ECRIN

EUROPEAN CLINICAL RESEARCH
INFRASTRUCTURE NETWORK

Requirements for Certification of ECRIN Data Centres

with

Explanation and Elaboration of Standards

Version 4.0
April 2018

Requirements for Certification of ECRIN Data Centres

with

Explanation and Elaboration of Standards, Version 4.0

Preface

This is version 4.0 of the 'Requirements for certification of data centres', published by ECRIN, the European Clinical Research Infrastructure Network (the first version was produced in 2011, the second in 2012, and the third in 2015). 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 2017 and 2018 by ECRIN auditors, members of ECRIN's data centre Certification Board, and invited experts from a variety of trials units in Europe. The full list of contributors is provided on page *iv*.

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 106 requirements, or standards, included in the current version are divided into 16 separate lists, some focused on IT, some mainly concerned with data management, and some that deal with more generic aspects of trial management. Each standard has a code, a title, and a single statement summarizing the requirement. This document provides, in addition, explanatory and elaboration material that attempts 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 Data Centres audit process, a glossary of terms, and a summary of the main changes from the previous version.

Contents

Contributors (version 4)

Alex Benalte Gasco, Hospital Clinic, Barcelona

Steve Canham*Ŧ, independent consultant in clinical trials IT, London

Will Crocombe, independent consultant in clinical trials IT, Leeds

Nancy De Bremaeker*, Luxembourg Institute of Health

Carlos Domingues*, AIBILI, Coimbra

Michael Faherty*, National University of Ireland, Galway

Maria Teresa Garcia Morales, Hospital 12 de Octubre, Madrid

Laura Houston, University of Wollongong, New South Wales

***Jens Lauritsen*¶**, Odense University Hospital, University of Southern Denmark

***Enrico Nicolis*¶**, Mario Negri Institute, Milano

***Christian Ohmann*¶¶**, ECRIN, Prinz-Georg-Str. 51, 40477 Düsseldorf

Catherine Pham*, Clinical Trial Consultants AB, Uppsala

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

***Christine Toneatti*§**, ECRIN Quality Manager, Paris

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

Steve Canham provided the original draft of proposed revisions and co-ordinated the review process. Other authors provided suggestions, comments and alternative proposals. All reviewed and approved the final proposals.

Additional contributors to previous versions

Luca Clivio, Mario Negri Institute, Milano, Italy;
Catherine Cornu*, Hospices Civils de Lyon, Lyon;
Jochen Dreß*, then at Centre for Clinical Studies (ZKS), Köln;
François Gueyffier*, Clinical Pharmacology and Clinical Trials Department, University Hospitals, Lyon;
Wolfgang Kuchinke, KKS, Heinrich Heine University, Düsseldorf*;*
Susan Lennon*,* then at Molecular Medicine Ireland / ICRIN, Dublin*;*
Gilles Palmer, Institut de Santé Publique, d'Epidémiologie et de Développement, Bordeaux ;
José Miguel Pêgo, School of Health Sciences, University of Minho, Braga;
Jadranka Rogan, then at Cyathus Exquirere Pharmaforschungs GmbH, Vienna.

*****ECRIN Auditor, **Ŧ** Chair of standard review process, § Secretary to the Certification Board,
¶ Certification Board Member, ¶¶ Certification Board Chair

Introduction and Background

This document describes the systems and functionality that a non-commercial clinical 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 (DM) practices, others concerned with more generic aspects of trial management, but all indicative of safe, effective and compliant data storage and data processing.

The 106 standards are divided into 16 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 [1]). 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 required to demonstrate compliance. 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 focus of the standards, the audit and the certification is the *IT and data management activities* of a clinical research unit, even though that unit will usually be involved in many other aspects of the research process — writing protocols, gaining regulatory and ethics 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 provides a public indicator or 'badge' of quality, backed up by public standards and a rigorous assessment procedure. 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 [2]).

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, pragmatic 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 [3].

The original version of the standards was used for audits within the data centre certification pilot phase, at Düsseldorf and Uppsala, in November 2011. The experience of the pilot phase

led to a substantial revision. The new version created (Version 2.2, July 2012), had 139 standards divided into 21 distinct lists. A description of the revision process, and a summary of the standards that resulted from it, can be found in [4], which also includes a full version of that version of the standards as a supplementary file. Version 2.2 was also translated into French [5].

Version 2.2 of the standards was used for certification audits in 2014 and 2015. In June 2015 a further review took place with input invited from auditors, from certification board members, and from specially invited experts in clinical trials IT systems. The result was version 3.0, published in October 2015. The changes were evolutionary and slightly simplified the requirements: 129 standards were now divided into 19 lists. After minor referencing errors and typos were corrected, version 3.1 was produced in January 2016, and this was the version used in audits in 2016, 2017 and early 2018.

The assessment process

The ECRIN standards are designed to be used as the basis of an on-site audit by appointed ECRIN auditors. 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.

ECRIN audits are planned to last up to three days, and normally involve a team of three auditors, all of whom are experienced trials unit staff. The audit results and auditors' recommendations are sent first to the audited unit (to allow them to comment and correct any factual errors) and are then passed to ECRIN's Independent Certification Board (ICB), who make the final decision about the certification of a unit as an ECRIN data centre. The audit is normally conducted in English, but ECRIN tries to ensure that the audit team includes at least one individual who can speak, natively, the language of the data centre, so that all evidence can be inspected.

A centre will be awarded certification if the ICB is confident all criteria (i.e. all standards) 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 follow up re-audit, to confirm that the required 'corrective and preventative actions' (CAPA) have been carried out, after which they will reconsider the certification decision. Otherwise the unit will need to re-apply at a later date.

Many trials units have experienced radical changes in their processes and procedures in recent years, so that data and IT management may be radically different from what it was only a few years ago. ECRIN auditors are interested in the arrangements made for *current* and *future* trials, so will focus their audit 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.

Note that the auditors expect to see trial specific processes (as described in current quality documents such as SOPs) demonstrated in the context of at least two trials.

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 and not as absolute requirements.

