management of patches and upgrades should be based upon a risk assessment – though here the options to consider include making the change, delaying the change, or not making it at all.

The data centre should be aware of when and how *all* patches and updates are applied, including those that it is not directly responsible for itself. It will often need to be involved in patches carried out by the parent or hosting organisation, partly to help warn users of any interruptions to services and minimise disruption, partly because only data centre staff are likely to have the expertise to test specialist systems after patches have been applied.

Evidence that the standard was being met would include:

- local controlled documents, and / or documentation from IT infrastructure hosts, detailing how patches / updates should be applied as safely as possible and who is responsible for doing what;

- specific patch / upgrade records that demonstrate that the patches identified as required, in the context of risk assessment, have been applied;

- discussion with the relevant staff about how the system works in practice.

IT02 Physical Security

The standards in this section deal with the physical security of a data centre's systems and data, including not just protection from intrusion and theft but also from environmental threats such as fire, and system threats such as power loss.

They apply to *all* the physical environments where the centre's data is stored, including those managed by external hosts, and including data stored in SaaS systems as well as on PaaS and IaaS machines. Where the centre does not have direct control of physical facilities, it first needs to *assure itself* that these requirements are met, but also be able to provide the relevant evidence to external auditors.

IT02.01 Locked server room

Servers must be housed within a dedicated locked room with unescorted access limited to specific roles, and with access arrangements known to the centre.

Servers must be located in a locked room, or rooms, specifically allocated for that purpose.

The standard states that the centre, even if it does not manage the server room(s) directly, should still know who is able to have unescorted access to the rooms (not the individuals but the names of roles or teams with such access). Non IT staff that might have unescorted access include maintenance, cleaning and security staff. The centre should know who, in terms of job title, level of experience and seniority, has access to the server rooms and what the relevant procedures would be (possible reasons for access, logging of individual visits etc.).

If the centre manages its own server rooms maintaining compliance with the standard and reviewing access will be straightforward. If this is not the case the centre should still be able to review both the list of those with potential access and the actual access, for instance every twelve months. It should be able to make any concerns known to the IT host organisation, seeking a revision of the list if necessary.

Physical inspection of the server rooms is useful but not essential in assessing the standard. More useful would be local controlled documents, or literature from an external hosting facility, describing the security measures in use and the access policies applied.

IT02.02 Secured power supply

The power supply to servers should be secured, e.g. by an uninterruptible power supply (UPS) unit, to allow an orderly shutdown on power failure.

Servers and related equipment need to be protected from loss of power, at least to the extent that they can be shut down in an orderly fashion. The uninterruptible power supplies and any other equipment used for this purpose should also be tested periodically (according to the manufacturer's recommendations) to ensure that they are functioning correctly.

Physical inspection of the server rooms is useful but not essential in assessing the standard.

More useful would be local controlled documents, or documents from an external hosting facility, describing the UPS and other power security measures, records of testing of the UPS or at least a description of the testing regime, and any records of and discussion about incidents when the UPS became necessary. Many UPS systems generate their own logs documenting tests and power failures, and these can be a useful source of evidence.

IT02.03 Server failure and response

Failure of any server directly supporting clinical trial activity, within normal local business hours, should result in alerts being sent automatically to relevant personnel

If a server does experience some sort of failure it is important that staff are aware of this straightaway, at least during normal local business hours.

In some situations, when supporting sites in a different time zone, it may be necessary to extend the hours covered by these arrangements, so that problems can be responded to and resolved quickly within the business hours of the sites. This may involve centre staff being 'on call' or even working additional hours, but such arrangements will be dependent on appropriate resourcing. The sponsor would therefore need to make the final decision on the time span to be covered and make funds available as required.

Note that this standard covers all servers 'directly supporting clinical trial activity', i.e. it excludes machines used exclusively for test and development, but includes all production machines and those used for immediate backup, e.g. mirrored or failover machines. Failure of a production machine is often obvious because the functionality suddenly disappears, but the centre also needs to be aware of 'silent failures' that may occur in a backup machine, and which may not become obvious until later - perhaps when that functionality is urgently required.

'Relevant personnel' means those that need to react to the failure and start any recovery or failover process. For externally hosted facilities the relevant staff would therefore normally be within those facilities. But wherever located the staff initially contacted should then normally inform the staff who need to liaise with end users, or send messages directly to end users themselves. In the event of a lengthy system failure they would then need to provide periodic updates on progress.

Evidence that the standard has been met could come from inspecting the server monitoring system(s) (or at least descriptions of those systems, in the case of an external hosting facility), looking at examples of any past alerts, and interviewing staff.

IT02.04 Controlled environment

Servers should be housed in a temperature controlled environment

Servers require controlled conditions of temperature and humidity for optimum functioning and any server room should at least be able to maintain temperatures within a defined range.

Evidence that the standard has been met comes from inspecting the server room system(s) or the specification / accreditation of the server room's features. Many environmental control systems can also provide logs of temperature and other variables, which could also act as useful evidence.

If the centre does not manage its servers itself it would still be expected to have the relevant evidence available during an audit, obtained from the organisation responsible for that management.

IT02.05 Hazard control - fire alarms

The server room should be fitted with heat and smoke alarms, monitored 24/7

Servers must be protected from fire, hence this requirement.

Evidence that the standard has been met comes from inspecting the server room system(s) or the specification / accreditation of the server room's features.

If the centre does not manage its servers itself it would still be expected to have the relevant evidence available during an audit, obtained from the organisation responsible for that management.

Other aspects of environmental and system control

Over and above the requirements listed within the standards, there are a range of other types of environmental control which are indicative of good practice. Most external hosting organisations would supply such facilities automatically, but they could be usefully included in locally run server rooms as well.

- An alternative power supply, e.g. from a local generator, to allow continued functioning during a lengthy power loss (UPS systems are usually designed only to last long enough for a managed shutdown).

- An automatic 24/7 intruder alarm system, providing alerts remotely (to security staff and / or senior IT staff) if triggered.

- Automatic 24/7 server monitoring, (rather than the limited hours requirement of IT02.03) with alerts being sent 24/7 to relevant personnel, so that failures can be picked up in the evenings and over weekends and national holidays.

- Success / fail messaging built into scheduled jobs, using for instance the messaging capabilities of PowerShell on a Windows server, or the built in

email services in a modern DBMS. This augments the hardware monitoring provided by server monitoring systems, and provides useful assurance that functionality is continuing as planned. Suddenly discovering that a nightly file transfer process has not worked for the last two months can be both embarrassing and costly!

- Full HVAC (heating, ventilation, and air conditioning) control systems installed in server rooms. These are usually found in commercial and dedicated server hosting environments, but may not be available where premises originally designed for other purposes are used to house servers and related equipment.

- Automatic fire response measures (e.g. inert gas or a misting system) as well as fire alarms.

- Protection against include water ingress (e.g. from external flooding or a burst pipes).

- Protection against infestation with insects or rodents.

IT03 Logical Security

The standards in this section cover protecting data from unauthorised access, from outside the data centre (controlling and differentiating access from within the centre is dealt with in IT04).

Variations between systems and the constantly changing nature of security threats mean that it is difficult to stipulate specific security measures for systems. What is essential, however, is an ongoing review of security risks, security mechanisms and incidents (hence IT03.01) as well as general commitment to the principles of data protection and access control (as illustrated by the other standards in the section).

IT03.01 Security management system

Regular reviews of security (practices, incident analysis, risk assessment, documentation etc.) should occur across all IT systems relevant to clinical trials activity, followed by any necessary corrective and preventative actions.

This standard is equivalent to implementing a *basic* Information Security Management System (ISMS). The term is borrowed from the ISO27001 standard on Information Security Management, though there is *no* expectation that that the centre or its parent organisation has obtained or is seeking full ISO27001 accreditation. The essential features of an ISMS are:

- Identification of security risks, together with an assessment of the potential damage to the centre from a failure in each case.
- Selection and implementation of security controls to reduce the identified risks and to meet the security objectives.
- Continued review and adjustment of security controls as circumstances change and incidents occur and are analysed

An ISMS ensures that "security controls are not merely specified and implemented as a one-off activity but are continually reviewed and adjusted to take account of changes in the security threats, vulnerabilities and impacts of information security failures…." (www.iso27001security.com/html/27001.html).

One would expect an external hosting facility to be able to describe / demonstrate such a review mechanism for IT security - indeed many will have ISO 27001 accreditation. Many universities and university hospitals operate security review groups at the institution level, which is fine as long as the data centre has some means of participating in or accessing that group. Data centres using their own on premise infrastructure will need to develop and demonstrate a security management system themselves.

Evidence that this standard has been met would include:

- controlled documents dealing with system security;

- minutes or other records of a periodic review process and any subsequent corrective or preventative action;
- records of incident analysis and any subsequent corrective or preventative action;
- interviews with staff to discuss how the system operates in practice.

IT03.02 Commitment to data protection

The centre and its staff can demonstrate compliance with and commitment to all relevant data protection legislation, including the provision of related training programmes.

A key component of system security relates to data protection legislation and policies.

Here 'relevant data protection legislation' means that which applies in the countries where trials managed by the centre are carried out, not just the legislation of the centre's own country. For instance, German and Danish data protection regulations would be relevant to a French centre if that centre was running a trial with centres in Germany and Denmark.

The expectation is that staff are made aware of their legal and ethical responsibilities under data protection, as part of their initial and continued training (whether carried out by the centre or external agencies). Controlled documents should also be available that demonstrates the centre's commitment to data protection and how they comply with relevant legislation.

One or more members of staff, in the centre or the parent organisation or both, should be identified as a 'data protection officer' and be available to provide both local support and guidance and advice to management, where necessary, on potential problems in complying with data protection legislation.

The evidence required to show that the standard has been met includes:

- controlled documents that describe how the centre implements data protection policies and the responsibilities of members of staff under those policies;
- one or more staff identified as having special responsibility for ensuring compliance to data protection legislation;
- records of training concerned with data protection (some level of training will be required for all IT / DM staff);
- interviews with staff to check understanding of data protection requirements and discuss how the systems work in practice.

IT03.03 External firewalls

External firewalls should be in place and tested to demonstrate that they block inappropriate access

A centre or (more normally) its host IT organisation should have external firewalls set up to block unauthorised access from outside the centre.

Exactly how the firewalls would need to be configured will depend on circumstances. A centre running eRDC, for instance, would normally have externally facing web server(s) placed in the 'demilitarised zone' or DMZ, logically outside the rest of the institution's network. Centres providing non web based remote access, e.g. through VPN or Citrix, will need to configure their firewalls to support this.

The firewall configurations need testing to check that they are effectively blocking access. But testing has to be against a specification, so there should also be a clear description of the access allowed / prohibited for each of the major systems.

Penetration testing is one possible method. Such testing can be done by commercial organisations but in the non-commercial sector could also be done by arranging mutual testing between institutions. Another possibility is an external audit of the firewall. All tests have to be documented accordingly.

It is also good practice to continually monitor traffic activity and to try and identify and investigate any hacking or denial of service attempts.

Evidence for the standard being met would include:

- explanation of how the firewall configuration worked to block inappropriate access;
- records of firewall specifications and related tests that demonstrate effective blocking of access;
- in the case of externally hosted facilities, equivalent documents that demonstrate appropriate external security. This might include certification against appropriate ISO IT security standards;
- audit certificates or records of penetration tests if applicable.

IT03.04 Encrypted transmission

Clinical data transmitted over the internet to or from the trials unit should be encrypted

All clinical data must be encrypted if transmitted to and from the centre over the internet, to prevent eavesdropping, tampering and 'man-in-the -middle' security attacks.

This will normally be in the context of eRDC, when the https protocol is commonly used to encrypt transmitted information. It may also take place in the context of a

VPN or Citrix connection. In the latter case the encryption should extend to the whole of the data transmission and not just the initial exchange of certificates.

An alternative approach is to encrypt the data before it is sent from the site, for instance using an AES algorithm built into the data capture system. In such cases the data is also stored in an encrypted form. This requires careful encryption key management but the transport mechanism can be plain http.

Centre staff will need to explain how the systems they use support encryption and provide the documentary evidence as appropriate - perhaps taken from the vendor's / developer's specifications of the CDMS.

N.B. Recently (2014) vulnerabilities (e.g. POODLE) have been exposed in the SSL algorithm used for encryption within https. The current recommendation is to only allow TLS based traffic to interact with web servers. Though not currently part of the standard, good practice would therefore be to disable SSL in an eRDC web server and only allow communication using TLS based encryption (so far as client browsers and the eRDC system itself allow).

IT03.05 Server administrator roles

Administrative access on servers should be restricted to specified members of IT staff, and subject to specific access management practices

Administrator level access to the centre's servers should be restricted to a small number of specified staff, usually IT staff within the centre and / or IT hosting organisation with particular responsibility for server management.

More senior staff within either the centre or the host IT organisation should not routinely have administrator level access unless they also have specific server management roles.

Administrator accounts should normally be subject to specific management practices (though these are not always described in a controlled document), so that the security of the access can be maintained over time. For example, it is often necessary to set up one or more *shared* admin passwords to allow easy access to servers or specific services outside normal hours. It might then be necessary to change all such passwords after key staff leave, especially if the leaving was not by choice.

From the point of view of business continuity, it may be a good idea to have some key administrative passwords stored off site (traditionally in a sealed envelope in a safe). This can conflict, however, with the need to periodically change these passwords to ensure that they are not compromised. There is no easy answer to this problem, though using a secure cloud based 'password locker' may work in some cases, as long as it is kept up to date.

The evidence that this standard has been met would include:

- the current list of staff with administrative access, or the relevant documentation / description received from any external hosting facility;

- interviewing staff - to allow them to describe management of administrator accounts and how it works in practice.

IT03.06 Internal blocks on data access

Inappropriate access to centre data from other users of the IT infrastructure should be blocked

Most centres are a part of a larger parent organisation, and share that organisation's IT infrastructure. Similarly, if they use external hosting facilities for some or all of their data they will be one tenant among many within the hosting facility, sometimes sharing the same servers with other tenants.

In either case there is a need to block access to the centre's data from users from other organisations or departments.

For a university, there is a particular need to block accidental or deliberate attempted access by student users, whilst for a hospital there is a need to prevent any unauthorised access into hospital systems from the centre, as well as vice versa.

One method to block access in this way is by using internal firewalls between different parts of the network, but other forms of access control (e.g. domain and user group management) may be used instead of or in addition to firewalls.
The evidence that the standard has been met will include:

- relevant controlled documents describing how access is blocked, or equivalent information from external hosting facilities;

- interviews with staff to confirm how the system works in practice.

IT03.07 Encryption of non-physically secured data

Clinical data relating to individuals should only be stored on protected servers and storage devices. It should not be stored on non-secured devices (e.g. on laptops, desktops, USB sticks etc.) unless encrypted

This standard says that *any* non-aggregated data, i.e. data that relates to individual trial participants, must not be stored on non-secured devices unless encrypted. This includes demographic, treatment and lab details as well as data relating to clinical signs and symptoms - anything that is an attribute of a single study participant or their experience.

Secured devices are servers and network storage devices that are physically secured by being in locked rooms, and logically secured by being within the centre's (or its IT host organisation's) firewall. *Non secured devices* include desktop PCs and laptops as well as USB sticks and CDs / DVDs, which are not encrypted. (Desktop PCs can easily be stolen, and frequently are, even from premises that were believed to be secure).

Please note: No distinction is made between data that contains obvious patient identifying data (PID) and data which does not. This is because PID is hard to

define and the distinction is not absolute. Obvious patient identifying data, like name, initials, and health system number stand at one end of a continuum. At the other extreme is anonymised data without any such items, or links to data that might contain them, and without localising data (either in space, such as hospital name, or in time, such as date of birth).

Some individual clinical data without obvious PID is so detailed, however, and / or so rare, that - especially with some localising data included as well - it *can* become potentially identifying. Such data stands somewhere between obvious PID and anonymised data. To keep things simple and safe therefore, the standard requires *all* data relating to individuals to be encrypted unless it is stored on a secure device.

The level of encryption required should match, as a minimum, the recommendations of the relevant national research or health organisation (128 bit AES in many instances, 256 bit in others). Many centres now routinely provide automatic 'whole-drive' encryption for laptops and USB sticks, which makes it much easier to demonstrate compliance with the standard. This does mean, however, that staff need to be aware that they should not use their own devices or USB sticks for data – only those that are issued to them by the centre.

Evidence for the standard being met can come from:

- the controlled documents describing the policy;
- direct examination of laptops and desktops;
- interviews with staff, e.g. to check their understanding of the relevant controlled documents.

IT04 Logical Access

The standards in this section cover the control and differentiation of access from within the centre (protecting data from unauthorised access from outside the data centre is dealt with in IT03).

The access being considered is to the data centre's own network and to 'all systems directly supporting clinical trial activity'. This most obviously includes the CDMS, but will also include (for instance) treatment allocation and trial administration systems. It excludes systems used exclusively for development, testing and training.

IT04.01 Logical access procedures

Controlled documents covering access control to all systems directly supporting clinical trial activity should be in place

This standard simply requires that controlled documents exist that govern access management, both to the network, which acts as the initial portal, and then to systems involved in directly supporting clinical trial activity.

Network access is often managed by the centre's host organisation, while the centre would normally manage access to its own systems. There will therefore often be two sets of controlled documents.

The evidence will be the documents themselves, which should include a summary of responsibilities, processes, outcomes and documentation involved in controlling logical access.

IT04.02 Access control management

Each system requiring access controls should have mechanisms, e.g. using roles, group membership, etc., that can be used to effectively differentiate and manage access

This standard requests that sufficient mechanisms exist to provide differential access, in terms of both allowed functionality and data. This might be by role assignment in a CDMS, or by explicit allocation of rights within a file management system, and would normally be done through managing group membership rather than on an individual basis.

The standard is concerned with 'each system requiring access controls', starting with the initial log-in to the centre's / parent organisation's network for internal staff, and including access to the CDMS for both internal and remote eRDC staff, but also including any other systems where access control would be expected because they directly support clinical trials activity.

Evidence that the standard has been met would come from controlled documents detailing how access control is implemented, plus direct demonstration of access control mechanisms and the inspection of systems, especially log-in processes.