Subcontracting and 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. Common examples are:

- IT infrastructure services provided by a university or hospital central IT department, rather than being housed within the trials unit itself,
- a SaaS (software as a service) version of a clinical data management system, where the system and the clinical data are hosted externally and all access, from the data centre as well as the clinical sites, is via the web.

In such circumstances it is important to remember that if the sponsor has delegated the responsibility for IT and data management to the data centre, *the centre still retains that responsibility* even if it has itself delegated some functioning to others. That means that *the centre must itself monitor its own service suppliers*, to ensure that their activity is regulatory compliant and all is functioning as it should be.

The recently revised version of the ICH's guidance document for Good Clinical Practice, E6 (R2), makes this point explicitly:

> "The sponsor should ensure oversight of any trial-related duties and functions carried out on its behalf, including trial-related duties and functions that are subcontracted to another party by the sponsor's contracted CRO(s)." [6]

In this context the data centres are taking the role of 'sponsor's contracted CRO(s)', and given that in most situations the sponsor also delegates the responsibility 'to ensure oversight' to the contracted CROs, the data centre now needs to make sure this happens to remain GCP compliant.

Unfortunately, the evidence is that oversight is not always performed as rigorously as it should be. The Inspectors Working Group of the European Medicines Agency have listed a wide range of issues that they have discovered in respect of sub-contracted services, [7], including missing

or out of date contractual agreements, poor definition of the distribution of tasks, lack of understanding of the location of data, a lack of understanding of GCP obligations by subcontractors, unwillingness to accept audits, poor understanding of reporting requirements, and confusion over outputs and actions to be taken at the end of the trial.

We have found similar problems within ECRIN audits, especially with regards to subcontracted IT infrastructure. Centres sometimes appear too willing to 'leave it to the IT people', with the result that they often poorly understand, and do not monitor, activities like backup and restore testing, ongoing system validation, and system patching and updating, even though they are still responsible for the proper execution of all these tasks.

Within the ECRIN standards we try and make it clear that, even if a centre is not carrying out the operational day-to-day tasks involved in an activity because they have sub-contracted it to some other organisation, or some other part of their own organisation, they remain responsible for its proper operation and must provide evidence that:

- **The sub-contracting organisation is performing its operations to the standard required,**
- **'the standard required' is based on a written and mutually understood definition of responsibilities, and**
- **the data centre has an oversight mechanism in place to ensure that the standard is being met.**

An 'oversight mechanism' means more than having good contractual agreements and SLAs in place, or even doing an initial supplier audit (though all of those are very useful) – it implies an ongoing risk based process of monitoring and / or periodic review. That in turn demands well defined communication channels, with clear understanding by all parties of the responsibilities of each, and a willingness by subcontractors to support the compliance needs of the data centres [8].

This point is re-iterated and expanded upon in the Explanation and Elaboration material for many individual standards, especially those dealing with IT, but the general principle applies to all types of activity and all of the standards.

The optional treatment allocation standards

Treatment allocation (TA) mechanisms are very important. Nevertheless, this group of standards on randomisation, minimisation and supporting systems have been moved into Appendix A in this version of the standards, and are now labelled as optional. Compliance with them is no longer required for certification as a data centre, because units vary so much in the range and sophistication of the treatment allocation services they provide. Including these standards within the certification process has therefore made the certification standard inconsistent.

The TA standards are provided for self-assessment, and / or inclusion in the ECRIN audit discussion *if a unit wishes*. In the latter case the unit effectively receives some additional free consultancy, and the discussion around these standards is not part of the formal assessment.

The *exception* would be if the unit decided it wished to provide TA services to ECRIN supported trials (or act as the 'lead CTU' in those trials, which would normally include the TA function). If that was the case ECRIN would then use these standards to assess the quality of treatment allocation systems, independently of any decision about data centre certification.

The intention is to modify the certification application process so that units can indicate if they wish the assessment of TA systems to be included in the audit or not.

There are other 'non-core but related' activities around IT and DM, in which units often vary enormously in the level of services provided. The most obvious are probably monitoring and pharmacovigilance, though management of laboratory derived data, the use of data standards and the preparation of data for archiving and data sharing have also been proposed. Similar optional sets of standards may therefore be developed in the future, with similar usage.

The standards – terminology and phrasing

The following 16 sections list the 106 ECRIN standards. In each case the standard code and title are followed by the requirement statement in bold text. The explanation and elaboration material, usually with notes on expected evidence, is provided below.

Note that several common terms (e.g. 'Centre', 'Site', 'Controlled documents') have specific meanings within these standards and the support material. In addition a few terms (e.g. 'CDMA') have been developed specifically for the standards. Please refer to the glossary at the end of this document for definitions of the terms used and explanations of abbreviations.

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 the standards this way, especially when the imperative is repeated many times.

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 DM04.

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 training (or 'induction') as well as ongoing training should be identified, organised, signed off and recorded.

The expectation would be that initial and ongoing training were tailored to the individual's role as well as their previous experience, that the SOPs and other controlled documents relevant to each role had been identified, that a mechanism existed to ensure that staff were familiar with the procedures and systems relevant to them, and that any changes in those procedures or systems were transmitted to the appropriate staff.

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. Initial training records should normally show how and when the role holder become familiar with the SOPs and controlled documents relevant to them, and include a final appraisal or 'sign-off' that clearly indicates the end of the 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 concerns – alternative pathways

Staff should know to whom they can go within the organisation to seek advice with ethical or legal concerns, if discussion with a line manager or trial management group does not resolve an issue.

In almost all cases, if a member of staff became aware of such an issue, they would initially take it to their line manager, and /or the trial management group, and the issue would be investigated and resolved – at least in the sense that trials unit staff all agreed on the chosen response and the reasons for it.

Very rarely, however, a member of staff may feel that the concerns they have raised have not been taken seriously (or perhaps even believed) by their line manager, or the management group to which they have reported their concerns, or that the response has been inadequate.

In such a situation, there should be a recognised 'alternative escalation pathway' that allows staff to go, if and when necessary, outside the normal reporting hierarchy to express their concerns, and all staff should be aware that it exists and how to access it. This might be via a particular office / post within the parent university or hospital. It might be direct to a sponsor representative if they were external to the trials unit. Whatever it is, staff should be aware that it exists and know how to begin the access process (e.g. by being given an email link and / or a telephone number). Such information could be given as part of the initial introduction to the unit.

Please note that this standard *does not* relate to 'ordinary' disputes between staff and their managers, for example about conditions of work, inappropriate requests, etc., which would need to be resolved by the relevant disciplinary and grievance procedures and the human resources department.

The standard *does* relate, however, to the need for a unit to accept that – however rarely it might be needed – it is necessary to set up an alternative escalation pathway that does not use the normal management hierarchy. Otherwise the power structure within the unit both investigates and judges every issue, without any appeal mechanism, which could potentially lead to the abuse of participants, and / or wasted research effort. It is accepted completely that the need for such a pathway to be used would be very rare, but it is not sufficient for the unit simply to claim that it would never arise.

The evidence that this standard is met would largely come from interviewing staff, discussing and clarifying the escalation pathways available and how staff are made aware of them. Although a unit may not have a formal controlled document dealing with this issue, 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.

IT01: Management of IT infrastructure

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 location, management and support of this central infrastructure through its life cycle, and the management of the physical environment in which that hardware is installed, normally a 'server room', including protection from intrusion, environmental threats such as fire, and system threats such as power loss.

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 the ECRIN standards.

We re-iterate that whatever the distribution of responsibilities for infrastructure management, it will be the responsibility of the data centre to have all relevant evidence *available* during an audit, even if it has not produced that evidence itself. 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.

IT01.01: Infrastructure location

Data storage and processing locations, including for backed up and mirrored data, must be known to the centre and the facilities used must meet all legal requirements for data protection.

If a centre manages its own servers, or they are housed within its own parent organisation, it should be straightforward to show compliance with this standard – the locations of the data will be known, and any local data centre will need to meet the local legal requirements. In particular, in Europe, from June 2018, it must be able to demonstrate compliance with the General Data Protection Regulation (GDPR) [9].

But if a centre uses a SaaS based CDMS, for some or all of its trials, it is important that the centre is very clear exactly *where* the data is located, and in particular, for a European unit, if it is in the European Economic Area, or EEA, or not. The same applies to back-up or 'mirrored' copies of the data. If not in the EEA, for instance a SaaS supplier uses an infrastructure in the US, the centre needs to know, and show, that the physical and logical security requirements in place are at least equal to those demanded by the EEA (e.g. using the provisions of the US-EU 'privacy shield'). This is because the GDPR imposes the same requirements on data controllers and data processors, if managing data about European citizens, wherever in the world that data is stored. The centre should also ensure, in these circumstances, that the patient information sheets a) clearly state that the data is stored outside the EEA, and b) provide an outline of how GDPR compliance has been assured (this level of transparency being itself a requirement under the GDPR).

The evidence in such cases would be the written assurances and explanations received from the SaaS suppliers guaranteeing GDPR compliance, plus example patient information sheets.

N.B. Because both institutions and national governments may change the rules regarding sensitive personal data and where it can be stored, and because the interpretation of those rules may also change, especially if challenged in the courts, ECRIN recommends that *EU based researchers do not store sensitive data outside of the EEA*, unless it is encrypted and the data centre itself controls the encryption process.

Encryption provided by the infrastructure or software vendor is still vulnerable to internal attack from the infrastructure or vendor staff, so only self-managed encryption is as secure as a local installation. But managing encryption (e.g. keeping a hierarchy of encryption keys secure, testing decryption mechanisms), especially over a long period, is not a trivial process. Although self-managed encryption does provide the freedom to store data everywhere, including on relatively cheap public clouds, the costs and effort involved should not be under-estimated.

IT01.02: Secured server room

Servers (and related equipment) must be housed within a dedicated locked room, or rooms, with unescorted access limited to specific roles, and with access arrangements known to the centre.

All servers, and related equipment such as SANs and routers, must be located in a locked room, or rooms, specifically allocated for that purpose.

The data 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). These would generally be a small subgroup of the IT staff, but it might include senior maintenance or security staff. The centre should also know the procedures for gaining access to the server rooms, and the possible reasons for access, logging arrangements etc., and be happy that those arrangements represented sufficient physical security.

If the centre manages its own server rooms maintaining compliance with the standard and reviewing access will be straightforward. If the servers are provided by a local hosting facility (e.g. the parent university or hospital's IT department) the centre should ensure that it can review the access procedures and access list on a regular basis (e.g. every year, or after a major perceived change in risk).

Even if a centre uses a SaaS based CDMS, it is important that it is satisfied that the data is housed in a secure physical environment. In general the centre will retain the overall responsibility for data management, delegated to it by the sponsor, including data security. In this case the centre should demand – from their SaaS supplier, who might in turn get the details from their infrastructure provider – the details of physical access control. Simply quoting ISO certification of an external hosting organisation is not sufficient – the centre needs to know the details of physical access control.

The servers may be physically distant from the data centre and the control of access may not easily allow escorted auditor entry, so physical inspection of server rooms is not essential in assessing the standard – though it can certainly be insightful if it is possible. More useful would be controlled documents and literature from the data centre and / or the server hosting facility, describing the exact location of data, the security measures in use, the access policies applied, the frequency of review and how the results of such reviews are communicated with the unit.

IT01.03: 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.

Evidence that this was the case would come from controlled documents, from a local or 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. Physical inspection of the server rooms may not be possible, and is not essential in assessing the standard.

If the centre manages its own server rooms than it is easy for the centre to provide the evidence described above. If the servers are provided by a local or external hosting facility (e.g. the parent university or hospital's IT department) the centre should assure itself that the power supply is secured, and that the mechanisms are tested, by obtaining the relevant information or logs from the hosting facility. Even if a centre uses SaaS, it is important that it is satisfied that the physical servers being used are protected from power loss in this way – by asking their SaaS supplier for confirmation that this is the case and for the supporting evidence.