IT04.03 Granularity of access

Access control mechanisms should be granular enough to allow compliance with the data centre's own policies on access control

This standard (which in practice would probably be considered together with IT04.02) emphasises the need to support granular access, i.e. to allow fine control over the access provided and the functionality provided with it, to different datasets and for different roles.

Granularity clearly applies to remote eRDC staff, who should only ever see their 'own' site's data, but it also applies within the centre, where staff should not be able to see data or other files that are sensitive scientifically, e.g. randomisation lists, or clinically / commercially, e.g. analysis results, unless they have a genuine need to do so.

Granularity may also be found in fine control over access to clinical data - for example a member of staff who works on one study should be able to see and edit the data for that study; her manager might be able to view that data but not edit it; a monitor might be able to raise and close queries for that study but not enter data, etc.

The granularity required should match the centre's policies on access control, themselves driven by the organisation of staff, tasks and systems.

Centres that store more obvious PID (e.g. patient names and addresses used to contact trial subjects in quality of life studies) will usually need to provide greater granularity of access, to protect that data, than centres that do not (or are not allowed to because of local data protection legislation).

Evidence that the standard has been met includes:

- controlled documents detailing how access control is implemented;
- direct demonstration of access control mechanisms and inspection of systems, especially with regard to particularly sensitive data types;
- discussions with staff about how and why the necessary granularity is supported.

IT04.04 Network log-in management

Network log-in management should be enforced on all users, usually including regular change and / or complexity rules for the log-in password

Protecting initial entry to the network for centre staff is a key part of managing access. Normally a process is established that enforces 'strong' passwords and a change after a fixed period (e.g. 90 days), but in some centres biometric devices or personal cards may be used, instead of or in combination with passwords ('2-factor authentication').

Evidence that this standard was met would come from:

- controlled documents detailing the management policies for network log-in;
- proformas and other documentation, and / or demonstration showing those policies being used;
- discussion with centre staff about how the local network log-in policy worked.

IT04.05 Remote access

Remote access should be controlled using the same principles as local access control, and should not normally include access to the host's network (unless the user has a pre-existing identity on that network).

Remote access is used here to mean direct access to a server and specific applications and / or the centre's network, e.g. using Citrix or VPN, rather than the browser mediated access of an eRDC system to data entry screens.

It may be provided for centre staff, who will usually have their own identity on the local network (for instance a monitor when working away from the centre) or for staff who are completely external to the centre, perhaps working for a collaborating organisation.

Remote access management should reflect this. It should prevent external users from gaining access to anything other than the specific applications and datasets that they have been authorised to use, and in particular prevent access through to the host's network. Internal employees may, in some systems, enjoy the same access as they would have if they logged in locally (more often a sub-set), and the remote access mechanisms should be able to manage this effectively.

Evidence for this can be obtained from:

- relevant controlled documents;
- from interviews discussing how any remote access is managed;
- demonstration of the remote access system's access control mechanisms and records, including relevant proformas.

IT04.06 Network lockout

Logins to the network should be locked after a locally determined inactivity period, requiring secured re-activation

When an employee moves away from their machines while logged into the network and / or a particular system, there is a risk that another user may use that machine, 'hijack' their access rights and gain unauthorised entry to systems. There should therefore be an *automatic* mechanism that locks the screen and which requires a password or equivalent mechanism to unlock.

The mechanism must be automatic after a pre-set time- not normally more than 15 minutes.

Requesting that users lock their machines manually does not provide a sufficient guarantee that it will actually happen, though those with particularly high access rights, such as senior staff, may be advised to lock their machines manually before the automatic time-out is triggered. The lock-out should apply to the network log-in and therefore lock the whole machine. Many CDMSs also provide an automatic log-out mechanism but on its own this is insufficient.

Evidence for this can be most easily obtained from direct observation, backed up by interviews with staff.

IT04.07 Administration of access to clinical data

Access rights to systems storing or processing clinical data should be regularly reviewed, changes to access requested and actioned according to defined procedures, with records kept of all rights, when granted, why and by whom.

This standard deals with the administration of access to clinical data systems. It requires that a system is in place to request and implement changes, to record when access rights were changed and by whom and that the rights are reviewed periodically (at least annually) to ensure that they are all still required.

Periodic review is particularly important for remote users, who are often employed by other organisations, and who may therefore leave without the data centre being made aware that they can drop access.

The standard only applies to those systems dealing with clinical data, but it would be good practice to extend the requirement and record *all* access requests / changes, including to the network and other (e.g. trial administration) systems.

Evidence that the standard has been met should come from:

- the relevant controlled documents;
- examples of the request and review procedures;
- the records maintained within the system itself.

IT05 Business Continuity

Business Continuity (BC) is the set of activities performed by an organisation to ensure that critical business functions will remain available to staff, customers, suppliers, regulators (etc.) after a major loss of function. The loss may be caused by a natural disaster (flood, fire, earthquake, hurricane, etc.) or be man-made (e.g. sabotage, walkouts) or be as simple as the sudden loss of key staff.

BC is *not* restricted to IT systems! It can include communicating with clients, storing copies of key material off-site, arranging alternative premises, hiring consultants or temporary staff and finding alternative service suppliers. The IT component of BC is Disaster Recovery (DR): the process of recovery or continuation of IT systems after a massive loss of functionality.

DR may include rebuilding and / or restoring data for applications, and re-establishing hardware, communications and other IT infrastructure. Key to any disaster recovery policy is the retention of copies of data, but so also is keeping copies of other key information (passwords, activation keys, scheduled jobs, user information etc.).

This section deals with business continuity in general (IT05.01) though the rest of the standards are focused on IT disaster recovery.

IT05.01 Business continuity planning

The centre should have or be developing Business Continuity measures and a process for regular review of those measures.

The usual method of trying to ensure business continuity is to develop a Business Continuity Plan (BCP), covering the likely actions in the event of a major loss of function (e.g. fire, long term power failure, full server failure, sudden loss of key staff).

It is recognised, however, that a BCP can take a relatively long time to implement, not only because additional funding may be required, but also because much has to be done in association with the parent organisation. The expectation of the standard is that the centre has such a plan, but it is accepted that it may still be a provisional document, not yet formally agreed with the host organisation.

Many BCPs include a listing of possible 'disaster scenarios', an estimate of the probability and impact of each, and the actions that would help the normal functioning of the centre to be resumed in each case.

Such actions fall naturally into two groups - those that can occur beforehand, as part of the *preparation* for business continuity, and the measures that must be implemented *after* the disaster scenario has occurred. Such scenarios should include a wider range of disasters than loss of IT function, though that may be a component of several of them.

In practice, for many units, the most likely threat to business continuity would be the sudden loss of one or more key staff. The usual mechanisms for dealing with

this (good documentation of activity, deputising arrangements, job sharing or shadowing etc.) would not normally need to be part of a BCP in any detail, but references to the relevant personnel or training policies / documents should be included.

A BCP should not be a static document: planned business continuity measures need to be regularly reviewed and updated as necessary, because situations and threats will change. The standard therefore includes this requirement. The expectation would be for at least an annual review, though again it is appreciated that agreeing any changes with a parent organisation may take time.

The evidence required is the BCP document itself, or documents that show that such a plan is in current development, and the plans for its regular review.

IT05.02 Back up policies

Policies for data backup and restore should match the centre's requirements, and the details of the procedures should be available to the centre.

The first part of this standard requires that the centre is clear about its requirements for data backup and restore. Issues that must be decided include:

- For how long should backed up data be retained? (or equivalently, from how far back should it be possible to retrieve data?).
- Is a nightly backup enough or should backup (of changed data and / or transaction logs) happen more frequently, to reduce the possible work in re-entering data?
- Do the backups need to be encrypted?
- How quickly should it take to restore individual files or databases, or whole machines?
- If the primary data centre goes completely off line, how long should it take to switch to a secondary centre?
- How much monitoring of IT operations (e.g. nightly backup) is required within the centre?
- When should restore operations be tested?

These questions have sometimes been seen as technical 'IT' issues, but in fact they are critical operational issues and need to be documented, considered and approved by the centre's senior management.

Developing matching procedures and controlled documents is relatively straightforward if the centre's own IT staff have direct control over the backup and restore processes. The relevant controlled documents, e.g. SOPs and work instructions, can be generated and approved in-house.

Increasingly, however, data centres use external IT infrastructure and staff. 'External' may mean a central IT department in a university or hospital, or a system vendor, or a completely independent commercial hosting facility.

Unfortunately, there is a tendency for some external hosts to provide a blanket assurance about data backup and restore without providing details. In such situations it is critical that the centre pursue the details of backup and restore arrangements in very specific terms, until they are sure that their requirements can be met. The unit may not be given (or need) access to the host's internal SOPs but they should insist on having all the information they need to make a decision.

The centre's requirements should also be included within any contractual and / or SLA agreements. If the centre's requirements go beyond one of the standard hosting 'package's than these agreements may need to include additional payments, but the key requirement is that it should be the data centre and not the hosting organisation that is determining the backup / restore regime. External hosts who cannot provide the necessary flexibility of provision should be avoided.

The evidence would be the controlled and / or contractual documents, plus discussion with centre staff (and if available staff from the external hosting facility) to explore how the arrangements worked in practice.

IT05.03 Back up frequency

Backups must be taken using a managed, documented and automatic regime that ensures new or changed data is backed up within 24 hours, and which allows the centre to check that the system is operating properly.

This standard on back up frequency reflects the fact that back up regimes are usually sophisticated enough to identify and only process data that actually needs backup because it has been changed or newly inserted.

If a centre is managing its own data backups it is relatively straightforward to monitor that the process is operating properly.

If backups are the responsibility of an IT host organisation the centre still needs to assure itself (e.g. by receiving reports or periodic copies of the logs) that the backup process is operating properly. Ideally this would also be every 24 hours but it is accepted that this may not always be easy to arrange. In such cases the centre will need to take a risk based decision on what level of monitoring is acceptable, given their knowledge of the internal systems within the hosting organisation and the contractual agreements that are in place. External hosts that are unwilling to provide any form of monitoring data or access should be avoided.

In practice there may be several different backup regimes, for instance one that applies to files on a SAN and another that applies to databases on a dedicated server. There may also be mechanisms for taking snapshots of virtual machines as well as (or instead of) conventional file based backup. The centre may therefore need to develop separate documents / monitoring regimes for each.

The evidence that the standard has been met includes:

- documentation describing the backup regime and how it is managed, either from the data centre or the IT host organisation;

- logs of the backup process and / or periodic summary reports indicating the backups are proceeding as required.

IT05.04 Back up storage

Back up media storage (location, protection, redundancy) should be sufficient to avoid data loss if there is a fire or other large-scale disaster.

Simply backing up data does not guarantee that it will survive a large scale disaster such as a fire, especially if it remains in the same location as the original data.

A variety of mechanisms exist to ensure that a such a disaster will not wipe out data, for instance secured off-site storage of tapes, on site storage in fire-proof safes, duplication of back up data to a mirrored site, and twinned but physically separate backup systems (e.g. at opposite ends of a large university or hospital campus)

This standard requires that one of these mechanisms, or something equally effective, is in place to ensure that if a large scale disaster happens at one of the data storage sites a copy of the data is still available. On site storage of tapes in fire-proof safes is a traditional approach but is rarely adequate - it usually only preserves infrequent copies and needs manual intervention. Given the low cost of electronic storage better alternatives are usually available.

Centres using external hosts should assure themselves that connecting to a secondary data centre is a realistic option and one which allows switching within a reasonable (to end users) time period. The problem is that if a whole hosting facility is destroyed there will be a queue of organisations demanding that access to their data is restored. Government agencies and large corporations will probably head that queue, and in practice it might be several weeks before data access was restored to a trials unit. The unit needs to obtain clarification about this and, if it was felt necessary, arrange and pay for a higher priority in the host's reconnection processes.

The evidence that the standard was being met would come from:

- controlled documents describing the procedures for storage of backups and the systems supporting this;

- discussion with staff to clarify procedures and explore how the systems work in practice.

IT05.05 Back up - Environment

Any necessary data management / administration data (access groups, log-ins, scheduled jobs etc.) should be backed up and restorable

Though the retention of copies of data is necessary for disaster recovery, so also is keeping copies of other critical information (passwords, activation keys, scheduled jobs, user information etc.).

This is particularly important for database systems, where the database server may hold a great deal of data management / administration information. This may or may not be backed up automatically by the IT host organisation's systems, and so may require additional agreements or scripts being run by the centre staff. The same sort of data is also necessary for file based systems but this is usually backed up along with all the other file material.

The much greater use of virtual machines, and the practice of taking regular 'snapshots' of these machines, re-applying them to hardware when necessary, is making this standard easier to meet for most centres, especially when using external infrastructure rather than on premise servers.

Nevertheless, it is necessary for the centre to be clear about the regime that is being implemented (see IT05.02, .03) and what components of the environment backup process, if any, remain the responsibility of centre staff, for instance by writing and running scripts.

Evidence that the standard had been met would come from:

- relevant controlled documents and / or details of procedures within external hosts;

- interviews with staff, including explanations and demonstration of the backup / restore mechanisms used.

IT05.06 Recovery Testing

Testing of restore or failover procedures should take place and be documented, at a frequency that reflects system and staff changes (for all servers relevant to clinical trial activity)

Back up is of little use without corresponding mechanisms for restoring data, and those restore mechanisms *must* be tested.

With single or small groups of files this is rarely problematic, but it can more difficult when the need is to rebuild a whole on premise server back to the state prior to failure, or to that of the night before, from the bare machine. Conversely, restore of a whole server is usually straightforward when using virtual machines in an external IT infrastructure, as data centres are increasingly doing, and indeed this is one of the major arguments for using such a facility.

The tasks of the data centre include:

- Identifying the possible restore operations that it might be required to carry out or request, at the level of files, systems (e.g. whole databases) and whole servers.
- Identifying the acceptable allowed time periods for successful restores of different types.
- Setting up tests of those restore processes, or - for external hosts - ensuring that the relevant restore processes are being tested.
- Documenting the test restore exercises (or receiving relevant documents from external hosts).
- Identifying any problems, and, if necessary, redoing the tests until they work without incident.
- Developing a mechanism to review and as necessary repeats test restores, for instance after major changes in the server configuration or back up regime or (for restore mechanisms that are the responsibility of the data centre's own staff) when there are changes to the staff.

Even when an external host organisation does most of the work of restoring files or systems, the data centre staff should still be clear about their own role in any restore process, for example knowing the information that needs to be transmitted to the hosting facility, or any information that needs to be given to end users.

For database based systems, mirrored servers or data duplication (using scheduled replication or transaction log shipping) allows a much more rapid failover if failure occurs and is generally regarded as good practice. It does, however, carry an additional administrative overhead as well as demanding additional hardware, or additional costs if delivered externally. In these circumstances 'restore' and its testing will involve a failover process, but may still include renaming servers or changing IP addresses to ensure that applications point to the right systems.

The evidence that the standard had been met would come from the documented restore requirements of the centre, and the records of test restores, together with a discussion with centre staff about how restore mechanisms are reviewed and repeated.

IT06 General System Validation

As used within the ECRIN standards and related material, 'validation' refers to the process of ensuring and documenting that a system or process is functioning as required. In other words, it should indicate whether or not a system or process can be relied upon to be 'fit for purpose'. This echoes the FDA definition of validation, which is:

> *Establishing documented evidence that provides a high degree of assurance that a specific process will consistently produce a product meeting its pre-determined specifications and quality attributes."*
> (FDA, 1987, Guideline on General Principles of Process).

This section looks at validation in general, of all systems used by the data centre. There are additional specific aspects of validating trial specific database systems (CDMAs) but these are covered in section DM02.

The standards in this section are designed to support a flexible approach to validation, one which stresses the underlying principles of validation more than any particular framework or methodology. Those principles, as conceived by ECRIN in the particular context of non-commercial data centres, are listed below, together with an indication of the standards which support them.

- Planned and documented validation of systems can represent a major investment in time and resources, especially for a small academic trials unit. It is important that the processes, implications and costs of validation are understood at all levels of the centre but especially by senior management. An overall validation policy needs to be endorsed by senior management, as indicated by approval of the relevant controlled documents (IT06.01).

- No organisation can validate every system or process that they use in detail. Resources must be focused on those systems where the impact of error or malfunction would be greatest and / or the likelihood of errors occurring is highest. The key to designing a validation regime is therefore risk assessment. A risk assessment methodology should be applied systematically to identify the systems that need to be validated and the level and type of validation required (IT06.02).

- Even if a system or process is not in the direct control of the data centre (for instance is a software service, or a hardware installation hosted externally) the centre still has a responsibility to ensure that the system has been validated. In other words, centres will need to obtain evidence of validation from the relevant external hosts and service suppliers, and should have that evidence available for inspection by external agencies (IT06.02).

- Validation almost always occurs when a system is first introduced into a centre, but systems change - are patched and upgraded etc. - and both the staff and the context, and thus the requirements on the system, can also change. Validation is therefore an *ongoing* process and centres should have a mechanism to review risk assessment and possible revalidation on a periodic basis as well as during planned change. This applies especially to externally

hosted services, where change may take place without the data centre's knowledge. Centres need a mechanism to assure themselves that the validation status of external services is retained over time (IT06.03).

- Validation of any particular system needs to be planned and then recorded, in detail, to provide the evidence for subsequent decisions. The complexity of systems and their usage means that absolute validation, i.e. of all possible inputs and situations, is impossible. Detailed testing should be sufficient, however, to give a 'high degree of assurance' that the system functions as it should, and that it can be relied upon to function as expected under normal demands. In practice system validation is often done in stages - IQ, OQ and PQ: installation, operational and performance qualification respectively (IT06.04).

- At the end of validation decisions need to be taken, signed and recorded, as part of the centre's overall quality control mechanism. Validation normally provides the basis of the decision to accept, maintain or reject a system for production use, but there is not always a simple link between the two processes. Verifying that a system performs as specified: 'does this system work as advertised?' is different from the acceptance decision: 'does this system work well enough for us to use it?'. The second question demands a risk based decision based on the answers to the first (IT06.05).