UPS systems are usually designed only to last long enough for a managed shutdown. Though not a requirement of the ECRIN standards, it is therefore good practice to have an alternative power supply, e.g. from a local generator, available to allow continued functioning during a lengthy power loss.

IT01.04 Controlled temperature 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.

If the centre manages its own server rooms than it should be straightforward for the centre to provide the evidence for this, either by direct demonstration and / or reference to the specification of the server room and temperature and other environmental records. If the servers are provided by a local or external hosting facility (e.g. the parent university or hospital's IT department) the centre should assure itself that temperature is properly controlled, by obtaining the relevant details from the hosting facility. Even if a centre uses SaaS, it is important that it is satisfied that the physical servers being used are being managed in an appropriate environment – by asking their SaaS supplier for confirmation that this is the case and for the supporting evidence.

Though not currently part of the ECRIN standard, most modern dedicated server facilities, and almost all commercial hosting facilities, go well beyond temperature control and have full HVAC (heating, ventilation, and air conditioning) control systems installed. This is the recommended practice, and may become part of the ECRIN standards at a later date.

IT01.05: Fire and smoke alarms

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

Servers and related equipment must be protected from fire, hence this requirement. Although heat and smoke alarms are commonplace, the key requirement here is that they are monitored continuously, (the monitoring may be off-site) and tested periodically.

If the centre manages its own server rooms than the centre can provide evidence by direct demonstration and / or reference to the specification of the server room, test records etc. If the servers are provided by a local or external hosting facility it should still assure itself that an adequate fire alarm system is in place, by obtaining the relevant details from the hosting facility. Even if a centre uses SaaS, it is important that it is satisfied that the servers used have fire protection – by asking their SaaS supplier for confirmation that this is the case and for the supporting evidence.

Automatic fire suppression systems (e.g. inert gas or a misting system) are highly recommended, and would normally be part of any commercial hosting facility, but are not essential for compliance with this standard.

IT01.06: 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.

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.

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 in the context of any particular trial, and make funds available as required.

Providing automatic 24/7 server monitoring, with alerts being sent immediately to relevant personnel, allows failures to be picked up quickly in the evenings, over weekends and national holidays. It is a service that is often available from external hosting providers, though is not currently part of the standard.

The inclusion of software monitoring, in addition to the hardware monitoring provided by server monitoring systems, is also seen as good practice, though is not currently part of the standard. This can include 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, 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!

IT01.07: Server support and recovery from downtime

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. Although in recent years the widespread use of virtual machines, which allow a server image to be transferred quickly to another hardware 'base' if and when the need arises, has made managing server failure easier, it is still important that hardware failures are dealt with promptly, to minimise actual or potential down-time.

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. As usual, even though a centre may 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 using SaaS systems would also need to be satisfied that their SaaS suppliers had similar arrangements in place, with their infrastructure providers.

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), the type of functionality being supported (e.g. an on-line randomisation service versus a CDMS, development systems versus production) as well as the degree of redundancy built into the system (e.g. using clustered servers, mirroring, or log shipping).

It is therefore important that the centre has considered the down-time that would be acceptable in different systems, and set up (or ensured that others have set up) mechanisms that allow those down-time requirements to be met.

Evidence that the standard has been reached include the documents and / or 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.

IT01.08: 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, regular or ad hoc 'snapshots' of server configurations may be taken.

If snapshots are taken infrequently, e.g. at initial build and before and after major changes, that 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.

IT01.09: Server software 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 — with the options including 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:

- controlled documents, and / or documentation from IT infrastructure hosts, detailing how risks associated with patches / updates are assessed, and how any changes are made, as safely as possible;
- 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: 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).

IT02.01: 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), ensuring that security measures are not specified and implemented as a one-off activity, but are periodically reviewed in the context of changing threats and risks. Reviews will necessarily include management and budgetary issues as well as technical discussions, and should therefore inform and / or involve senior management. The term is borrowed from the ISO27001 standard on Information Security Management [10], though there is *no* expectation that that the centre or its parent organisation has obtained or is seeking full ISO27001 certification. 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.

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 certification. 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.

IT02.02: 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. Note that within the EU the introduction of the General Data Protection regulation (GDPR) may reduce but not remove differences between countries, because large parts of the regulation dealing with sensitive data have been left to the discretion of national legislatures to implement.

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.

A member of staff, usually within the parent organisation rather than the data centre itself unless the data centre is a separate legal entity, will be identified as the organisation's Data Protection Officer, as this is now a mandated requirement under the GDPR for all data controllers or data processors. That person should be available to provide local support and guidance to the data centre staff, and may be involved in providing training input. Given the specialist regulatory requirements surrounding trial data, however, it can also be useful to identify one or more individuals within the data centre who can also develop expertise in this area, liaising with the institution's DPO as necessary.

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;
- identification of the institution's Data protection Officer, plus one or more staff identified within the units as having special expertise in data protection legislation;
- records of training concerned with data protection (some level of training will be required for all IT / DM staff); the expectation would be that the introduction of the GDPR would trigger new training input.
- interviews with staff to check understanding of data protection requirements and discuss how the systems work in practice.

IT02.03: 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.

IT02.04: 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. In 2014 the SSL algorithm sometimes used for encryption within https was shown to be vulnerable to the POODLE man-in-the-middle attack. The current recommendation is to only use the more recent TLS algorithm on all traffic interacting with web servers (as per various articles in the computer press [e.g. 11], it should be the most recent TLS 1.2, properly configured). 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 1.2 based encryption (so far as client browsers and the eRDC system itself allow).

IT02.05: 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.

IT02.06: 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.

IT02.07: 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.

IT03: 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 IT02).

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.

IT03.01: 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.

IT03.02 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.

Traditionally a process is established that enforces 'strong' passwords, with a variety of rules defining what 'strong' means: e.g. length > 8, at least one upper and lower case letter, at least one digit, one punctuation character etc., and a change after a fixed period (e.g. 90 days). Increasingly, however, some security managers recommend much longer pass-*phrases*, typically 12-20 characters, but easier to remember and retained for longer periods, e.g. a year. There is little empirical evidence to say which approach is best, so either is acceptable.