- The need to take decisions about systems highlights one of the great values of validation. It is not just about testing a system's functionality. It also allows a subgroup of staff, normally those that will be the system's main users, to fully understand that system and its relative strengths and weaknesses, and to become the local experts in how the system operates. Even though full operational qualification of a system can take some time, this is often an essential first step when introducing a new system to a centre.

- Planned change within systems should be governed by policies that stipulate how those change should be managed, and the responsibilities and workflows involved. In particular, the policies should require a risk assessment of the impact of the change (IT06.06). The risk assessment and any subsequent revalidation plan, together with the results of that revalidation and the resulting decision, should all be recorded. (IT06.07).

- In a busy data centre it is easy for additional system components to be introduced without being validated. This applies particularly to data reports and extractions, which are often added on an ad hoc basis throughout the life time of systems. Again a risk-based approach should be used to validate, as and when necessary, these data outputs (IT06.08).

IT06.01 Validation policies

Controlled documents should be in place covering system validation approaches, responsibilities and processes

This standard requires the centre to have developed controlled documents that describe a validation strategy. Typically, this description would include:

- The *general* principles and approach(es) taken towards validation.
- The scope of validation, i.e. the types of systems considered (but not the individual systems, see IT06.02).
- The method(s) used for risk assessment (see IT06.02).
- Who should do what, in term of the roles within the centre.
- The overall workflow of validation processes.
- The expected outputs.
- The quality control and sign-offs within the process.

The document will often include reference to particular frameworks and models for validation and risk assessment (e.g. GAMP, PIC/S) but they should *not* include detailed descriptions or discussions of those frameworks. 5 pages summarising GAMP 5 does not constitute a validation policy!

Similarly, there is no requirement for any particular framework to be used - partly because those frameworks are themselves evolving, partly because most have their origins in the pharmaceutical industry, and often in manufacturing and laboratory practice rather than the specific validation requirements of data management systems. Existing frameworks can certainly be very useful, but they work better as a starting point for developing local ideas and systems rather than being 'dropped in' as complete, fully formed solutions.

The scope of validation should normally include a***ll the types of system used by the centre to directly support clinical trials***, and not just the obvious ones like a CDMS, or systems classified as 'falling under GCP'.

In most cases scope would *exclude* infrastructure software like operating systems, commercial databases and web server systems, but the decision should be up to the centre. A web server in an unusual configuration, for instance, might be seen as requiring validation and being in scope. A CDMS would almost always be in scope, but so also would any IT based treatment allocation systems, trial administration and eTMF systems, if they are seen as directly supporting clinical trial activity.

The question is more complex for systems hosted outside the centre, perhaps by the parent institution, or the system vendor, or by a commercial hosting facility used by the vendor. Such systems will not always be directly accessible to the centre staff and their initial deployment and validation will normally have been done by someone else. Despite this these systems can (and usually should) remain in scope for validation - but the nature of the evidence will change. Rather than being generated directly by the centre the evidence, or some summary of it, will usually need to be obtained from the hosting organisation (see IT06.02).

In summary, this standard is about the centre showing it is clear about its overall approach to validation, and that the approach has been endorsed by management. It is *not* concerned with individual systems and their validation, which is addressed by IT06.02.

The evidence would be the relevant controlled documents.

IT06.02 Validation system inventory

The centre should have an inventory of all the IT systems in scope for validation, the risks associated with each, and - in summary - the validation strategy for each.

Given the decision about the types of systems in scope for validation (see IT06.01) the logical next step is to list each of those systems and carry out a risk assessment for each. That in turn allows the level and type of validation required to be described explicitly.

This list may form a single document (when it is often known as a '*Validation Master Plan*') or it may be distributed across several documents. The standard only requires that this 'inventory' exist within the centre, where it can be used to direct validation activity.

N.B. 'Systems' can include processes that may not be associated with a specialist software package, but which are associated with specific tasks within the centre. This could include, for instance, data transfer between externally and internally hosted infrastructure, or data extraction and processing to support query management. Such processes may use standard operating system features or standard office or statistical software, but the context in which they are used means they can represent a distinct 'system' as far as the centre as concerned.

The documentation should identify the risks associated with each system and thus the types and level of testing required. It may also indicate who will be involved, when, and what tools they will use, and the nature of the outputs of the validation process.

Usually only a paragraph or two is needed for each listed system - the key requirement is that a risk assessment has been carried out and the validation requirements have been identified. Factors that can influence the risk assessment include:

- *The type of software:* A classification scheme in common use (from GAMP 5) divides software systems into four types (N.B. there is no longer a type 2):
  - 1 – Infrastructure software including operating systems, Database Managers, etc.
  - 3 – Non configurable software including commercial off the shelf software
  - 4 – Configured software, including CDMS, treatment allocation systems.
  - 5 – Bespoke software

This classification is often used to allocate different validation regimes to systems, but it is a very blunt instrument, and many other factors need to be taken into account. Some of those are listed below.

- *The potential impact of malfunction:* A component that contributes to data integrity, or GCP or other regulatory compliance, or is otherwise involved in maintaining patient safety, clearly has a higher potential impact - on patients, the scientific conduct of the trial and the reputation of the data centre - if it operates incorrectly than (for instance) a module allowing users to easily reset their own passwords or a report that gives a breakdown of accrual figures by site / month.

- *The possibility of silent failure:* Some problems in systems are obvious as soon as they appear. They will disrupt work but are unlikely to be allowed to impact the study's results in the longer term because they will be resolved. Other problems are less obvious and may introduce errors without the users being aware of the problem until much later. The costs of resolving the problem, and the potential impact of the issue, are correspondingly greater.

- *The numbers of other users:* Though systems should always be validated in their own local environment, systems developed by established vendors and in common usage will normally carry less risk than specialist, often locally configured systems. Systems with a large user base are usually extensively tested by their vendors, and there will also be a user community that can identify and publicise potential issues.

- *The resources used to develop the system:* Systems that are developed by companies with extensive development resources, and well established quality management practices themselves, are likely to carry less risk than systems created by new and / or small development teams, and especially by a very small in-house development team. (On the other hand the responsiveness of the development team in fixing identified problems often varies in the opposite direction).

External systems, like a CDMS hosted by the system vendor or a web based treatment allocation system, present a particular problem because they will probably not be within the direct control of the data centre or directly accessible for testing. Nevertheless, the centre still has a responsibility to ensure that these systems have been validated and are fit for purpose.

It will therefore need to obtain evidence of validation from the external hosts and / or service suppliers, acting almost as a quality inspector for its own suppliers. That evidence should then be made available for inspection as necessary by external agencies, usually with prior agreement of the system suppliers and if necessary with confidentiality agreements in place. Service suppliers who cannot or will not provide such proof of validation should not be used. The approach taken should be summarised within the validation system inventory.

The evidence that this standard has been met would largely be the validation system inventory itself, as well as discussion with centre staff explaining how risk assessment was applied in practice.

IT06.03 Periodic review of validation

The centre should have mechanisms in place for periodic reviews of the risks associated with systems, with possible subsequent revalidations.

Validation almost always occurs when a system is first introduced into a centre, but systems change - are patched and upgraded - and the context, and thus the requirements on the system, will also change. Centres should therefore have a mechanism to review risk assessment and possible revalidation on a periodic basis - over and above the risk assessment that takes place within managed system change. Any revalidation required will often not be for the whole of a system - just those components perceived as affected by changes need to be retested.

It is worth stressing in this context that a 'system' or 'process' will normally involve hardware, software, and people, and often supporting sub-systems and workflows. For example, a system may be valid with expert users, but not fit for purpose if the users are novices. Even though most *system* changes will trigger a risk assessment (see IT06.06) this is not necessarily true of organisational and contextual change - hence the need for periodic review of the 'whole system'.

At some point a review may indicate that there is less risk involved in retiring and / or replacing a system than trying to continue to use it. Validation is therefore an ongoing process that should last, and ultimately govern, the lifetime of the system.

The need for ongoing review applies particularly to externally hosted services, where major change may take place without the data centre's knowledge. Centres should therefore have a mechanism to ensure that they know of both changes in external services and how the validation status of those services has been maintained over time. Ideally they would receive periodic updates confirming the validation activity of the service supplier.

The system inventory or Validation Master Plan (see IT06.02) can provide a good place to record the dates of validation exercises for each system (just the dates and perhaps a summary of the scope of the validation), and so provide good evidence for this standard. Other evidence for the standard would come from supporting statements within controlled documentation and discussion with staff about how review was implemented, together with related records.

IT06.04 Validation Detailed Evidence

Detailed validation documents should exist for any particular system, detailing the validation carried out, including any test data and protocols, and the results obtained

Each system validation exercise should generate a set of retained detailed validation evidence - i.e. the descriptions of the tests and their results. The documents should also indicate who carried out the tests and when. In some cases these may be electronic rather than paper documents.

It is impossible to test every possible set of inputs into a system, so judgements need to be made on the level of evidence required to show, with 'a high degree of

assurance', that any particular system component is functioning properly. Again those judgements should be made on an (informal) risk assessment, with more effort being made to test the more critical parts of systems.

Many data centre staff are familiar with validation and the associated terminology from the V-model of GAMP 4 - i.e. initial, operational and performance qualification, and use these to structure their validation processes. The three types of qualification are defined below:

- *IQ (Installation Qualification*): checks that a system's installation is correct with respect to the vendor's (design) specifications - i.e. everything is in the right place and the various components / modules are interconnected properly and can be accessed as required.

  IQ is the normal initial step in validating systems. In practice IQ scripts usually check installation by verifying a core sample of functionality, that confirms that all components in the system are accessible and available.

- *OQ (Operational Qualification)* checks that a system is functioning correctly, i.e. against the system's functional specification for commercial systems or the design team's specification for local systems.

  In practice this means establishing, documenting and running through a series of test cases, often supplied for commercial systems by the vendor as an OQ script, that examines each aspect of the claimed functionality. OQ for a major system like a CDMS may take several days or even weeks.

- *PQ (Performance Qualification)* is the process of checking that the system, over a range of 'real world' conditions, continues to perform as intended.

  PQ is an important additional stage because OQ, especially if only using a vendor supplied list of test cases, may not fully reflect the intended usage. It is one thing to confirm that a module works as advertised with 1 user and 20 patients, quite another to check that performance is still acceptable with 50 users and 5000 patients, or to discover that intrinsic limits prevent work with populations (of data items, subjects, logic checks etc.) greater than a certain size.

In practice PQ can often be partly integrated with OQ by designing additional test cases with realistic loads. The context of PQ should also mimic, as far as is possible, actual usage - in particular real users should be involved in some aspects of the testing process. In other words PQ should include some User Acceptance Testing (UAT).

The balance between OQ, PQ and the sign off into production use is another risk based decision process. In low risk scenarios it might be OK to start to use a system after successful OQ, after which the system would be tested / monitored against a steadily accumulating range of real usage conditions. In higher risk scenarios some PQ / UAT will usually be done as well, with users being given access to the system, deployed as it would be for production use, and asked to run additional tests.

There is always a balance between the time and resources spent on validation and the risks involved in not confirming a system's functionality in different scenarios. In the end the validation that is carried out will be a function of the perceived risks associated with a system, including the possible impacts of a malfunction, and the costs and time required for the validation process.

The evidence for compliance would be the detailed validation documents themselves, against a range of different systems.

IT06.05 Validation Summaries

A signed and dated summary of the results of each validation should exist

As well as the detailed results (see IT06.04) any validation exercise should also generate a relatively short summary (often one page) of the validation, signed off and dated by one or more key staff, that confirms that validation has been completed and which indicates its result.

A system that failed a validation exercise would normally then have further documents listing the 'corrective and preventive actions' (CAPA) to be taken to remedy the problems identified. A later and more focused revalidation exercise would then confirm that these actions had been successfully carried out.

The 'result' of validation is not always a simple pass / fail. Often it is about whether the system can go into (or stay within) production use or not, which is not the same thing. For instance, even if a system fails some components of its OQ / PQ testing it still may be acceptable for use if the problems are not critical (i.e. do not affect GCP and regulatory compliance), or a workaround is available, or the system vendor / designer can be persuaded to quickly add or fix the missing functionality. The reality is that the time and money spent on assessing and procuring a system, or building one in-house, and then installing and validating it, are usually far too high for a non-commercial data centre to be able to quickly switch to another system.

The summary documentation should make both the responses to both questions clear: did the system pass or fail the validation exercise and if it did not what are the problems and subsequent CAPA? Is the system suitable for production use, and if so are there any caveats or workarounds that need to be implemented?

The evidence for compliance would be the summary statements themselves, against a range of different systems.

IT06.06 Change Management Policies

Controlled documents should be in place defining risk-based change management mechanisms.

All systems are subject to change, for instance from user requests or vendor upgrades and patches, and those changes should be managed for systems to retain their validation status.

This standard requires that there are controlled documents that should specify the change management process and procedures, as well as the roles and responsibilities involved, and how it is documented. It also requires that the process is risk-based, i.e. that the change management includes a risk assessment of the possible effects of the change on the system, and thus the possible revalidation that might be required.

The documents are often augmented by sample proformas for requesting changes, carrying out a risk assessment, approving the changes, and documenting how and when they were carried out (see IT06.07).
The evidence that this standard has been met would be provided by the controlled documents themselves, together with the associated proformas.

IT06.07 Change and risk evaluation

Changes in IT systems in scope for validation should be documented, and include a documented risk assessment as well as any necessary revalidation results.

If IT06.06 requires that policies are in place that govern change management, this standard simply requires that those policies are used in practice and that there is documentary evidence of this. It also seeks to guarantee that re-assessment / re-validation is integrated into the change management process.

Many centres use a 'check-list' approach to change management that allows common issues to be identified and the decisions taken in respect of each to be easily documented. Questions could include:

- How critical is the functionality being changed?
- Who will be affected by the change and in what context?
- What are the possible impacts on other aspects of the centre's functioning?
- Will documentation and / or training need to be revised to reflect the change?

The response to the first question in particular will dictate how much re-validation of the relevant parts of the system will be required.

Any re-validation would normally generate detailed documentation that would indicate if the relevant parts of the system still functioned as intended, or not, plus a signed and dated summary statement to that effect. Subsequent CAPA based changes would be documented in a similar way.

Evidence that the standard has been met would include:

- change management documentation that clearly reflected this method of working;
- structures (e.g. test systems in which changes can be rehearsed) that supported it in practice;
- discussions with staff to clarify how the systems worked in practice.

IT06.08 Validation of extracted data

Reports and data extractions should be validated for their accuracy, using a risk based approach.

The reports and data extraction facilities that many systems come with 'out of the box' will almost always be validated as part of an initial OQ exercise. The problem is that reports and data extractions are often added on an ad hoc basis during the lifetime of a system, and it is easy for them to slip through the validation net unless there is a deliberate policy to systematically assess the need for possible additional testing.

The approach should be risk based. Relevant questions might include:

- How are the reports / data extractions used? Are they providing critical clinical data (e.g. SUSAR details), quality management data (e.g. query rates by site) or administrative details (accrual figures)? The possible impact of any error in the output will be a major factor in determining the validation effort required.
- How similar or different are the reports / extractions to others that have been shown to work?
- Are there any special characters or values in the data that might cause existing extraction or reporting mechanisms, even if they are well established, to work incorrectly?
- How have the reports / data extractions been constructed? Are they standard reports built in to the system and used (and therefore checked) by a wide variety of users, or are they ad hoc reports only available at a single centre, and perhaps only used by a few individuals at that centre? Do they involve scripts and code generated in-house rather than by the system vendor?
- How complex are the outputs? Are they simple listings or do they contain complex derivations and sub-totals?
- How much transformation of the data was necessary to produce it in the structure and format required? A system that pivots, splits or aggregates data from various sources, or transforms it into another format altogether (e.g. to XML) is more prone to errors than one that simply dumps pre-existing tables into flat files.
- How easy are the outputs to cross check? Would errors be obvious, e.g. by visual cross checking with the data in the databases or with data from other sources, or could errors slip through if not checked in detail?

It should be stressed that not every report / extraction needs to be validated, but every distinct report / extraction should at least be *assessed* to decide if some form of validation should occur.

Many reports can be parameterised, so part of any validation process would be deciding what range of parameters should be checked.

As with all validations, the results should be documented and available for inspection. The relevant policies, records of risk assessment and the validation documents themselves would then form the evidence that the standard had been met.

IT06.09 Validation of data transformations

Data transformation processes should be validated, using a risk based approach.

Reports and data extractions often include data transformations when they are generated, but such transformations can also occur in isolation, for instance changing the format of extracted data (e.g. from XML to SAS, or from the internal database structure to SDTM or ADaM) before transferring it to another institution, or in preparing data prior to importing it into the system (e.g. into CSV files ordered in particular ways).

Like reports, extractions are often added to the centre's processes after initial validation exercises have been carried out on the associated systems. There is therefore a similar risk of them being used without any formal evidence that they have been properly validated.

As with other validation tasks, the process should start with a risk assessment, focusing on the process(es) in which a transformation is used, and how critical those processes are to the overall scientific and data management of a study, and taking into account the same types of factors as listed within IT06.08.

As with reports, when transformations can be parameterised, it is also important to consider what range of parameters should be checked.

As with all validation, results should be documented and available for inspection. The relevant policies, records of risk assessment and the validation documents themselves would then form the evidence that the standard had been met.

IT07 Local Software Development

The three standards in this section only apply to those centres that develop their own software in-house.

'Software' in this context means all types of systems, utilities, code and scripts used to support data management, for instance extraction and reporting routines, complex stored procedures within databases, and trial administration, coding and treatment allocation systems. In some centres the CDMS itself may have been developed locally.

Note that the scope *excludes* statistical scripts generated used for analysis.

In-house systems are subject to the same risk-based validation requirements as any other system, as described in IT06, but they also have specific requirements relating to their development. In particular, it is vital that the centre has the resources to develop and maintain systems properly, and that the systems created are well documented, so that they are not dependent on the staff who created them.

Hence the focus of these three standards is on documentation (IT07.01 and IT07.02) and resourcing (IT07.03). In addition, a number of suggestions for 'good practice' in software development are provided.

IT07.01 Documentation of in-house software

Technical documentation should cover system architecture and deployment, configuration details and the characteristics and purposes of individual modules, files and / or classes.

The focus of this requirement is for a top-down overview of any locally produced system and its architecture, including a brief description of each constituent module, file or class (different structures will be relevant to different types of software). That should include at least a description of the function of each module / file / class, and (if not provided by inline comments) the nature of inputs and outputs. The documentation should complement but not duplicate the more detailed comments that will be found in the code itself (see IT07.02).

Details of deployment, configuration and dependencies (especially if not integral to the build) are especially important, because these are often difficult or impossible to discover from the code itself. They may include details of web server settings, configuration files and their contents, and runtime dependency requirements. Build processes should be scripted or described in sufficient detail for them to be replicated easily.

In total, the level of documentation should be sufficient - when used with the in-line commenting described in IT07.02 - for another competent developer to make sense of the program, start to work on it and deploy it successfully in a reasonably short time (days rather than weeks).

A detailed functional specification is not required by the standard (though one is always useful!) because it is assumed that users would be able to describe the system's functionality if necessary.

The evidence would be obtained from examining the relevant documentation. The auditors' judgement is necessarily a subjective one and it is accepted that it is difficult to agree on what is 'sufficient' documentation. There is also an element of risk-assessment required here – standards of documentation may be set higher with more critical systems. It is relatively easy, however, for auditors to identify systems where documentation levels are clearly too low. For that reason, and because of the importance of documentation in supporting any software project, this standard has been included.

IT07.02 In line Commenting

All code, scripts and procedures should include in-line documentation explaining non-obvious aspects of program execution.

The focus of this particular requirement is bottom-up in-line commenting, so that program execution, particularly when it involves non-obvious processing, is adequately described and the function of individual components can be easily identified.

'In-line' here also includes the headers often found above function or class definitions, describing purpose, input and output parameters, and - in the case of functions - when and from where the code is called. There is no expectation that every function or class is so described, or that every action requires explanation - but anything where the function is not obvious from the code and name should be decorated with comments.

Full descriptive names for functions, classes and variables are strongly recommended as a way of drastically reducing the need for additional comments in code.

The level of documentation should be sufficient that - when used with the overview documentation described in IT07.01 - another competent developer could make sense of the program, start to work on it and deploy it successfully in a reasonably short time (days rather than weeks). Different programmers will have different styles of documentation, so some might use in-line commenting for some information which others would put in separate documents (though in the latter case it would be reasonable to expect in-line references to those documents). The auditors are therefore asked to consider the total documentation available when assessing this and the previous standard.

The evidence would be obtained from examining the relevant code. The judgement is subjective but worth attempting because of the importance of this type of documentation. In addition, it is easier, and arguably more important, to identify missing or clearly inadequate commenting, accepting that 'sufficient' is harder to define.

IT07.03 Resources for software development

The unit should have access to sufficient staff and other resources to support in-house development in the long term.

Within relatively small academic trials units the resources available for IT development can be very limited, sometimes limited to one or two people. This can represent a huge risk for the unit - sudden loss of those staff can (at best) freeze development of the systems and (at worst) lead to systems being abandoned altogether.

Good documentation, of both systems and processes, can do a lot to reduce the risks, but too often a small IT team is under such pressure that they do not have the time to produce that documentation.

Note that the centre only needs to 'have access' to IT staff, they do not need to be part of the unit. They could come from a central pool of IT staff, or from a loose co-operative of developers from different departments or even different institutions, all working on the same system. Centres that use and contribute to open source projects also have access to a greater pool of expertise.

'Other resources' refers to things like training and tools, as well as other physical resources - space, machines, infrastructure support etc. - all of which contribute to the development and maintenance effort.

This standard asks the auditors to make a judgement about the resources available to the centre to support its locally developed systems in the longer term, and the risks it might be exposed to by having too much expertise concentrated in too few people.

As with the other standards in this section the judgement is a subjective one, and the resources required will depend on the extent of in-house development. A unit with a single developer *may* be adequately resourced if all that developer is doing is writing, and fully documenting, reports and extractions on an open source system with an active user community, all contributing similar components. That single developer would be a completely inadequate resource, however, if they were responsible for an entire CDMS system. In fact, trials units that could not guarantee sufficient developer resources should probably be encouraged to use commercially available CDMS systems, because in the longer term the total costs of ownership (which are usually dominated by salaries) may be lower.

As with the other standards in this section, IT07.03 has been included more to allow auditors to point out the dangers of *clearly inadequate* resourcing rather than to trigger long debates about the exact levels of resourcing required. In the context of ECRIN certification, the key requirement is that a centre can maintain continuity of system development and maintenance, even with loss of key staff. It would be difficult to recommend certification of any centre where that was felt not to be the case.

Good practice in Software development
Though not required as a standard, there are a variety of development techniques which would help to indicate high quality practice and which should make systems easier to develop and safer to maintain. Some of these are listed below.

They would not all be applicable to all situations, and it is accepted that opinions can differ (sometimes strongly!) about the relative merits of some of these approaches. In addition, some might be beyond the resources of a small development team. Nevertheless, the presence of some of these techniques would increase confidence in the quality of the in-house development process.

- Design techniques that promoted clear 'separation of concerns' between different parts of a system.
- Use of a source control system that allows branching and release management.
- Programming against interfaces rather than concrete fixed components, with dependency injection.
- Programming against data repositories rather than fixed data sources.
- Use of a unit testing framework and / or integration tests.
- Continuous integration of a test regime with a source control system.
- Use of a library of user controls / common modules across systems.
- Regular code reviews and walk-throughs; shared coding.
- Use of a bug tracking system.
- Use of a scripted build and / or deployment scheme.
- Use of scripts for constructing and modifying databases.
- Consistent and effective error / exception handling techniques.
- Consistent and comprehensive logging techniques.

DM01 CDMAs - Design and Development

A CDMA, or Clinical Data Management Application, is a system supporting data entry and management *for a specific trial*. It includes the databases and files used to store the data and associated notes and queries. It also includes the CRFs (paper and / or electronic) used for data entry and the trial specific data validation checks, skipping logic and derivations that those CRFs contain.

The standards in this section deal with how CDMAs and CRFs are specified and constructed, and how CDMAs in development are isolated from those in production. In addition, several examples of 'best practice', that can make CDMA development and CRF design quicker and easier, but which do not form part of the formal ECRIN requirements, are listed at the end of the section.

DM01.01 CDMA development policies

Controlled documents covering the development of CDMAs and CRFs should be in place

Developing CDMAs and the CRFs within them must be done using defined procedures, with tasks and responsibilities clearly delineated for design, development, testing and deployment. Controlled documents should therefore exist covering these areas.

The evidence that the standard has been met would be the relevant documents.

DM01.02 Requirement specifications of CRF

Processes exist to ensure the CRF specification fully supports the outcome measures and safety requirements in the protocol but does not ask for unnecessary data.

A fundamental requirement is that the centre works with the sponsor to ensure a clear link between the protocol and the set of CRFs, with the CRFs capturing sufficient data for the outcome measures described in the protocol, but avoiding redundant questions and data that 'might possibly be useful one day'.

The centre should be able to describe and demonstrate how the CRFs are developed and / or checked to ensure they match the protocol in this way. One approach is to first use the protocol, and in particular listed outcome measures, to specify the data points that the statisticians will need to carry out the required analyses, and then use these *analysis data requirements,* in addition to the relevant safety parameters, to drive the CRF specification. At the moment many centres start CRF design by re-using sections of existing CRFs, which is understandable but does not always mean the CRFs exactly match the new protocol's requirements. Identifying analysis requirements first, before looking at pre-existing CRFs, may be more efficient and result in significantly streamlined data sets.

The more this aspect of CRF design is made explicit the easier it will be to demonstrate the standard. A review by the study statistician (and whoever else is

involved in the analysis) is therefore key, and ideally their sign off would be accompanied by an explicit statement that the specification meets the data requirements but does not include unnecessary data collection.

Note that ECRIN auditors are not expected to assess the *outcomes* of this process, i.e. to assess this CRFs against the protocol in the context of specific trials (partly because in any particular case the sponsor or investigator will usually have the final say about CRF design). The requirement is that the centre should be able to describe and demonstrate the *processes* of CRF construction and review, and how that is linked to the requirements of the protocols.

DM01.03 Functional specifications of CRFs

CRF design and functional specifications exist identifying each data item on each CRF (including field names, types, units, validation logic, conditional skipping)

A key aspect of developing an eCRF is creating a full functional specification, characterising all the data items and associated validation (i.e. data checking), skipping and derivation logic.

The specification may contain an 'annotated CRF' (though on its own this is unlikely to contain all of the required information and will usually need to be supplemented by other documents) or it may exist as an entirely separate set of documents, for instance as a set of spreadsheets or as a database report.

CDMA programmers can use the specifications to accurately build the eCRFs. This is not necessarily a single 'specify then create' process - often an iterative approach will be used - but the CRF should still be clearly based upon a specification. Without such a specification it becomes very difficult to properly validate, and document the validation, of the final eCRF.

Evidence that the standard has been met would come from:

- inspection of CRF specifications;
- discussion with staff to clarify how the specifications are developed;
- relevant controlled documents.

DM01.04 Cross-disciplinary approval of the functional specification

CRF design and functional specifications are signed off and dated by signatories representing a cross-disciplinary team

Once CRF / CDMA specifications have been constructed they will need to be formally approved and signed off by the key individuals involved with the trial.

The final specification may not be signed off until at the end of the CRF construction process (i.e. it does not always have to be fully approved before the CRF is begun) but the CRF should still be clearly approved against a specification, in order to provide a clear baseline for validating the system.

Because developing CDMAs and the CRFs within them should involve the various users of the system, or key representatives of those users, the final sign off should represent a cross-disciplinary team.

As a *minimum*, the expectation is that a representative of those collecting the data (i.e. the trial's data management staff) and those analysing the data (i.e. the trial statistician) sign off the functional specification.

Others that are usefully included in the sign off are those setting the CDMA up (i.e. the IT staff), the chief investigator and / or sponsor representative, and possibly a quality assurance or operational manager.

Evidence that the standard has been met would come from:

- inspection of the relevant controlled documents;
- discussion with staff to clarify how the CDMAs were developed;
- the range of names and signatures involved in signing off CRF specifications.

DM01.05 Isolation of development CDMAs

CDMAs in development should be isolated from CDMAs used productively

A CDMA should be developed within an environment reserved for development and test activity only. The development and production systems should be isolated from each other - there should be no possibility of any problems in a developing CDMA spilling over to affect any production system, or of users inadvertently confusing the development and production instances.

This could be done by having distinct data stores (e.g. different databases or even database servers) for the development and production environments, or possibly two distinct instances of the CDMS.

The evidence that the standard had been met would come from:

- explanation and demonstration by centre staff of how the CDMAs in development were kept isolated from production systems;
- inspection of relevant controlled documents.

DM01.06 Isolation of training eCRFs

Access to the CDMA for training purposes is managed to ensure that is isolated from clinical data

Users need to be trained on CDMAs, generally using dummy or test data, and it is important that this data is kept separate from actual study data.

User access for training purposes must therefore be managed to ensure that this is the case, sometimes by using a completely different CDMS instance and / or data store for training than for production, sometimes by setting up dummy

'training sites' within the production system (the data from which is excluded from analysis).

The evidence that the standard had been met would come from:

- explanation and demonstration by centre staff of how the data generated in training was kept separate from actual study data;
- inspection of relevant controlled documents.

DM01.07 Production of interim CRF

For trials / sites using eCRFs, procedures should be in place to generate accurate iCRFs (interim CRFs) for sites, if and when necessary

A centre should be able to generate so called interim CRFs or iCRFs, if required and if the sponsor agrees this would be appropriate.

These may be needed in eRDC systems if direct data entry into the system is not possible or desired during initial data collection. Anecdotal evidence suggests that this is a common situation, especially as many site staff find it difficult, and rather unsympathetic, to interview subjects and use an eRDC system at the same time.

In such circumstances the research staff at the site are far safer using structured paper documents that match the eCRF to note down responses and other data, rather than blank sheets of paper or whatever else might be available. The system should therefore be able to produce such iCRFs, ideally directly at the site ('system' being all available systems and processes, including but not limited to the CDMS).

In some cases the iCRFs can be as simple as screen shots of the eCRF screens, though they should include a mechanism for noting the subject's name, number or similar unique identifier. The important thing is that they allow data collection to be structured in the same way as if the eCRF was directly available, and safely stored before it is transferred to the eRDC system.

The evidence that the standard had been met would come from:

- explanation and demonstration by centre staff of how interim CRFs could be created;
- inspection of relevant controlled documents, detailing the procedures to be followed.

Further Indicators of Good Practice in CDMA Design and Development

Listed below are several examples of 'best practice' in CDMA development and CRF. They do not form part of the ECRIN requirement but their usage provides greater confidence that procedures for CRF creation are well developed and applied consistently.

- *Using libraries and metadata repositories:* Having libraries available of items and forms, or a more formal metadata repository, enables reuse of data items and a consistent approach to coding and naming, especially if backed up by local guidance documents. Such libraries can also promote the consistent use of repeating question groups (or alternatively lists of single questions) within particular domains.
- *Consistent local coding systems:* Common principles applied to item design and metadata (e.g. preferred coding systems, even for 'yes' and 'no', styling and numbering of items, the coding of different types of missing data, preference for positive formulated questions, etc.) can all make systems more consistent and easier to use.
- *Using standard coding systems (e.g. CDISC CDASH):* In some domains international standards are available for data item codes and definitions, especially those defined by CDISC within the Clinical Data Acquisition Standards Harmonisation (or CDASH) project.
- *Using standardized questionnaires and instruments:* Using validated questions, scales or standard instruments (e.g. for quality of life questionnaires) improves the reliability of the final results and, if already available in a library, speed development. Decisions about the use of such validated instruments are ultimately the sponsor's, but a data centre should have them available and be able to promote their use.
- *Local design and guidance documents:* Local documents specifying good design practice and preferred orientation, colours, fonts, graphics, positioning etc. (so far as the CDMS allows variation in these) can promote consistency and a 'house style'. Consistent and sensible use of dividers and sectioning, and white space, can also add to consistency and the ease of use of systems.

DM02 CDMAs - Validation

Once a CDMA has been constructed it must be validated to ensure that it works as intended and is fit for purpose. The validation required will follow the general principles described in IT06, but there are CDMA specific aspects of validation that are commonly applied, and which are collected together in this section.

Validation in this context does not refer to *data* validation, i.e. the process of checking that data contains reasonable values and is logically consistent. That process is better referred to as (logical) data checking. It does apply, however, to the process of verifying that the data checks operate correctly, and indeed this 'checking the checks' usually forms a significant part of CDMA validation.

Because CDMAs are specified and built in-house, they can normally be amended relatively easily until they meet their original specification, though occasionally the validation process itself may trigger last minute changes to the specification. The validation process would normally be shared by IT and data management staff, as well as end users (see DM02.04), to ensure that the system and its constituent eCRFs were fit for purpose.

DM02.01 CDMA validation policies

Controlled documents for CDMA validation are in place

There should be a general procedure for CDMA validation, specified in controlled documents, detailing procedures, responsibilities, outcomes etc., even though each individual CDMA will need its own specific validation documentation.

Evidence that the standard had been met would come from the controlled documents themselves.

DM02.02 CDMA Specific test plan

A trial-specific test plan and a test documentation set exists for each CDMA.

Each CDMA will require its own set of specific validation documentation. These will usually be based on the general procedures but list the specific study parameters (e.g. uniqueness checks), and eCRF logic checks, skips, and derivations in test documents. Such documents can then be completed with the result of the tests recorded for each individual element.

Some CDMAs may require additional testing, for instance to check access from particular sites, or particular functionality (like coding, or message triggering) that is not found in other study applications. The testing of these should form part of the validation plan. In other words validation should, as usual, be based on a risk assessment and identification of the elements that need to be tested.

Evidence that this standard had been met would come from examination of the CDMA specific validation documentation for a range of studies.

One approach to CDMA validation involves completing test pCRFs (or iCRFs), inputting them into the CDMA, and then exporting them again in a form that is readily comparable with the original data.

This has the advantage of testing overall usability as well as many of the functional components of the system, and more importantly it also means that the extraction / reporting functions are tested as well - something that may be more important if locally built routines are used for part or all of the extraction.

The main disadvantage of this approach is that - unless enormous care is taken in preparing a large set of test data - not all functional components of the system will be systematically tested. If used, the method should therefore probably be seen as an addition to the detailed testing of each component above.

Evidence that the standard had been met would come from examples of trial specific test documentation, along with discussion with staff to clarify how it was used in practice.

DM02.03 CDMA testing against functional specifications

Testing with sample data against functional specifications is carried out for each CDMA before deployment to live environment

One of the key aspects of CDMA validation is the detailed testing of each CRF against its functional specification.

This will include checking the correct data items are there, of the right type, with the specified codes and code lists, and in the right order. Most of the testing effort, however, will be centred on the logic built into each CRF - the range and consistency checks, the skipping (or enabling / disabling) logic, and the generation of any derived values. Each of these checks should be separately documented, with - for the logic checks - input values and the system's response.

Ideally, the system should be able to generate some of the necessary test documentation itself, for instance it should be able to generate a listing of all the logic checks on a particular CRF. These might then need further processing to create a proforma (or a database) for recording the test results.

Alternatively, some data centres use their system for generating and recording the CRF functional specifications to also produce the test documentation, and to record the results of those tests - something that is relatively easy to do with specifications stored in databases rather than spreadsheets.

The evidence that standard has been met will come from the detailed test records for a range of CDMAs.

DM02.04 Assessment of CRFs by users

Procedures are implemented to secure feedback from selected end-users, on the practicality and ease of use of specific CRFs.

This standard considers the CRFs as a whole and how easy they are to understand and work with in terms of raising or responding to queries. To be valid, such checking needs to involve a sample of actual users (ideally with different levels of experience).

Some aspects of a system's ease of use will be fixed by the design of the underlying CDMS (for example the method of navigating between screens and patients) but within those constraints there is usually some variation, within a specific CDMA design, that allows different question ordering, skipping, coding and captions, as well as on screen prompts. It is these that need to be checked with users, to ensure that they can be interpreted and used correctly.