In some centres biometric devices or personal cards may be used, instead of or in combination with passwords, a process known as '2-factor authentication' (in fact this is now demanded in some countries).

Evidence that this standard was met would come from:

- controlled documents detailing the management policies for network log-in;
- a description of current users group and how their access rights are distributed.
- 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.

IT03.03: 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.

IT03.04: Remote access (not using a browser)

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.

IT03.05: Access control management

All systems that directly support clinical trials activity and that require 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 all systems 'that require access controls', starting with the initial log-in to the centre's / parent organisation's network for internal staff, but including in particular access to the CDMS for both internal and remote eRDC staff, and any other systems (e.g. treatment allocation, coding, pharmacovigilance) that directly support clinical trials activity. In general remote site staff should 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 access control for centre and site staff, across the different systems
- demonstration of the access control system, especially for the CDMS.

IT03.06: 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 IT03.05) 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.

IT03.07: 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. This could risk data integrity, especially if the leaving was not voluntary. A variety of mechanisms are available to try and reduce the time lag between someone leaving and their access being revoked. None are 100% reliable, but using two or more together can reduce the risk of unauthorised access by ex-staff. These include:

- Monitoring of access, to identify staff who have not logged into the system for some time
- Asking monitors and other site visitors to check the access required at each visit.
- Regular reminders to site senior staff to let the data centre know of staff changes
- Coupling any requests for new access at a site with a check on the existing accesses required

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.

IT04: 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 (IT04.01) though the rest of the standards are focused on IT disaster recovery.

IT04.01: 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.

IT04.02: 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 reentering 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 clarify the details of backup and restore arrangements, so that 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 the information they need to make that judgement.

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.

IT04.03: 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.

IT04.04: 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.

IT04.05: 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 IT04.02, IT04.03) 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.

IT04.06: 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.

IT05: 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." [12]

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 DM01.

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 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 (IT05.01).
- 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 (IT05.02).
- 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 (IT05.02).
- 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 (IT05.03).

- 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 (IT05.04).
- 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 (IT05.05).
- 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 develop expertise in operating the system. 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 (IT05.06). The risk assessment and any subsequent revalidation plan, together with the results of that revalidation and the resulting decision, should all be recorded. (IT05.07).
- 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 (IT05.08, IT05.09).

IT05.01: 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 IT05.02).
- The method(s) used for risk assessment (see IT05.02).

- 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 **all 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 IT05.02).

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 IT05.02.

The evidence would be the relevant controlled documents.

IT05.02: 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 IT05.01) 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.

One of the key factors determining the risk assessment is the *software type*. A classification scheme in common use (from GAMP 5, [13]) 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

Although this classification is often used to allocate different validation regimes to systems, it is a very blunt instrument, and many other factors need to be taken into account. Some of these 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.

IT05.03: 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 IT05.06) 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 IT05.02) 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.

IT05.04: 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 the V–model approach to validation and the associated terminology from GAMP 4 [14]: i.e. initial, operational and performance qualification, and use these to structure their validation processes. The three types of qualification, together with the equivalent terms from GAMP 5, are defined below:

- *IQ, Installation Qualification (= Configuration testing in GAMP5):* 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 (= Functional testing in GAMP5):* 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 (= Requirements testing in GAMP5):* is the process of checking that the system, over a range of 'real world' conditions, continues to perform as required.

  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.

IT05.05: Validation Summaries

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

As well as the detailed results (see IT05.04) 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.

IT05.06: 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 IT05.07). The evidence that this standard has been met would be provided by the controlled documents themselves, together with the associated proformas.

IT05.07: 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 IT05.06 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 reassessment / 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 revalidation 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.

IT05.08: Validation of extracted data

Extracted data, however formatted, and the underlying data extraction processes, should be assessed using a risk based approach to decide upon the level of validation needed to ensure accurate extraction.

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 system validation 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.

If, in addition, an extraction process involves locally constructed processing of some kind then that processing will also need validation, and/or the data in the extracted set or report will need to be compared with the original data in the CDMA to check that they match.

This applies most obviously to the extraction process that generates the datasets for analysis. Although the extraction process would be expected to be the same for all trials on the same system, the volume and / or type of data in any specific trial should be considered to see if more detailed testing might be necessary in any particular case, especially for the first trials extracted from the system.

A common practice is to ensure that the extraction process for the analysis datasets generated the correct numbers of subjects, distributed correctly between sites, and that data from the first and last participants in the trial, and possibly in each site, appeared to have been correctly extracted.

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.

IT05.09: 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 IT05.08.

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.

IT06: 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 IT05, 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 (IT06.01 and IT06.02) and resourcing (IT06.03). In addition, a number of suggestions for 'good practice' in software development are provided.

IT06.01: 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 IT06.02).

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 IT06.02 — 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.

IT06.02: 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 IT06.01 — 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.

IT06.03: 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, IT06.03 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: Data Management Planning

This section is unusual in that it only consists of a single standard, but it is one that might be considered fundamental to data management. It requires not just that data management plans are used, but that there is also a suitable local template available for constructing such plans. The reason for asking for a template is that it shows the data centre is co-ordinating the use of data management plans and has ensured that they meet its own requirements, rather than leaving it to each individual trial team to create a document with an ad hoc structure.

DM01.01: Use of data management plans and template

Each study should have a data management plan section within its Trial Master File, describing the study specific elements of data management, structured using a locally created template.

Although different SOPs and other controlled documents will describe the *generic* procedures for various aspects of data management, there is also a need to describe the *study specific* aspects of that management. This is supported by a data management plan (DMP), very often a distinct document, but as a minimum a defined section within the trial master file.

One function of the DMP is to provide a record of the study specifics of data management, for example the systems used for the databases, the exact locations of files, or the versions of coding systems and data collection instruments used, with this data being added throughout the life of the study. But it also provides an important mechanism for planning, and therefore resourcing, various aspect of data management: the nature of any data transfers, the use of data standards, the methods to be used for data cleaning, the way in which the analysis files will be constructed, the plans for long-term storage, etc. It is therefore important to have a locally defined template for the DMP, to structure this planning activity and ensure that all important aspects are covered.

There are two main reasons why data management plans have become more critical in recent years:

- The fact that trials are often more complex and involve data collection from a variety of sources, especially in translational research where specialist laboratories may be used to carry out bio-assays or measure genomic expression. The need to plan how such data should be aggregated with traditional clinical site data has therefore increased.

- The increasing demand from funders and journal editors for researchers to make their datasets available to others, with the data itself becoming an important product of research rather than merely an intermediate step in the production of a published paper. This demands greater planning for long-term data storage, including possible data preparation steps (e.g. de-identification) and transfer to a dedicated repository.

The template needs to contain sections covering the entire life of the data. The list given below gives some commonly used headings:

- Timescales of data collection and analysis. Target accrual figures and expected approximate volumes of data.

- The different sources of the data (clinical sites, the participants themselves, machine generated data, electronic health records, images etc.) and the nature of the source documentation in each case. Instances where there will be no obvious source documentation need to be highlighted.
- The use of standards for data items (e.g. CDISC CDASH) and data collection instruments (e.g. questionnaires, standardised tests). The versions being used should be clearly indicated.
- Study specific training and guidance materials for data entry and / or management.
- The exact systems, including versions, used for the collection and storage of the data, including their physical location. This may reference other more detailed documents.
- Any variations from the normal roles, responsibilities and processes described by generic SOPs, in the construction of systems and in data management.
- Storage locations of the relevant data and metadata files (within the file system).
- The study specific rules for any 'self-evident correction' procedures.
- How data quality will be assured – e.g. by using data validation checks on data entry, double data entry or visual monitoring of core items, central statistical monitoring to identify outliers, etc. The relative importance and role of each method.
- The nature of data transfers to and from the centre and the mechanisms for merging data from different sources.
- The systems to be used for any coding of data, including versions and any study specific coding rules.
- How the analysis datasets will be constructed / extracted, and the differences in definition, if any, between the analysis data and the data as collected.
- The expected location and duration of storage of data in the long term, and the different responsibilities of the organisations involved.
- Measures to prepare the data for possible secondary re-use, assuming study participants' consent allows this, including planned use of de-identification techniques, organisations involved etc.

It is stressed that this is only a sample list and that any centre should construct their own DMP template, reflecting its situation and the types of studies that it manages. The existence of a template does not mean that every DMP must contain exactly the same content, but it does mean that every DMP should at least consider the same core set of issues, and expand the relevant sections as required, adding extra sections if necessary.

Evidence that the standard was met would include the existence of the template and demonstration of its use in at least two studies.

DM02: CDMAs – Design, Development and Validation

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 electronic CRFs (eCRFs) used for data entry and the trial specific data validation checks, skipping logic and derivations that those eCRFs contain. As depicted in figure 1, CDMAs are built upon an underlying clinical database management system (CDMS), such as Macro, OpenClinica, RedCap, TrialMaster or Rave.

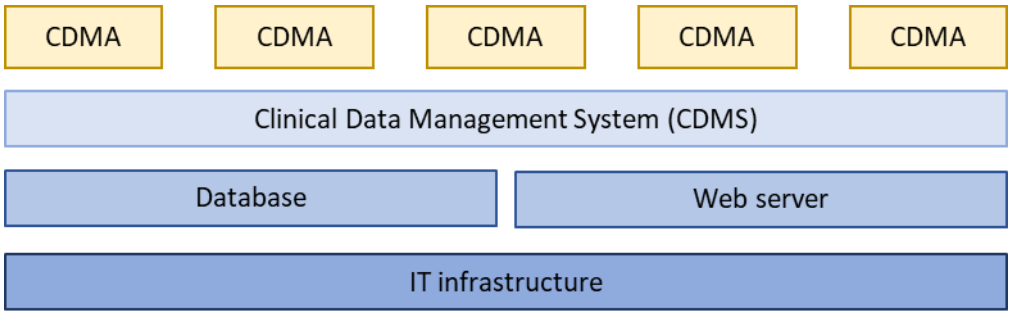


Figure 1: The clinical trial technology 'stack'. CDMAs are built as trial specific applications on top of a generic CDMS. The CDMS itself uses database and access services (usually via a web server), themselves assembled on the underlying IT infrastructure.

The standards in this section deal with how CDMAs and the eCRFs within them are constructed and then validated to ensure that they are ready for use. They therefore span the development of a CDMA from initial conception through to the final sign off of a production ready system. This cross-disciplinary process is depicted as a flow chart in figure 2, and necessarily falls into two distinct stages:

- The design phase, in which the CDMA is specified, almost always in a series of iterative steps, and constructed (very often at the same time, as successive prototype versions, but if not than at the end of the specification process). The final design and specification should be formally approved by the key staff involved.

- The validation phase, which involves detailed, systematic testing of the newly constructed CDMA against the approved specification. After the validation process is finished, and any errors or omissions are corrected, the system can be formally signed off as ready for use.

The whole process needs to be controlled by SOPs and policies, and supported as necessary by forms, spreadsheet templates and – where necessary – more detailed system specific technical guidance. DM02.01 therefore requires that such policies exist.

The basis for the CDMA for any particular trial is the trial protocol, and the process of ensuring that the CDMA matches the requirements of the protocol, but does not collect unnecessary data, is considered in DM02.02.