The standard does not require every CRF in every study to be tested by users - it would be up to the centre to assess the novelty and / or complexity of CRFs, as well as the prior experience of site staff, before deciding which users needed to be involved and for which CRFs.

The centre may want to retest CRF designs, however, if the user group itself changes or includes new types of staff. For example, if users from a new country or language group are involved then the CRFs should probably be tested against a sample of those users.

Note that user feedback could also be integrated with initial user training.

Evidence that the standard had been met would come from explanation by centre staff of how user feedback is organised and gathered, plus inspection of documents including user feedback and / or sign off, against specific CDMAs.

DM02.05 CDMA approval

Each CDMA should be formally approved, dated and signed by the relevant signatories, before production use.

Once CDMA validation has been completed it needs to be signed off, normally by a small cross-disciplinary team but as a minimum by the trial or project manager who will oversee the use of the CDMA in supporting the study. In most cases a single sign off will cover the whole CDMA, but some centres may have each CRF signed off separately.

Evidence that the standard has been met will be appropriate dated signatures confirming that the CDMA is OK to be used as a production system.

DM02.06 Validation detailed findings

All validation results, including any test data and protocols, are retained for each CDMA

All the detailed test documentation / systems, as well as the results, and any scripts, dummy data, listings etc., used for any particular validation should be retained. Much of this may be in electronic form rather than on paper.

The evidence that the standard had been met would be:

- an explanation and demonstration by centre staff of how and where the detailed test results were retained;
- inspection of actual results against a range of CDMAs.

DM03 CDMAs - Change management

Even after a CDMA has been successfully validated and moved into production changes will be requested. Such changes must be carefully managed to ensure that the system retains its validation status.

The change management required follows the general principles outlined in IT06 (standards IT06.06 and IT06.07 in particular) but CDMA change is relatively common, and its proper management critical to data management, so a separate section of standards is justified.

DM03.01 Change management of CDMA

Controlled documents for CDMA change management are in place

Controlled documents should be in place dealing with CDMA change management, detailing procedures, roles and responsibilities and documentation.

Evidence that the standard has been met will be the controlled documents themselves.

DM03.02 Documenting change requests

Individual requests for change to CDMAs are justified, itemised and documented

The initial step in the change management process is to ensure that any requests for change to the CDMA are properly described and authorised. This would normally involve a paper or screen based proforma being completed with the necessary specification of and justification for the request.

Evidence that the statement had been met would be from inspection of such proformas.

DM03.03 Change and risk analysis

A risk analysis is conducted and recorded when considering any change

The change management process must include an assessment of the *potential impacts and risks* associated with a proposed change. For relatively trivial changes (addition of additional categories to a code list for instance) these impacts may be small; for large changes - e.g. the addition of a new eCRF - they may be considerable.

Changes that would risk orphaning data already in the system (e.g. dropping questions or categories) or making existing data invalid (e.g. changing the type of a question) should not normally be allowed and the change request should be rejected.

Any change will impact the CDMA itself, but there may also be impacts 'downstream', for instance on the data extraction process or the scripts used

during statistical analysis, or on system documentation and / or user training. A CDMA change may also imply a change to the protocol (see DM03.06).

It is important that all these aspects are taken into account. Some centres use a 'change checklist' approach to structure the assessment of risk and to help with its documentation.

Evidence that the standard had been met would be the inspection of the risk assessment documentation against a range of proposed CDMA changes.

DM03.04 Testing of CDMA changes

Any change is tested in the development / test environment and the test results are recorded

The risk analysis (see DM03.03) will determine the amount and type of re-validation required. This should always take place in the development / test environment and the results recorded.

In a busy data centre it is sometimes tempting to make and inspect trivial changes in the production environment, but then the flow of versions between the two environments is disturbed, and the next import of a study definition from the test environment will overwrite the earlier change.

All changes should therefore be implemented in the development environment first, and the revised system then exported to the production environment. This also makes it easier to store each version of the study definition metadata file for future reference.

Evidence that the standard had been met would come from inspection of the detailed test results relating to changes.

DM03.05 Versioning of CRFs

CRF development and change management should include clear versioning of all relevant documents, including the (e / p) CRFs themselves.

As part of the development, deployment and change management processes different versions of CRFs and associated documents will exist and need to be carefully and clearly managed. The management should include clear records of when new versions were signed off and introduced into the system (possibly on a site by site basis), as well as clear indications of the different versions on all documents.

Evidence that the standard has been met would come from inspection of the CRFs and relevant specification documents, and a discussion of version management in the centre.

DM03.06 Communicating changes

Mechanisms are in place to inform relevant staff and users of changes, and provide support and explanatory material as required

The potential impact of any change on users should also be considered. In most cases data entry staff will need to be informed of changes and why they have been introduced, and so mechanisms should be in place to allow this to happen consistently.

For substantial changes there may also be a need to provide additional training, and the communication should reflect that.

Evidence that the standard had been met would come from explanation by centre staff of how the system worked, from the relevant parts of controlled documents and from examples of the mechanism in action.

DM03.07 Changes and protocol revision

Mechanisms should exist to ensure any requested CDMA change that implies a protocol amendment is identified.

An amendment to the study protocol can often generate changes in the study's CDMA. That is normally a straightforward process, because it is the direction in which change would be expected to flow.

From time to time, however, a requested CDMA change may represent or imply a change to the protocol, even though it may not have been presented or recognised as such. The centre should have some mechanism in place to ensure that any change that implied a protocol amendment (that had not already been proposed) would be identified. The amendment would then need to be managed before the CDMA itself was changed. For instance, any necessary re-approvals would need to be obtained before the CDMA change was implemented in the production system.

It is recognised that for many centres this type of change request would be very rare, but there is no harm in including a checking mechanism within the process of reviewing and approving requested changes, and recording the decision made (for instance as part of a 'change checklist').

[It might also be useful to record, as part of the change management process, the more normal situation where a requested CDMA *follows* a protocol amendment, and if so which one].

Evidence that this standard had been met would come largely from inspection of the relevant controlled documents and associated proformas, together with discussion of any examples of the mechanism being used in practice.

DM04 Data Entry and Processing

The standards in this section deal with data entry into the CDMA. Most modern CDMSs make this very straightforward but, as one of the core processes of data management, it still requires a framework of policies and procedures if it is to be carried out consistently to agreed standards.

DM04.01 Data entry policies

Controlled documents for data entry and corrections are in place

Some of these documents may be generic (e.g. general policies on using self-evident corrections) but others may be trial specific and usually found within the Data Management Plan for the trial (e.g. the specific self-evident corrections that have been agreed as acceptable)

Evidence that the standard had been met would be the controlled documents themselves.

DM04.02 Access control for data entry and review

Access control is fully implemented; data entry / review is only accessible to authorised personnel and according to need

Data entry must take place in the context of controlled access, i.e. adhering to the centre's own policies on access control.

This is a special case of the access control already required under IT 04.02. It is included here partly to provide an additional emphasis on access control within the CDMS, partly because access control for data entry is often a joint responsibility of IT and data management staff, and partly because it is often the subject of specific policies and controlled documents.

An important aspect of being able to access data only 'according to need' is that remote site staff only have access to the data (and related material like queries) of their own site.

Control of access should also include access to reports, data extraction and other review mechanisms, i.e. users should only see the data that they have a right to see and be able to run the reports that are relevant to their role within the system.

The evidence that the standard had been met would come from

- the controlled documents dealing with CDMS access control for centre and site staff;
- demonstration of the access control system.

DM04.03 Management of missing data (eRDC)

Mechanisms are in place to identify and report on missing or late eCRF data

(This standard only applies to centres running eRDC trials.)

Monitoring what data has arrived is part of the data entry process, so that sites can be contacted to request missing or late data. Some eRDC systems make this straightforward, with the system set up to identify missing data and the centre able to send messages to sites to query that data. Others focus on data collection rather than the workflow, so data may need to be exported and processed, perhaps using statistical scripts, before missing or late data can be identified.

The exact mechanism is therefore likely to depend on the sophistication of the eRDC system(s). A useful feature of scheduling systems within eRDC system is the ability to suppress missing data messages when notification is received that the subject has died or is lost to follow up. This avoids irritating sites by requesting data that will never exist.

Evidence that the standard had been met would come from:

- the relevant controlled documents;
- demonstration of the missing / late data management system(s) and explanation of their use in practice.

DM04.04 Management of missing data (paper CRFs)

Mechanisms are in place to identify and report on missing or late paper CRFs

(This standard only applies to centres running paper based trials.)

With trials using paper CRFs there is often a lag (from several days to several weeks) between CRF receipt and the addition of the data to the CDMS, so that the CDMS cannot be used reliably to monitor receipt of data. It is therefore necessary to have a separate CRF tracking system in place, unless the lag time can be guaranteed to be limited to a few days.

A useful feature of CRF tracking systems is the ability to automatically truncate a subject's schedule when notification is received that the subject has died or is lost to follow up, or at least allow easy manual amendment. This avoids irritating sites by requesting data that will never exist. This is not currently part of the standard but is regarded as best practice.

The evidence that the standard had been met would come from:

- the relevant controlled documents;
- demonstration of the pCRF tracking system and its outputs.

DM04.05 Handling patient identifying information

Inappropriate patient identifying information submitted to the centre is obscured or removed

One of the problems that can occur in data entry is patient identifying information being inappropriately added to, or retained on, submitted data. For instance, with paper CRFs, site personnel may add the patient's name or initials to a safety report, or annotate a CRF or image file with local identifiers. With eRDC, sometimes names are entered in error into comments, notes, and query responses, etc.

In some cases this may contravene national regulations, in others the policy of the centre and / or sponsor. In either case the identifiers should be removed or (more normally) blocked out on paper CRFs and the site reminded of the requirement to omit such identifiers. Many centres simply use black marker pen to cover the identifiers and make them illegible, annotating the action on the CRF. For eCRFs query mechanisms can be used to ask the site to remove the identifying data.

In either case the centre should be able to demonstrate general and / or study specific policies describing the appropriate actions to take, and their application in use.

The evidence that the standard had been met would come from:

- relevant controlled documents;
- discussion with staff and demonstration of the blinding being put into action.

DM04.06 Audit trail

All transactions in the CDMA (insert, update, delete) must have an audit trail, covering the date and time of the input, the person making the change and the old and new values

Providing an audit trail of the CDMS transactions is a regulatory requirement. For instance the FDA (CFR 21 (11), section 11.10(e), 2010) requires the

"Use of secure, computer-generated, time-stamped audit trails to independently record the date and time of operator entries and actions that create, modify, or delete electronic records. Record changes shall not obscure previously recorded information."

Modern CDMSs normally support such an audit trail.

The audit trail requirements do not include a 'reason for change' (RFC) as a mandatory data item, though many CDMSs support this as well. Some data centres like to make use of this feature, others are less convinced of the utility and accuracy of the data recorded.

Evidence that the statement had been met would come from demonstration of the audit trial being created in a test database.

DM04.07 Timestamp control

Sites using eRDC should not be able to change the CDMS's time stamp

Because an accurate time stamp is an integral part of the audit trail, it is important that there is no ambiguity about the time recorded against data activity. In particular it should only be possible to set this time centrally, i.e. at the data centre, and not at the remote sites.

Most CDMSs support this feature automatically, and also record both the local time at the data centre and the time at the remote site inputting data, usually as the data centre time +/- *n* hours according to the site's time zone.

Evidence that the standard had been met would normally come from the CDMS documentation and demonstration of the use of local / site times within the data.

DM05 Managing Data Quality

A data centre should be able to run checks on the accuracy and consistency of the data it contains, during and after data entry. There should also be mechanisms, involving raising queries with the clinical sites, to resolve the discrepancies, or potential discrepancies, that are found.

The standards in this section cover this area, but they are concerned only with the data quality activity that take place at the data centre - they exclude those that take place at sites, and specifically *they exclude monitoring and source document verification* (SDV) even though these are important mechanisms for checking data quality. They do include, however, support that the centre might provide for SDV and monitoring.

Data checking may take place during data entry into a CDMS, using pre-configured consistency and range checks, after data entry but still using tools within the CDMS, after data entry but using manual checking of source paper records and database values, or double data entry, or after data export and subsequent analysis, usually by scripts written in statistical software. The standards cover all these types of checking mechanisms though of course only some of them would be used by any particular centre.

Query management is usually integrated into modern CDMSs, with queries raised, annotated, responses reviewed and the queries closed all on screen, the CDMS acting as the transport medium between centre and sites. For paper based trials queries must be raised and tracked separately, in some centres using IT systems developed for the purpose, in others more basic tools like spreadsheets. The standards in this section apply to both types of query management.

DM05.01 Data quality policies

Controlled documents are in place describing the various ways in which the centre maximises data quality.

These documents will cover (for instance) data checking mechanisms, both within CDMAs and outside them, query generation, tracking and resolution, and the support of site monitoring (but not the monitoring process itself). For centres managing paper based trials there should also be policies about quality control of the transcription process from paper CRF to the CDMS.

It is recognised that in any particular case the details of the data checking regime might be modified by the sponsor and / or trial management team (and be described in the study specific data management plan) but there should be default policies and procedures in place.

The evidence that the standard had been met would be the controlled documents themselves.

DM05.02 Data checks during data entry

It should be possible to include data checking mechanisms within the data entry process

As a minimum it should be possible to apply range checks on numeric and date data items. These may be either 'soft', i.e. they generate a warning (e.g. 'the weight value seems unusually high'), or 'hard', i.e. they reject the data value entirely (e.g. 'but that date is in the future'), or some combination of both. The use of 'hard' checks in a paper based studies is unwise, because it may stop a received value being input into a database, but they can be useful within eRDC systems.

Data entry checks would also normally include other conditions that could be easily set up on a single data item, such as set membership (e.g. 'value is one of 1,2,3,4 or 5') or matching a regular expression ('this does not appear to be a valid email').

Of course many CDMS systems allow much more complex checks than these to be set up. Many allow data items on different forms and visits to be compared for consistency and also allow complex expressions to be evaluated. There is, however, a debate about the time and effort it takes to set up and test complex checks in many CDMS, compared (for instance) to doing them within scripts in a statistical package.

The level and complexity of checks used will vary from study to study, and will also tend to vary inversely with the number and complexity of checks carried out *post* data entry. Data centres exhibit wide variations in the emphasis they put on checking data during and after data entry, but they should have mechanisms available to do both, and be able to demonstrate both in action.

The evidence that the standard has been met would be:

- demonstration of simple checks on a variety of eCRFs;
- discussion with centre staff explaining their use of data entry checks.

DM05.03 Data checks post data entry

Pre-programmed data checking procedures are available to be used post data entry

This can involve a variety of mechanisms. The most flexible method is to periodically export the data so that scripts can be run against it, usually using a statistical package such as SAS, R or Stata, to identify outliers, inconsistent values, missing values etc.

Many CDMS also allow pre-planned validation checks to be run against datasets, often referred to as 'batch validation'. This may happen periodically, but it is particularly useful if a new data entry check is added to a system, and needs applying to the data that has already been metered.

The more traditional method is to export selected data points into a simplified format, often in a spreadsheet, to form 'line listings'. These can then be visually inspected for inconsistent or extreme values. Unfortunately, used in isolation, such a method is unreliable, but it is sometimes used to supplement the other methods described above. There is nothing wrong with line listings *per se*, but they should be checked by some form of automated process rather than manually.

The standard requires that the centre can carry out data review, post data entry, using some form of pre-programmed process (i.e. over and above visual inspection of data).

The level and complexity of checks will vary from study to study, and will also tend to vary inversely with the number and complexity of checks carried out during data entry. Data centres exhibit wide variations in the emphasis they put on checking data during and after data entry, but they should have mechanisms available to do both, and be able to demonstrate both in action.

The evidence that the standard has been met would be:

- demonstration of checking procedures and / or scripts, and documentation of their use;

- discussion with centre staff explaining their use of post - data entry checking.

DM05.04 Query creation

Queries can be created - automatically and / or manually - based on any of the data checking mechanisms employed.

There are two main mechanisms for creating queries:

- during data entry, as a function of the omissions and discrepancies noted by data entry staff, usually prompted by the validation messages generated by the check logic in the CDMA;

- after data entry, as a result of checking data, e.g. by batch validation or statistical methods, using values flagged in some way by the checking process.

In either case there should be clear procedures in place that guide when and how the queries are generated. Though not a requirement of the standard, ideally the centre would be able to always send queries to the clinical sites in the same way, whatever the query generation mechanism.

For instance, most CDMS include a mechanism for on-screen query generation, triggered by data entry checks. It should be possible to manually add new queries, as identified by checks run on exported data sets, into the same system. The sites then only see the queries as presented by the CDMS. Conversely a data centre running paper based trials, where the queries also have to be delivered to the sites on paper, by post or courier, should be able to send the same query

proforma for queries generated by the CDMS (used in-house for data entry) as for queries generated by statistical checking of datasets.

The evidence the standards had been met would come from an examination of queries generated and a discussion with staff about how the relevant procedures worked in practice.

DM05.05 Tracking of queries

Responses are recorded when returned, identified when outstanding and queries resent if necessary

Having sent the queries out, through an eRDC system or by post or courier, the centre needs to be able to track the responses to them and identify those for which no response has been received, or for which the response is unclear, resending the query or generating a new one if necessary.

If queries are sent out through the eRDC system, that system will normally have such tracking functionality built in. For trials using pCRFs a separate query tracking mechanism will be necessary. For best practice it would be linked to the query generation process and include functionality to prevent duplicate queries being sent out to sites, though this is not a formal requirement.

Evidence that the standard has been met would be demonstration of the query tracking system(s) that showed how queries were recorded and tracked.

DM05.06 Actions in response to queries

Query resolution is tracked, and appropriate actions taken and documented.

Once a query response has been received a decision is made as to whether it is fully answered or not, and a supplementary query sent if necessary. If the issue has been resolved values in the CDMA may need to be changed.

For most eRDC systems with integrated query management the link between the query, its response and the value in the database, whether or not it has been changed, will be obvious and visible on screen. For pCRF based trials with separate query management, many centres use a comment or 'reason for change' field to link the data value to the query or queries associated with it (for instance storing a query ID number).