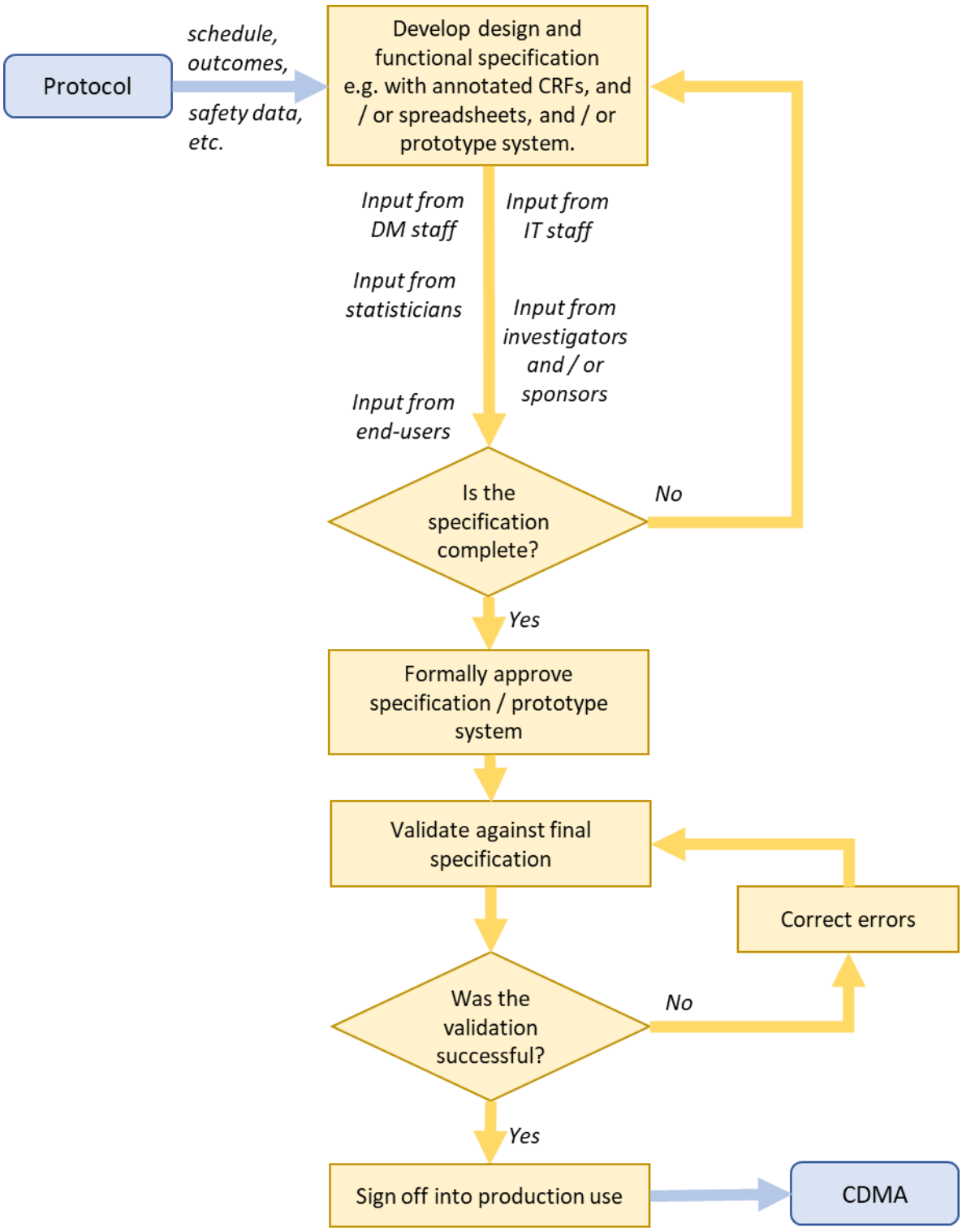


Figure 2: The workflow for CDMA development

The requirement for a full functional specification is the subject of DM02.03. Developing such a specification demands input from different professional groups. One of these groups is the system's end-users, whose specific input is considered in DM02.04. The process also demands a particular development environment, as described in DM02.05. The specification and construction process should end with a formal approval process, as required by DM02.06.

The validation phase of CDMA development will be based on the functional specification, as required by DM02.07. It will lead to a set of detailed documentation recording the validation that has taken place and ultimately to a summary validation report, that should be signed off to indicate that the system is ready for production use (DM02.08).

DM02.01: CDMA development and validation policies

Controlled documents covering the development of CDMAs and CRFs, including their validation, should be in place.

Developing CDMAs and the CRFs within them must be done using defined procedures, with tasks and responsibilities clearly delineated and with approval processes clearly described. Supporting quality documents (e.g. forms, templates and checklists) should be available as required.

The policies could be integrated into a single SOP and related documents, or split into different SOPs dealing with different aspects or stages of the process – the details are not important as long as all aspects of the development process are covered. It is usually a good idea to keep system specific technical details in separate guidance documents, partly because CDMS systems change between versions, partly because a unit may have, or introduce, more than one CDMS. This allows the SOPs to be kept relatively 'high-level', and therefore less likely to need frequent revision.

The evidence that the standard has been met would be the relevant controlled documents, together with CDMA specific documents that showed that the policies had been applied in at least two instances.

DM02.02: The CDMA and the protocol

Processes exist to ensure the CDMA 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 within the CDMA, with the CRFs capturing sufficient data to support the analysis of outcome and safety measures described in the protocol.

Making the CDMA specification a cross-disciplinary process is an important way of ensuring this happens, with input from the investigator(s) and statistician(s) particularly important in this respect. One approach is to first use the protocol to specify the data points that the statisticians will need to carry out the required analyses, and then use the annotated protocol, or even a formal set of analysis data requirements, to drive the CRF specification.

In practice, the danger is less often that insufficient data is collected, but rather that *too much data* is requested, with data points included not because they are necessary to answer the protocol's questions, but because the data 'might possibly be useful one day', or even because they are part of a eCRF re-used from an earlier, similar trial.

Collecting too much data runs counter to *data minimisation*, a principle emphasised in the General Data Protection Regulations (GDPR) of the EU: "Personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which those data are processed." (GDPR Rec.39; Art. 5(1)(c)) [9]. At least within the EU, collecting unnecessary data is therefore illegal as well as unethical.