Either way the record of the query and its resolution should be linked to the data item, either in the CDMS or in a separate query management system, effectively making the query part of the audit trail.

The standard would be met if this is shown to be the case.

DM05.07 Self-evident corrections

Clear guidelines and procedures should exist to identify and carry out self-evident corrections

In some cases the data on a paper CRF is obviously incorrect and would fire a warning or reject message if input, but it is clear what the correct data should be - the error has been caused by a common omission, addition or transposition. An example would be '07/11/209' or '07/11/20009' for '07/11/2009', or the omission of a response to the 'Any Adverse Events?' question followed by a report of three adverse events.

In such cases it does not make sense to query the site, and a self-evident correction (or an 'automatic obvious data modification') can be used to amend the data. The use of such self-evident corrections (SECs) must be tightly controlled however.

- They should be restricted to a pre-agreed list of situations where they could be applied, normally agreed at the level of the individual study (often starting with a default list maintained by the data centre).

- There should be a clear procedure to follow when self-evident corrections are applied, including instructions on how the source document should be marked to indicate that the correction had been made.

It can be useful to send each site's final list of SECs back to the site at the end of the study, so that they are aware of changes made and can check them. This process should not be used to *authorise* the changes, however, because retrospective checking is difficult and there is no guarantee that it will be carried out thoroughly. A pre-defined and pre-agreed set of pre-conditions for SECs, as described in (a), should be used instead.

Self-evident corrections *could* be applied to eRDC systems as well. But data entry checks should pick up the sort of obvious error that would call for a SEC and, even if something looked like it needed a self-evident correction, it could simply be sent back to the site as a query. SECs make sense for paper based studies because queries are relatively expensive, but they are usually much quicker and cheaper to resolve in an eRDC study.

The evidence that the standard had been met would include:

- the relevant controlled documents (e.g. examples of data management plans with self-evident correction instructions in them);

- discussion with and demonstration by the centre staff of the procedure in action.

DM05.08 Quality checks of data transcription

There should be policies and procedures in place to provide a quality check (QC) on the transcription process from paper CRFs to the database system.

(This standard only applies to centres running paper based trials.)

Various approaches can be used. Some centres use double data entry of some form, for some or all of the data entered from paper sources. Others check

accuracy retrospectively, for example selecting a sample (e.g. 10% of data, or particular visits / forms) and compare the database values with those on the original CRFs - a type of 'internal SDV'. If the error rate exceeds a particular threshold - say 5% - the check is then usually extended to a larger size sample.

The standard requires that the centre has mechanisms in place to carry out this QC of transcription in paper based trials. They might vary from one study to another, and therefore might be described in study specific data management plans rather than generic controlled documents such as SOPs.

The evidence for the standard would be the descriptions of QC mechanisms used by the centre, both in documents and as obtained from discussions with staff.

DM05.09 Quality check documentation

There should be detailed results available from the QC of data transcription

(This standard only applies to centres running paper based trials.)

The checks carried out of transcription accuracy, of paper CRFs, need to be documented. This includes the results, i.e. discrepancies found and decisions taken, of any double data entry.

The expectation is that at least a summary report would be available as part of the trial's documentation. The detailed data would often be available in electronic form and / or as a report from a system, but it should still be available on demand.

The evidence for the standard would simply be summary and detailed QC results.

DM05.10 Supporting source data verification

The centre has procedures for supporting source data verification, as a minimum providing access to its data for those implementing and conducting the SDV

The sponsor will normally determine both the SDV strategy required and decide who will be doing the SDV. Pharma sponsors may, for instance, want to use their own monitors for SDV. Even non-commercial sponsors may wish to use a different trials unit for the monitoring / SDV function than for the data centre function.

What a data centre does need to do is *support* the work of monitors carrying out SDV, by making the trial data available to them. There should therefore be procedures in place for allowing monitors access to the data so that they can inspect and assess it, and for exporting and presenting data on demand, on a subject by subject basis, to monitors.

It would be good practice, though not currently a formal requirement, to further support SDV with reports detailing query rates and late data (or any other indicators of problems during data entry) on a site by site basis.

The evidence that the standard had been met would be the controlled documents describing the relevant procedures, together with explanations from staff about how they worked in practice.

DM05.11 Supporting central statistical monitoring

The centre can generate reports to support central statistical monitoring

One of the key requirements of risk based monitoring is central statistical monitoring of data, specifically to identify clinical sites that have relatively high query rates for their data, and / or who are consistently late with data. The monitoring can also be used to identify particular data forms and even items that appear to give rise to problems in data collection, possibly prompting a redesign of the CRF.

The statistical monitoring may be carried out by using statistical packages and scripts against exported data, or it may come from reports built into trial administration systems if they handle data tracking and queries, or in some cases it may even come from reports in the CDMS itself.

The evidence for the standard being met would come from demonstration of the relevant reports and a discussion of how they were used in practice.

DM05.12 Removing fraudulent data

Data deemed invalid (e.g. produced fraudulently) can be safely removed from the analysis data set.

Though rare, it sometimes happens that a site is shown (or is strongly suspected) to have produced data fraudulently, or is otherwise guilty of misconduct. In these situations, the sponsor may decide to disregard all the data from that site.

The expectation is that the centre could describe how (in a technical sense) the data could be safely removed, at least from the data being analysed - it would normally stay in the source data - and how (in an administrative sense) it would document the removal process.

Given the rarity of the event it is not expected that a centre necessarily has a controlled document in place describing these procedures, but it should be able to provide an explanation of how such a situation would be handled.

DM06 Delivery and Coding of Data for Analysis

The standards in this section deal with the ways in which trial data is prepared, checked, fixed in some way, and then extracted in the format required for analysis.

The specific processes used for generating analysis datasets will vary, depending on the longevity and type of trial as well as the purpose of the analysis. For example, for a self-contained study where there will be no further data collection, the database is often locked down (or 'frozen', though the exact definition of 'locked' and 'frozen' varies between systems) so that no further data entry or amendment is possible. For a longer term study where data collection may continue for many years after the primary analysis, or where various interim analyses are necessary, it would be more usual to export a 'snapshot' of the data state.

Note that there is no requirement relating to the format of the extracted data. That will normally be as agreed with the statisticians that carry out the analyses - examples include CSV, XML, and SAS, R and SPSS native formats.

DM06.01 Policies for database locking

Controlled documents should be in place dealing with taking a snapshot of the trial data, and / or 'locking' and 'unlocking' that data

All processes by which data is prepared and extracted for analysis should be governed by clear procedures, documented within controlled documents.

The relevant evidence would be the controlled documents themselves.

DM06.02 Data completion

All relevant data (or all except for a pre-defined / pre-agreed fraction) should be received prior to data extraction for analysis

Extracted data need to be as complete as possible. In some cases database lock is dependent upon completion of data entry, in others a snapshot is taken once all data expected by a certain point is in, or at least - e.g. for an interim analysis - all that can be reasonably expected in a given trial at a given time.

The evidence that this standard was being met would be:

- the relevant controlled documents;
- examples of communication and / or a checklist relating to database lock / snapshot and the levels of data required.

DM06.03 Query resolution completion

All queries (or all except for a pre-defined / pre-agreed fraction) have been resolved prior to data extraction for analysis

Queries will also need to be resolved before database lock or snapshot. In some cases this will mean all queries, while in others some exceptions may be allowed. The rules governing any exceptions should be explicitly defined and agreed.

Data consistency checks will also often generate additional queries during the final phase of preparation for analysis, leading to an upsurge in query generation with, very often, faster timelines for their resolution (see DM07.05).

The evidence that this standard was being met would be:

- the relevant controlled documents;
- examples of communication and / or a checklist relating to database lock / snapshot and the query resolution required.

DM06.04 Data reconciliation

All external data (e.g. safety database, lab data) has been reconciled prior to data extraction for analysis (or all except for a pre-defined / pre-agreed fraction)

Data preparation may also involve reconciliation of the data input through the CDMA with that received from elsewhere - for example between expedited SAE reports and the more routine adverse event reporting, or between sample and laboratory result data. This should be brought up to date before the database is locked or a snapshot is taken. If exceptions to data reconciliation are allowed, they should be defined, agreed and documented.

Where data coding has been used (see DM07.08, DM07.09) it would be normal for that coding to be reviewed as part of the data preparation. In some instances a data quality check may also be done, especially if one has not yet been performed on this data. Whatever the detailed arrangements specified by the relevant controlled documents, a check list dealing with the different aspects of data preparation can be a convenient way of ensuring all the aspects are covered and recorded.

The evidence that this standard was being met would be:

- the relevant controlled documents;
- examples of communication and / or a checklist relating to database lock / snapshot and the need for data reconciliation.

DM06.05 Post lock data amendment

Controlled documents should be in place detailing procedures to be followed if data needs to be altered after the snapshot or DB lock

Despite the best planning and preparations, there may be occasions when amendments are required to the data after the database has been locked, or to snapshots after the extraction has actually taken place - perhaps to correct errors that come to light at the last moment, or to incorporate late returned query data. In such cases it is essential that the unlocking / amendment process is tightly controlled and documented in any given instance, as demanded by this standard.

The evidence that this standard was being met would be:

- the relevant controlled documents;
- documented examples of post lock data amendment.

DM06.06 Read only retention of analysis data

The data provided for analysis is retained within a read only regime, and is available as a reference data set for any future re-analysis or audit

There will be a need to arrange the long term retention of any extracted data, partly for audit or inspection purposes and partly to allow - if necessary - the reconstruction of any analysis using the same extracted data. This would normally be done by placing the relevant files within an area of the centre's storage capacity that is read only (except for the IT staff that do the transfer).

The evidence that this standard was being met would be:

- the relevant controlled documents;
- demonstration of read only retention for a range of extracted data sets.

DM06.07 Extracted data validation

The data generated for analysis should be validated against the data in the clinical database, or the extraction process itself is validated

The processes used to extract and, if necessary, transform data for analysis will need validating (see IT08.03 and IT08.04). If the extraction method is part of the normal functionality of the CDMS the validation will probably already have been done, as part of the OQ / PQ of that system. If it involves additional, locally constructed processing of some kind then that processing will need validation, and/or the data in the extracted set will need to be compared with the original data in the CDMA to check that they match.

Evidence that this standard was met would come from the detailed records and summary statement(s) relating to the validation of the extraction process(es).

DM06.08 Policies for coding

If data coding is carried out, controlled documents are in place detailing the procedures to be used

In many data centres some data is coded using international standard systems, usually as an aid to reconciliation, classification and analysis of data. The best known example is MedDRA for adverse events (and in some case medical history) coding, but other coding systems include the WHO ICD system for mortality and morbidity data and the WHO Drug Dictionary sometimes used for concomitant medications.

Using such systems involves more than the simple application of codes to matching terms. Code allocation may be ambiguous, and the standards exist in different versions, so policies and procedures must be developed to support consistency in coding and to stipulate the versions to be used, or at least how decisions about version should be reached.

Autocoding mechanisms generate much discussion. While they may make the coding process quicker many staff feel they can too easily blur the distinctions that often have to be made between coding in one trial and in another. For that reason some staff prefer to use autocoding only within one trial at a time, and others are suspicious of them in general. Clear policies should therefore also exist to govern the use of autocoding mechanisms, if any are used.

The relevant evidence would be the controlled documents themselves.

DM06.09 Coding training

If data coding is carried out, it is carried out only by personnel trained on the relevant systems with access to authorised trial specific support material

Because applying codes is not straightforward the staff that do it need to be properly trained to carry out that task. In addition it is often necessary to supply such staff with support material - e.g.in MedDRA coding, a list of commonly linked symptoms that should be coded as a single entity, and a list of such symptom pairs that should be coded separately.

Common adverse events which can be classified in different ways (i.e. in MedDRA terms allocated to different system organ classes) may need to be listed against the classification that should be used - usually on a trial by trial basis. The responsibility for authorising such support material would normally fall to the sponsor / investigator, but the centre needs to ensure that such material is prepared and that the staff know how to use it.

Evidence that this standard had been met would be:

- relevant training records for the staff involved;
- examples of authorised trial specific material to support coding.

GE01 Centre Staff training and support

The standards in this section are concerned with the initial and ongoing training and support for the data management and IT staff that directly support the data centre. In most cases such staff will be based in the centre, though some IT staff may be based in IT host organisations. The standards do *not* apply to site based staff - training and support for these is dealt with in Section GE02.

During an audit the focus will be on the IT / DM staff and the documentation (e.g. training records) associated with them. The expectation would be, however, that the controlled documents and processes concerned with training and support would apply to *all* centre staff. There is no requirement for IT / data management specific policies or procedures.

GE01.01 Policies for training

Controlled documents are in place describing initial and continuing training requirements, policies and procedures

Having properly trained and competent staff managing trials and related systems is a GCP requirement. While it is not possible or appropriate for auditors to assess the competence of staff in the course of a short audit, it is possible for them to check that a centre has the proper mechanisms in place to promote and monitor staff competence.

Appropriate controlled documents should therefore exist that cover this area, detailing how initial induction as well as ongoing training should be identified, organised, signed off and recorded.

The expectation would be that induction and training was tailored to the individual's role as well as their previous experience, and which SOPs and other controlled documents any particular individual should be familiar with would be identified, so that they and their manager could ensure that they had familiarised themselves with them.

The evidence that the standard had been met would be the controlled documents themselves.

GE01.02 Documentation of training

Records of initial and continuing training and development are kept for all IT and DM staff.

All training should be documented, to show that staff have been properly prepared for their role. This includes the initial training of new staff, as well as ongoing courses, study days, workshops, webinars etc., etc. Initial training programmes should normally indicate the SOPs and controlled documents that the role holder should become familiar with during their initial training period.

Although some generally applicable training input (e.g. GCP updates) may be organised and recorded on a unit wide basis, in most cases it is far better to document training on an individual basis, for example using a separate folder for each member of staff. This is more flexible, allows greater detail to be captured, and allows training to be monitored much more easily (see GE01.03).

Training records should include, as a minimum, the dates and titles of training, but details such as duration and training provider are also useful. Individual folders can often include attendance certificates and programme details as well, and may be combined with job description(s), CVs, records of publications etc. to create a comprehensive training and development portfolio. Such folders could be held and maintained centrally or by the members of staff themselves.

The training of IT staff associated with (but often not part of) the trials unit should ensure that they are also aware of the additional data protection and GCP requirements linked to handling clinical trial data, at least as they apply to their role.

N.B. Although the standard is specifically about IT and data management staff, it is expected that the training systems would be the same for all staff within each of the relevant departments in the organisation.

The evidence that this standard had been met would be the training records themselves.

GE01.03 Managing training requirements

Mechanisms exist to review, plan and document training and development for individual IT and DM staff, with the time between successive reviews not normally being greater than 1 year.

Training requirements change, as a function of both general or organisational change (e.g. revised regulations or new systems) and individual development. In addition, training may not always be possible when initially scheduled, or become irrelevant or superseded.

Training and development needs must therefore be kept under review, and to be effective this must be done on an individual basis. A mechanism to identify needs and requests should exist and the results of that process should be documented.

In many units this will form part of an annual staff appraisal, but in others it may be part of an annual exercise in setting and allocating training budgets. The requirement for an annual review is a minimum - there will be many situations when changes in an individual's role generates a training or development need on an ad hoc basis.

As with GE01.02, the use of individual training folders or portfolios makes the training review process much easier to both manage and document.

The evidence the standard had been met would come from inspection of the relevant records, as well as discussions with staff.

GE01.04 Managing ethical or legal concerns

Staff know to whom they can go within the organisation to seek advice with ethical or legal concerns.

From time to time relatively serious problems or uncertainties about a trial's conduct may occur that cannot be resolved by normal informal discussion, for example staff may become aware of behaviour at one of the clinical sites that appears to be outside of GCP, or even the law, or there may be disagreement about the ethics or legality of a proposed action.

In such cases staff should be aware of how they can raise these issues, and with whom, to try and resolve them. Though in most cases the first stage would involve taking the issue to their line manager, there should also be a recognised 'escalation pathway' that allows staff to go, when necessary, above their immediate manager - for instance if they feel the manager is not taking their concern seriously or not acting upon the information provided.

When the problem arises outside of the trials unit the escalation pathway will normally end with senior staff within the unit or the trial's management group. In the much rarer case of a perceived problem within the unit itself, or someone feeling that senior staff are ignoring a reported issue, the escalation pathway should extend to an individual or group within the parent organisation (if there is one), allowing the individual to seek further advice.

Please note that this standard does **not** relate to 'ordinary' disputes between staff and their managers, which would need to be resolved by the relevant disciplinary and grievance procedures and the human resources department.

The evidence that this standard is met would largely come from interviewing staff, discussing the organisation of the centre and its governance, and clarifying the escalation pathways available and how staff are made aware of them.

A unit may not have a formal controlled document dealing with this issue, but some form of information (e.g. as a document for new staff, or on a web page) should be available to all staff describing the options available to them.

GE02 Site Management, Training & Support

These standards apply to the preparation and support of site staff by the staff of the data centre, with regard to data management and IT systems, and data entry and query management in particular. They are not directly concerned with overall site management issues such as site regulatory or ethical approval (though this is an indirect issue in GE02.04).

GE02.01 Policies for site opening and support

Controlled documents for opening and supporting a site for data collection are in place

Preparing and supporting site staff is a key function of any data centre and must be covered by relevant controlled documents. These would need to deal with (for instance) the training and preparation of site staff, the triggers that allowed access to production systems, the provision of documentation and ongoing support for sites.

The evidence would be the controlled documents themselves.

GE02.02 User training for data entry

User training with data entry instructions or guidelines, for pCRFs and / or eCRFs, is provided for site staff

Site research staff will need adequate preparation to correctly use pCRFs and / or eCRFs, delivered by preparatory training sessions, and / or self-study training material, written guidance, onscreen prompts and help documentation. The amount of preparation will vary with the experience of the site staff and the complexity and / or novelty of the study

The evidence that this standard is met will come from the records of training sessions and the distribution of training materials, and discussion with staff to clarify how the training is applied in practice.

GE02.03 Test or production environment

There is a clear and consistent on-screen indication to the user if they are working on a test or training eCRF.

For eRDC systems users should have the opportunity to familiarise themselves with a particular trial's CDMA within a test or training environment. It is vital, however, that the test and production environments are clearly distinguished, so that any staff member will not mistake one system for the other, and carry on putting test data into the real system:

In particular test or training eCRFs should be consistently and clearly marked, annotated or coloured to make this distinction clear. In some systems using a graphic able to reference a different image in the test and production systems might work, and make the deployment of the two systems easier.

Though included in the standards for site staff, the same consideration also applies to internal centre staff who input data for pCRF based trials, and who need initial familiarisation with the trial's CDMA.

Evidence that the standard had been met would come from demonstration of the differentiation between production and test / training eCRFs.

GE02.04 Site access to production system

A site is given access to a production CDMA only once the sponsor, or the sponsor's representative, has confirmed that all relevant preparation, permissions and agreements have been completed

For eRDC trials the production CDMA should not be available to a site until that site has been fully prepared and approved. That normally means that all contractual agreements have been signed, normally by both the site and the sponsor (or the data centre acting on the sponsor's behalf) and the relevant organisational and ethical approvals are in place.

Individuals, assuming they are properly prepared themselves (see GE02.05), should only be given access to the production system after the overall site preparedness has been confirmed.

It is the sponsor's responsibility to make the decision about a site's preparedness. The data centre may be part of the same organisation, or be acting for sponsor in this respect, but in general the sponsor needs to inform the centre when a site is 'ready to go', and policies and procedures should reflect this.

For paper based trials the 'production CDMA' at the site is effectively the set of pCRFs, which may be delivered during the preparatory phase. pCRFs should not be accepted from the site, however, until it has been officially opened.

The evidence that the standard has been met would come from the relevant controlled documents, and demonstration by centre staff of how and when actual sites have been opened.

GE02.05 Individual access to production system

Individuals have access to production data only when they have been trained with the CDMS and the specific CDMA.

The centre should be confident that the site staff can use the system properly and accurately in the context of any particular CDMA. There is no requirement for a formal exam or test. The input could be:

- Training provided at the site by data centre staff or monitors.
- Demonstrations across the web, or pre-recorded videos.
- Training material and manuals. Many centres create a generic training manual for their system(s), and then add study specific data entry instructions to that for each study.

- Provided at the site by more experienced or specialist site staff ('super-users') who can then provide guidance and training for new or less experienced staff.

In practice two or three of these methods are often used together.

To allow the competence of staff to be assessed, and to allow the staff to develop confidence themselves, most centres provide a training version, or – more normally – a dummy 'training site' for each study. Initially users are given access only to the training site, where they can add dummy patients and try out different data values, see how the system operates, how alerts and messages work etc. Of course, when the data is extracted for analysis any subjects and data in the dummy site are removed.

This scheme allows the users to demonstrate they have entered data for a few patients in the dummy system, and that they are happy with using the system, before they are given access to their normal site data. If necessary, the centre can check the accuracy of their input. If the user comes across things that they don't understand in the dummy site, they are able to input different values to see the effects of that, and / or contact the data centre for guidance.

It is difficult to describe a system that will fit every situation. Many centres specialise in trials of a certain type or disease area, and often use the same clinical sites repeatedly. In these cases only a small amount of training might be required, just to cover any trial specific aspects. On the other hand, if a centre has set up a very complex trial and is using some sites for the first time, users will probably need more training and checking before they are allowed on the production system.

The evidence for the standard being met would include the centre demonstrating it had systems in place for controlling access and for determining the most appropriate training and checking methods for any specific study, and the demonstration of some of those methods in practice.

GE02.06 Site documentation

Processes exist to update and redistribute site documentation when this is required as part of change management

A site will need to store documentation relevant to the trial - particularly the protocol and guidance material related to completing the pCRFs / eCRFs. Should the protocol and / or CDMA change those documents will need revision and redistribution to sites, and mechanisms need to be in place to support this.

Evidence would come from demonstration of the mechanisms in action, usually within the CDMA change management process (see DM03.05).

GE02.07 Responsibility list

Processes exist to assure that up to date information of who can do what at each site, including entering data and / or signing off CRFs, is available to data centre staff

Centres need to know not only which staff at each site should have access to the production system, but also what the responsibilities of those staff are within the trial, allowing them to check that only properly authorised staff carry out tasks - for instance completing CRFs, carrying out the treatment allocation procedure, or completing a SAE form. If staff leave or are away for a reason (particularly the site's principle investigator) the centre needs to know to whom his or her duties have been delegated.

In short the centre needs to keep what is often known as a 'delegate log' covering the staff for each site in the trial. How that log is maintained will differ from centre to centre - some may use monitoring or other staff visiting the centres to keep the centre informed of changes, others may ask site staff to send the details in directly to trial managers. Either way the requirement is that a list is available to data entry and trial management staff.

Evidence that the standard had been met will be:

- the presence of lists of staff and responsibilities for sites;

- controlled documents that describe how such lists are obtained and kept up to date as much as possible.

GE02.08 User Support - prompt response

The centre is able to provide Help Desk support and / or web based support (details as agreed with sponsors) to provide a rapid initial response to site requests

User support needs to be maintained during the course of the trial, and that includes the prompt response to queries or requests for help from site staff. Such support might involve a telephone hot line or it may be a web based system.

The precise nature of this support will depend on the centre's and trials sponsor's judgement about what is required, and the resources that have been made available to provide it. The requirement is that the centre is able to provide some form of prompt user support when resourced to do so.

As evidence that this is the case the centre staff would normally be expected to provide examples of current support agreements and mechanisms.

GE02.09 User Support - in English

Help desk / web support can be provided in English as well as the data centre's native language

With multinational trials user queries and requests may arrive in a variety of languages. No centre can be expected to support all the potential languages staff might use in a cross European trial, but there is a requirement that they can provide such support in English at least.

Evidence would come from direct observation.

GE03 Treatment Allocation

These standards deal with all forms of treatment allocation, i.e. both traditional randomisation, normally using permuted-block allocation, and minimisation and other deterministic methods. They are also concerned with the whole treatment allocation process, not just the parts supported by IT systems or IT and data management staff. Input from statisticians, in particular, is included in the scope of the standards.

If a data centre uses an external agency to provide some or all of its treatment allocation services, then it needs to have the evidence available that the external agency, where necessary, also complies with the relevant standards.

GE03.01 Procedures for treatment allocation

Controlled documents are in place dealing with the set up and management of treatment allocation

Whatever the treatment allocation methods used, there should be clear policies and procedures in place governing how treatment allocation should be set up and then managed.

The relevant controlled documents would provide the evidence this standard had been met.

GE03.02 Policies for ensuring blinding

Controlled documents exist covering the preservation of blinding (where used)

Though not all trials can be easily blinded (e.g. surgery and radiotherapy trials, and oncology trials involving chemotherapy) most trials that involve only oral medication will be double blinded.

In such cases it is necessary to have clear policies about how blinding is established and should be maintained (these will often cover distribution of the labelled drug as well).

The relevant controlled documents, together with explanations of how they are applied in practice, would form the evidence that this standard had been met.

GE03.03 Policies for Unblinding

Controlled documents are in place to support rapid and safe unblinding of blinded treatments when required

Clear procedures are required, in the context of blinded trials, that describe how - when the need arises -blinding can be removed. Unblinding policies should normally cover the unblinding sometimes necessary for individuals, e.g. in the context of a SUSAR, and that sometimes requested for whole treatment groups, e.g. in the context of a data monitoring committee meeting.

The relevant controlled documents, together with explanations of how they are applied in practice, would form the evidence that this standard had been met.

GE03.04 Algorithms and supporting systems

Systems used for treatment allocation are documented to show how they provide allocation sequences as specified and effective concealment of allocation.

The systems used for treatment allocation may vary considerably in sophistication, but they should be documented so that:

- The underlying algorithms are clear (or if published are referenced).
- The technical details of how those algorithms are implemented locally are available.
- The general (i.e. non study specific) validation of allocation systems is described, with reference to detailed results as necessary. This should include ongoing validation as the systems develop.
- The way in which the systems support allocation concealment, to investigators at clinical sites, is clear.
- The way in which allocation sequences are generated and managed inside the data centre, to ensure restricted access as appropriate, is also clear.

In other words, the standard requires that detailed scientific and system documentation, probably generated by statisticians and IT staff, is available for the treatment allocation systems. This is in addition to the material included within the related SOPs (the latter would traditionally deal with responsibilities, timing, outcomes etc.), or study specific requirements and implementation (see GE03.05).

The documents should cover the range of allocation scenarios the centre provides, e.g. blinded and open label trials, permuted blocks and minimisation, etc. It is recognised that in some cases allocation systems may be relatively simple and that the documentation will reflect that. Nevertheless, there should be some statements about the aspects listed above.

If the centre uses one or more external allocation services, they should still provide and demonstrate familiarity with the technical / system documentation as described above, even if parts of that may have been obtained from the allocation service providers.

Evidence that the standard has been met would come from the documentation available.

GE03.05 Specification documentation

The treatment allocation system for any specific trial should be documented, tested and approved.

The broad methodology to be used for treatment allocation will normally be included in the protocol, but each trial will also have its own detailed specification, usually determined by the trial statistician (though the sponsor will have the final decision) dealing with such things as block size, stratification factors, or the random element within a minimisation scheme.

Once the allocation method has been fully specified it can be set up, either in-house or using an external service supplier, but in either case it will then need testing. The amount of testing required will be based on a risk-assessment, taking into account, for example, the complexity of the allocation specification, its similarity to previous specifications and the previous use of / confidence in the allocation system. In most cases testing should be carried out by a statistician not directly involved in setting up the allocation system.

Once successfully tested there should be a documented sign-off against the specified allocation mechanism.

The evidence for standard compliance would be the relevant specification, testing and approval documents.

GE03.06 Problem Management in Treatment Allocation

Any problems or errors that arise in the treatment allocation process are logged and the subsequent actions recorded

Occasionally errors can arise in the treatment allocation process - subjects being allocated twice, or, if stratification or minimisation criteria were not collected accurately, being allocated to the wrong treatment group. Such cases, and the actions taken as a consequence of them, should be recorded.

The documentation of the allocation errors and the subsequent actions, together with relevant controlled documents, provide the evidence that the standard has been met.

GE03.07 Treatment Allocation Training

All staff who handle allocation requests are adequately trained for each specific trial randomisation process

Treatment allocation is often complex and cannot always be completely automated. Where staff are involved, even if it is just noting down stratification criteria, they must be adequately trained so that errors do not occur (or are at least minimised).

Evidence that the standard had been met would come from records of training and explanation about how treatment allocation is distributed amongst staff within the centre.

GE03.08 Record of Allocation

Records of all allocation material generated and all allocation decisions made must be maintained

The treatment allocations made during a trial are a vital part of that trial's history and must be retained, for as long as the trial data is retained.

This means keeping the original randomisation lists, and the minimisation decisions in their correct order (i.e. context), and not just the resulting treatment allocations. Controlled documents would normally specify the process by which this data was stored, as well as the access control required.

These controlled documents, together with examples of the lists themselves, would provide the evidence that the standard had been met.

GE03.09 Failover to Manual

System(s) must be in place, supported by training, to deal with a loss of IT based treatment allocation (if used)

When treatment allocation uses IT there is always the problem of what to do when for some reason that IT system is unavailable. Treatment allocation should still be able to continue if subjects are presented for inclusion. A centre must therefore have systems in place to cope with this situation, for all trials being allocated at any one time, with the staff involved suitably trained to use whatever methods have been identified as suitable.

Manually allocating treatments from permuted block lists is usually fairly straightforward, but manually applying minimisation algorithms can be complex, and may demand specialist expertise. In either case there will be the need to ensure that once restored the IT based systems are brought up to date with any allocations that may have occurred when they were down.

The relevant controlled documents, training records and discussions with staff would form the evidence that the standard had been met.

GE04 Transferring Data

Transferring data refers to sending the data out of the centre, not just removing it from the CDMS (which is data extraction or export).

Transferred data will leave the centre's IT network completely and be sent to another institution. It may occur in the context of a collaboration or meta-analysis, or sending data to a statistician or investigator based elsewhere for analysis or review. For an industry sponsored trial it may include sending data to that sponsor.

GE04.01 Data Transfer Procedures

Controlled documents dealing with the transfer of data from the data centre should be in place

This standard requires that there are controlled documents that describe the principles to be followed when transferring data, including the documentation required.

Final decisions about who to send data to and when will rest with the sponsor or a trial management group acting on the sponsor's behalf. A centre should still, however, have procedures in place to ensure that it transfers the data safely and accurately and records the entire process.

Probably the most common form of data transfer, and in many ways the simplest, is sending data back to a sponsor organisation. But, at a time when transparency in research, including making data available to other researchers, is increasingly being emphasised, it is important that units have procedures in place to deal with *all* sorts of data transfer requests.

Procedures may vary, depending on how much a data centre is involved in the sponsor's decision, ranging from 'not at all' to acting as the sponsor's full proxy. Even in the former case, however, the data centre needs to consider the potential risks of identification of any sensitive data that it transfers. Those risks have potential implications not just for the study participants and the sponsor, but also for its own reputation, and that of its parent organisation. A centre should design its procedures accordingly. For this reason, some data centres request that applicants for data sharing complete a proforma, asking for details on why the data is requested, how it will be stored, etc., as well as covering technical details like the fields and formats required.

The procedures and documentation required in any particular centre will depend upon the nature and frequency of data transfer out of the centre, and those procedures should be reviewed if the types of data transfer change.
The evidence that the standard had been met would come from the controlled documents themselves and the documentation associated with specific transfers.

GE04.02 Encryption of Individual Data

Any file(s) transferred out of the data centre that include data relating to individuals should be encrypted

If transferred data includes data relating to individuals it must be encrypted, to the level considered as good practice by the national regulatory authority (currently 128 or 256 bit AES encryption). This reflects the difficulty in distinguishing patient identifying data from other data relating to individuals (see IT02.03).

Encryption may occur before transfer, which would be required if the medium is a portable device like a USB stick or CD. In some cases, encryption may take place *during* electronic transfer, when using a secured system, as is sometimes the case when sending data to industrial sponsors.

Because transferred data may be commercially sensitive even when it does not include individual level data, it may be safer and easier to routinely encrypt *all* such data. Sending encrypted data electronically as an attachment is now very difficult because recipient systems will normally remove it as an unknown and therefore potentially malicious file. Encrypting all data allows the development of a single procedure for the physical transfer of all data.

Evidence that the standard had been met would include the relevant controlled documents and explanation of how encryption of transferred data was carried out in practice.

GE04.03 Format of Data Transfers

Procedures should be in place for agreeing, specifying and documenting the format of the transferred data

The format of the transfer will depend on whatever is agreed between the centre and the recipient. The centre only needs a mechanism to agree the best format(s) to use.

A data transfer proforma that can be sent to the recipient for completion allows this and the other data required to be captured in a structured way, making discussion around and documentation of the whole process easier.

Evidence that the standard has been met would come from the relevant controlled documents and from the documentation associated with data transfers.

GE04.04 Records of Transfers

Details of any specific data transfer should be logged, and include a summary description of the data, sender, recipient and transfer method, and the date sent

Once the transfer takes place it needs to be recorded, and include the data listed in the standard.

Evidence that the standard had been met would come from the documentation associated with data transfers.

GE04.05 Retention of Copies

Copies of the data sent should be retained within a read only regime and be available as a reference data set for audit / reconstruction purposes.

After the transfer takes place the centre should keep copies of all data sent, for audit purposes and in case it needs to be sent again, in a read only section of its storage capacity (i.e. read only apart for the IT staff who need to put the data in that location).

Evidence that the standard had been met would come from demonstration of transferred data in an appropriate read only environment.

GE04.06 Retention of post-processed data

If data is processed before being transferred, copies of the data as extracted before post processing should be retained as well as copies of the data actually sent

It is not uncommon for data to undergo some form of pre-processing before it is transferred outside of the data centre, for instance to the format agreed by the sender and recipient if that is not the same as that generated natively by the CDMS (e.g. conversion to CDISC ODM).

In these circumstances it is important that the data as originally extracted (as well as that which is finally transferred) is retained in a read only environment, so that a complete history of the transfer process is available and, if necessary, the post processing can be checked and / or repeated.

Evidence that the standard had been met would come from demonstration of both extracted and transferred data in an appropriate read only environment.

GE05 Receiving and Uploading Bulk Data

Centres often need to upload and import bulk data from a variety of external sources: laboratories (e.g. biomarker data), instrumentation (e.g. the settings from a radiotherapy machine), collaborators (e.g. data from another set of sites), or even the sponsor (e.g. SAE reports).

There are usually two stages to the process - firstly data receipt into the centre, i.e. of the data files as sent from the source, and secondly data upload or import into the data centre's own systems. In many cases the data will need to be processed in some way before the second upload stage can begin.

The data uploads may be directly to the trial's CDMA, or they may be to a data repository system that itself receives data from the CDMA, allowing aggregation of all the data, whatever its source, before a combined extraction for analysis.

GE05.01 Import Procedures

Controlled documents dealing with receiving and uploading bulk data should be in place

The receipt / upload process should be governed by pre-specified generic procedures and processes, as required by this standard.

In practice each upload process will also probably need its own more detailed procedural guidance if consistency is to be maintained, especially if - as is often the case - the data needs transforming in some way before it is imported.

The relevant controlled documents would provide the evidence that this standard had been met.

GE05.02 File Retention

The original files received should be retained within a read only regime, and be available as a reference data set for audit / reconstruction purposes.

To ensure a full audit record, and in case the import needs to be repeated for any reason, it is important to keep copies of the data as originally received.

Evidence that the standard had been met would come from demonstration of the original data in an appropriate read only environment.

GE05.03 Retention of post-processed data

If imported data has to be pre-processed before upload to the CDMS, copies of the data actually uploaded should be kept within a read only regime

When data must be processed before it can be imported then, unless that processing is trivial, predictable and can be quickly repeated, the data as it stands after processing should *also* be kept, i.e. the data that is actually imported into the system.

Processing in this context may not just take the form of a consistent transformation, e.g. of format or data type. Especially when data arrives in a relatively unstructured form such as a spreadsheet, it may also be necessary for the data to be scanned for any values that appear to be errors or out of normal range, and make the necessary corrections in a more ad hoc way.

Evidence that the standard had been met would come from demonstration of both received and imported data, i.e. each side of the pre-processing, in an appropriate read only environment.

GE05.04 Logging of receipts and uploads

Each receipt and upload process should be documented and logged

The receipt and upload process itself should be logged (though recording the actual import may be an automatic function in some CDMSs). Logging should normally include the source organisation, a summary of the contents, the location of the copies of data and the date, as well as any problems that arose during the import.

If importing into a CDMS, it is very useful if the system can apply the normal validation logic that is used during manual data entry to the incoming data, generating warnings etc., and allowing a review of any problematic data items (though this is not a formal requirement).

Evidence that this standard had been met would come from the receipt / upload logs themselves.

GE05.05 Format of received data

Procedures should be in place for agreeing, specifying and documenting the format of the received data

As with data transfer out of the centre (see GE04.05) the format of the data received should be agreed between the centre and the source organisation. The centre therefore needs a mechanism to agree the best format(s) to use.

Evidence that the standard has been met comes from the relevant controlled documents and from the documentation associated with arranging data imports.

GE05.06 Direct amendment of data in the database

Procedures should exist to deal with requests for direct changes of data in the database

This standard deals with a relatively unusual situation, and one that some centres may never experience. It involves the need to directly change data in the back end database or file store, rather than going through the normal CDMA user interface.

Such situations can arise if data is imported to a CDMA in bulk (so there may not be an eCRF corresponding to it in the system) and then needs correcting. If the

data import is a regular event and designed to over-write existing values there is little problem - the data can just be re-imported with the corrected values.

If the original input was intended as a one-off, however, then any amendments required will need to be done manually on an ad hoc basis. An example might be an imported treatment allocation list (i.e. subject trial ID against treatment received, A or B) that had to be amended because one or two subjects were found to have received the wrong treatment.

A centre should be prepared for such a situation, or prohibit it entirely and insist on another method of editing the data (e.g. by repeated bulk upload according to a specific procedure). If the centre *does* allow direct data amendment, then because each change request will be different there is little a centre can do other than have a very generic procedure, for instance that identifies how the change request would be considered and by whom, who would carry out the action decided upon and how the whole process should be fully documented.

If direct amendment of data does take place then it must be recorded, with all details noted and communications (emails etc.) retained, perhaps as a file note in the trial's master file, and this should be made clear in the associated controlled documents.

The evidence for compliance would be the procedure itself and the records of any data amendment.

GE06 Long Term Data Storage

Trials eventually reach a point when data is no longer being input, all outstanding queries have been resolved and all the anticipated papers have been written. Direct access to the trial data, in paper or electronic form, is either no longer required or limited to occasional read only access. At this point the trial enters long term data storage.

The trial is not necessarily formally 'archived' or curated at this point. It could be, though very few data centres appear to have mechanisms in place to provide a full digital curation service for electronic data, even if many have separate long term storage facilities (which may or may not be called an 'archive') for paper based data.

The characteristic of long term storage is restricted access and thus protection from change. The trial's electronic documentation and its data become hidden or read only (though some at least of the IT staff need to retain access in order to resurrect the data to active use if necessary). Its paper data records are moved away from the normal storage locations and into a special store reserved for old, no longer active records, which may not be at the same physical location as the rest of the centre.

In the future keeping electronic data over the long term may also mean changing the format of that data, to make it less dependent on proprietary systems that may disappear in the future. Possible target formats are CSV files or XML, e.g. using the CDISC ODM format. The latter has the great advantage of being able to include metadata definitions as well as the data. Anonymising the data so that storage can be encryption free (which avoids the difficult issue of long term key management) is another useful technique for long term curation.

At the moment the standards do *not* include such active data transformations, though they may in the future, especially as long term curation becomes a more prominent issue and these techniques become more common.

GE06.01 Policies for long term storage

Controlled documents are in place concerning long term storage of both trial documents and electronic data

Moving a trial's data and documents to long term storage should be the subject of controlled documents that describe the overall approach and procedures, roles and responsibilities.

Evidence that the standard had been met would be the controlled documents themselves.

GE06.02 Long term storage of documents

Measures are in place to ensure secure storage and controlled access to paper based records in long term storage.

Long term storage for paper based records should be secure and include environmental protection for documents (against fire, damp etc.). Ideally, there would also be the ability to lock individual cabinets or shelving so that access to one group of documents does not mean access to all. In some cases the centre might make use of external archive facilities, or a service provided by their parent organisation, rather than storing documents within their own premises.

Access to the data in long term storage should be controlled, usually with designated staff acting as the 'gatekeepers' to the stored material. This allows access and any retrieval of documents to be recorded and monitored.

Evidence that the standard had been met would be provided by inspection of long term storage facilities, discussion of the access procedures, and the records of access and / or retrieval.

GE06.03 Long term storage of electronic data

Measures are in place to ensure secure storage and controlled access to electronic based records in long term storage.

Long term storage of electronic data is usually managed by removing access to it from users, except for the IT staff themselves, effectively isolating the data. In most cases data in electronic long term storage therefore stays within the normal storage capacity of the centre, but is just not visible to normal users.

Though such data no longer needs to be part of a regular backup procedure (because it is no longer changing) there is a need to ensure that independent copies of the data exist and can be accessed relatively easily if ever required. 'Ordinary' backup systems are usually configured to provide relatively short term redundancy and security and are **not** intended to cope with long term storage. Other mechanisms may therefore need to be used to provide redundancy in the long term.

Access to the data in long term storage should be controlled, usually by IT staff acting as the 'gatekeepers' to the stored material. This allows access to individuals or groups to be managed and recorded, with restrictions re-applied when required.

Evidence that the standard had been met would be provided by discussion of the storage and access regimes for long term electronic storage, by the procedures described in the relevant controlled documents, and the records of access.

GE06.04 Length and Content of Storage

Procedures should be in place to agree with the sponsor the length and content of long term storage.

The material (paper and electronic) that is placed in long term storage is there partly so that the data can be consulted when necessary, partly so that the trial

itself can be re-examined and its design, implementation and results can be fully understood.

This does not mean that it is necessary to be able to restore the trial's data and systems to exactly their original state - systems will move on, change versions, even whole applications - but it does mean being able to inspect the sequence of the major events within the trial, as well as the data and the generic and trial specific documents that formed the framework within which it operated.

The key to being able to do this is selecting the right material at the beginning of the long term storage process. Ultimately this will be the sponsor's decision but the centre, if it is responsible for carrying out the long term storage, should have procedures in place to agree that content with the sponsor.

Evidence that the standard had been met would be the relevant procedures, as described in controlled documents, and examples of their use in practice.

GE06.05 Final fate of data

Procedures should be in place, or be being developed, to decide with the sponsor the final fate of physical and electronic data

Eventually, data and documents will either be destroyed or archived, the latter often preceded by preparatory steps such as anonymisation, the creation of associated metadata, and transfer to specialist media.

The final decision will be taken by the sponsor, acting in the context of national regulations, but the centre's procedures should include mechanisms to identify with the sponsor the retention periods and the nature of the final fate of the data (which may be different for paper compared with electronic data) for any particular trial.

This is a rapidly changing area, however, especially with the recent interest in data repositories for individual level trial data, which may impact how trial data is archived in the future. Furthermore, some younger centres will not yet have reached the stage of destroying or archiving data. The standard therefore allows a centre to be still developing relevant procedures - the key thing is that the centre is addressing the issue of the 'final destination' of their trial data and examining how that will be decided.

Evidence that the standard had been met would be by discussion of current and any planned processes and procedures.

3. Glossary

This section provides explanations of some of the terms and abbreviations used within the standards and supporting material. Many of these terms are relatively common but because of that are often ambiguous. A more precise definition is therefore provided, at least for their usage in this context.

AdaM: The Analysis data model is a CDISC standard for describing and documenting analysis datasets, particularly in the context of regulatory submission. The underlying principle is that the design of analysis datasets, and the associated metadata and documents, should together provide an explicit description of the content of, input to, and purpose of any submitted analysis dataset (see *CDISC*).

Aggregated data: data only about groups of study participants, as provided in statistical summaries and the research papers derived from the study.

Anonymised data: clinical data from which the obvious PID (participant identifying data) has been removed. While such data often contains a unique identifier for each participant, that identifier *cannot* be linked to any identifying data. Anonymising data is a one-way process - once done the data cannot normally be linked back to individuals (see also *Pseudo-anonymised data*). It is difficult to *guarantee* anonymisation of data – in some cases clinical details, especially in the context of rare diseases, and / or linked geographical information, and / or linked genomic information, may allow the individuals that provided the data to be identified. Data is considered anonymised when the practical barriers to identifying individuals are so high that the process is impractical.

CDASH: The Clinical Data Acquisition Standards Harmonization is a CDISC standard designed to help standardise data collection, by providing predefined data fields for 18 domains, e.g. adverse events, demographics and others that are common to most therapeutic areas and phases of clinical research (see *CDISC*).

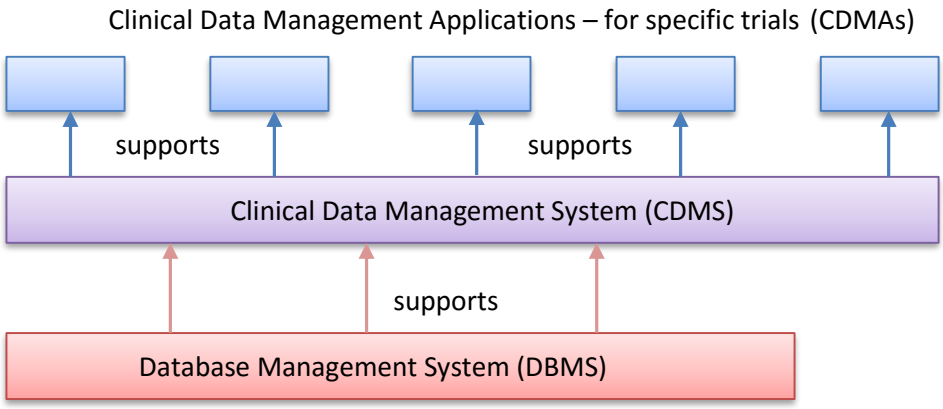
CDISC: CDISC, the Clinical Data Interchange Standards Consortium (http://www.cdisc.org/), is a global non-profit organization that has established standards to support the collection, exchange, submission and archive of clinical research data and metadata. The CDISC mission is “to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare.” (see also *AdaM*, *CDASH*, *ODM* and *SDTM*).

Centre: is used to refer to the organisation or team seeking certification as an ECRIN data centre, even though it may call itself a trials unit, a research centre, a clinical research department, a trials and statistics co-ordination centre, or any one of the many variations on these titles. If there is a risk of ambiguity the term data centre is used.

Clinical data (or ‘individual data’, or ‘data relating to individuals’): is used to refer to any data that is associated with an *individual* trial participant, whether or not it describes a clinical symptom or situation. In particular, it could include demographic, treatment and lab details – anything that is considered as relevant to the study and which is an attribute of a single study subject or their experience.

Clinical Data Management Application (CDMA): refers to the specific system established to hold the data for a *single* trial. As well as the data itself, the CDMA contains the schedule

and check logic for that trial, and the specific data collection instruments, i.e. the eCRFs, that have been set up for the trial. A CDMA is therefore a specific application of the underlying CDMS. The relationship between CDMAs, the CDMS and the DBMS is diagrammed below (see *CDMS* and *DBMS*).



Clinical Data Management System (CDMS): Within centres, the system (or collection of systems) that holds the clinical data gathered during trials. CDMSs are specialist software systems and are often purchased from specialist vendors, but may be built and maintained in-house. Examples are Medidata Rave, OpenClinica, InferMed Macro, Omnicomm TrialMaster and RedCap. Within the CDMS, each study will have its own logically separate CDMA (see *CDMA*).

Controlled Documents: is the generic term used for *all* quality management documents that are authorised (i.e. signed off as correct and designated for implementation) by one or more people, and which are version controlled. They include SOPs and work instructions, and most policies. Most organisations keep their controlled documents within electronic filing systems and apply document management to differentiate the various versions. Because different units designate different controlled documents differently within their quality management systems the standards always use the generic 'Controlled Documents' rather than the more specific SOPs, work instructions etc.

CRF: is the generic term used for all types of Case Report Form (see *pCRF, eCRF, iCRF*).

Data relating to individuals: = clinical data, as defined here

Database Management System (DBMS): This refers to the underlying data storage system for a CDMS, often known as the 'back end' database. Almost all CDMSs use a commercial database system for data storage, e.g. Microsoft's SQL Server, Oracle, PostgreSQL, or MySQL. Most use a relational table structure and some variant of SQL (Structured Query Language) to access and edit data and table structures.

eCRF: In the context of eRDC the electronic screen based case report form, used for direct input into the CDMS from the clinical site. eCRFs normally include validation and range checks so that unlikely values can be flagged, and errors corrected, during initial data entry.

eRDC: is the term used here for electronic remote data capture, i.e. data entry direct from sites. In most eRDC systems access for data entry will be via a web browser.

Guidance notes: see *Work Instructions*

iCRF: (interim CRF) In many cases it is not practical for research staff to access eRDC systems while interviewing patients and / or collating information, and in any case many staff prefer not to do so, feeling it is disruptive to the interview and uncomfortable for the patient. In such cases it is useful to have a paper version of the eCRF, to capture data in a structured and accurate way, rather than simply making notes freehand. This paper CRF, probably printed from the eRDC system and used / retained within the clinical site, i.e. not sent to the trials unit, is here referred to as an interim or iCRF.

individual data: = clinical data, as defined here

IT host organisation: is the organisation responsible for managing a particular component of the centre's IT systems – exactly which component will vary with the context. To keep things simple, the body providing the IT component, which might be the centre itself, it's parent organisation or an external host, are all referred to as the IT host organisation.

MedDRA: acronym for Medical Dictionary for Regulatory Activities, used as a coding system for pathologies and adverse events in most clinical trials.

ODM: The CDISC Operational Data Model (ODM) is an XML format for interchange and archive of clinical research data. The model includes participant data along with associated metadata, administrative data, reference data and audit information. Unlike SDTM, which imposes its own structure on the dataset, the ODM can describe the meta- and clinical data in their original forms, for instance as stored within or extracted from a CDMS (see *CDISC*).

Parent organisation: is used to refer to that organisation (or organisations) to which the centre belongs – normally a university or a hospital, sometimes both. In some contexts it may mean in practice just that part (e.g. faculty, clinical directorate) which directly contains the centre, in others the whole organisation.

PID, Participant or Patient Identifying Data: any data within clinical data that could potentially be used to identify subjects, either directly or by linkage to other systems. PID obviously includes names and initials, but also hospital system IDs or national health service / insurance IDs, numbers which in conjunction with those systems would identify an individual. Dates of birth can be PID, though normally not in a large data set and without other associated data (e.g. identifying source hospital) when identification would be difficult. *There is no absolute definition of PID* - it depends on the size of the data set and what data is present. Any clinical data can be PID if it is rare, in a small data set, or linked to other information (e.g. geographical location).

pCRF: The traditional paper based case report form, distributed by the trials unit to the sites and then returned completed, usually by post or courier.

Policies: Fairly general statements of the aims of the organisation with regard to a particular aspect of functioning. Policies will usually be distinct documents approved by a senior manager or committee, and may or may not include a broad brush description of how the policy should be carried out. Some policies may only be written down only as minutes of meetings, however, so not all will necessarily be formerly controlled documents. Policies would normally trigger the production of supporting SOPs (see *SOPs*).

Pseudo-anonymised data is data from which the *obvious* PID (participant identifying data) has been removed, but which contains a unique identifier for each individual subject. That

identifier not only groups and labels the data for a single subject, it can also be used as a key to link the data back to the subject's identifying data, if and when necessary. The identifying data must be stored separately (and normally more securely) from the pseudo-anonymised data. (see *anonymised data*).

Remote access: as used here, is *not* the same as eRDC. It refers instead to the process whereby collaborators (including other trials units) and centre staff working away from the centre premises gain access to the CDMS using technologies like Citrix, Terminal services or VPN, as well as browser based methods. This may involve data entry, but could also include other functions like entering monitoring results, or even CDMA design. Remote access is therefore a more general term than eRDC, and can include a wider range of access methods and functionality.

SDTM: The Study Data Tabulation Model is a CDISC standard for presenting data for regulatory submission, and in particular to the FDA. It imposes a particular structure on the data, dividing it into specified 'domains' and specifying field names for data points within those domains.

Site: is used for the various clinical and other data collection locations that are participating in a trial and that provide the data to the centre.

SOPs (Standard Operational Procedures): Controlled documents, with version control and relevant authorisations, application/review dates etc., which provide a description of procedures to be followed, describing and assigning responsibilities for the tasks and subtasks, and identifying the ordering, inputs and outputs of the processes involved. An SOP should be specific enough to be auditable and provide the necessary guidance to staff. They can often overlap with policies in scope, but are usually more specific (see *Policies*). SOPs normally form the backbone of any quality management system, with more detailed documents like work instructions and forms being linked to them.

Systems directly supporting Clinical Trials: This phrase, and minor variations of it, refers to all systems that store or process trial clinical data or analyses, trial administration and financial data, or trial specific documents (e.g. protocols, agreements), i.e. all things that directly support trial activity and that would stop or disturb that activity if they malfunctioned. It *excludes* systems exclusively used for development, testing and training, and systems that only store non trial specific documents and data (e.g. general centre inventories, staff and budgetary information). It *includes*, however, mirrored or back up servers, even if they are normally passive partners, that could be called into immediate action as part of a failover mechanism.

Work Instructions (WIs**):** also known as Procedures or Guidance Notes, are the detailed procedural documents (or web pages) that describe how to actually carry out tasks. They are usually linked to, and referenced by, one or more SOPs. These documents should also be controlled (i.e. there should be a clearly defined current version) but may not require the full review / authorisation procedure of an SOP. For instance, an IT work instruction may be better revised and distributed by the IT manager, in conjunction with his or her team, rather than the full quality management team (see *SOPs*).