Data may of course be deliberately collected for purposes beyond answering the immediate research question – for instance to ensure a disease specific 'core dataset' is available, that could be integrated with similar datasets from other sources in the future. But when that it is the case it should be explicitly mentioned within the protocol and the participant information sheets, so that the consent for trial participation and data processing is fully informed. (If the exact details of post-trial data use are unknown, a separate consent for data sharing should also be obtained [15]).

One way of reducing unnecessary data is to mandate a detailed review of the CDMA's data points by the study statistician, as the chief 'consumer' of that data. Such a review can result in significantly streamlined data sets, leading to reduced CDMA development and validation times.

Whatever the detailed mechanisms used, 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. The more this aspect of CRF design is made explicit the easier it will be to demonstrate the standard.

Note that ECRIN auditors are *not* expected to assess the *outcomes* of this process, i.e. to assess CRFs against their source protocol, partly because in any particular case the sponsor or investigator will usually have the final say about CRF design, partly because of lack of time. 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.

DM02.03: Creating a full functional specification

A CDMA design and full functional specification should exist identifying each data item on each CRF (including field names, types, units, data checking logic, conditional skipping, and derivation logic).

The CDMA development is almost always a multi-disciplinary, iterative process, as shown in figure 2. The data management and IT staff may be most concerned with the detailed specification and design on a day to day basis, but input from statisticians and investigators, and others on the trial management team, is also important, as is input from end-users (see DM02.05).

There is a huge variation between trials units in how such a specification is generated. Many start the process with 'annotated CRFs', even for studies that will be entirely eRDC, because they are easier for most people to assess and discuss. More detailed spreadsheets are often introduced in the later design stages, especially to describe the detailed data validation logic, because they give a more precise formulation of the database.

Perhaps the easiest way of developing a system specification is to build it, as a series of prototype CDMAs. This is especially useful if the system can generate its own metadata, as most now can, for more detailed examination when required (e.g. of data validation and skipping logic). That way all involved in the design process can see the CDMA taking shape, including the visual design elements like layout, colours and prompts, and can comment upon it, and the detailed specification is guaranteed to match the current state of the CDMA database.

For the avoidance of doubt, building a CDMA in this way is not an 'agile' development strategy, other than in the relatively minor sense that the visual layout of elements can be more easily negotiated. The users' specification of the system is fixed and is represented by the protocol and the context in which the CDMA will be delivered. Gradually building the CDMA simply offers an easier way for people to monitor development and check the specification is being interpreted correctly, in a series of incremental steps rather than all at once at the end. The cross-disciplinary approval process (see DM02.06) always anchors the system against the protocol's requirements.

If the prototyping approach is not used, and the specification only exists on paper, then at the end of the specification phase the system will need to be built from that specification, initially to allow end-users to examine it, and then in order that it can be validated. Conversely, if prototyping has been used, then the specification can simply be generated from the final prototype. In either case the specification will include details of the data collection schedule, data points, skipping, derivation and data validation logic, as well as additional study specific aspects like support for coding, e-signatures, source document verification, or email alerts. The specification and the system are then used as the basis of the validation that makes up the second stage of development.

Evidence that the standard has been met would come from:

- inspection of at least two detailed CDMA functional specifications;
- discussion with staff to clarify how the specifications are developed;
- relevant sections of controlled documents.

DM02.04: Isolation of CDMAs in development

CDMAs in development should be isolated from, and clearly differentiated from, the CDMAs used in production.

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, including IT staff with elevated privileges, inadvertently confusing development and production systems.

This means that a process needs to exist that exports the completed CDMA from the development environment and imports it into the production system, as a 'study definition' of one or multiple files. It also means that development systems should be clearly marked as such on screen (e.g. by the use of different colours and labels).

In an age of virtual or containerised machines, 'isolated from' means logically isolated rather than necessarily on different physical hardware. That means distinct URLs for web based systems, distinct connection strings and other access mechanisms for database servers, and different users and access control systems on the different types of server.

In most cases, because virtual machines are relatively cheap, it is easier to have at least a pair of virtual database servers, one development and one production, and a pair of virtual CDMS servers (usually web servers) again one being development and the other production.

That makes the distinction clear and relatively easy to manage – for instance a production server may require more frequent backups than a development server. It also makes it much easier to keep the production system simple and 'clean'. That in turn gives less room for human error, clearer access control, and less suspicion that the validation status of CDMAs might be compromised by some unknown side effect from one of the other systems (e.g. a resetting of a server wide attribute) however unlikely that might be in practice. Development servers, in contrast, may and often do accumulate multiple versions of systems, or additional reporting, administrative and test systems.

It *might* be possible to have different development and production versions of the same database / system on the same server, but the procedures and processes would have to be very clear to show how isolation was maintained between the two types of usage, and how the potential for human error and the risk to data integrity was minimised.

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 logically isolated from production systems;
- inspection of relevant controlled documents.

DM02.05: Input into CDMA development by end users

Procedures are in place to secure feedback from selected end-users, on the practicality and ease of use of the CDMA, and to decide when and how such feedback will be sought.

'End-users' are staff outside the trials unit who, in an eRDC context, will have to use and input data into the system. In general it would not be realistic to expect such users to carry out detailed, systematic testing of the CDMA, but it is reasonable to expect a sample of end-users to provide some feedback on the system's ease of use, for instance how easy it is to understand or navigate, on the practicality of providing the requested data, and on the ease of raising queries etc.

How much feedback will be necessary, and from whom, will be very much a function of the particular CDMA and the sites in which it will be used. A relatively simple CMA, deployed only to sites that already have experience of very similar trials, will probably need little or no additional feedback from end users. On the other hand a CDMA that includes novel features or patterns of data collection would benefit from end-user feedback, ideally from different sites with different levels of prior experience. If new sites are being used for a trial, especially if they are from a different country or language group, then user feedback can be very informative in clarifying how the eCRFs will be interpreted and in identifying potential problems. (User feedback could also be integrated with initial user training.)

There is a question about *when* such feedback should be sought. The feedback can only be obtained at or near the end of the design and CDMA construction process, because there has to be a system to demonstrate – even if using dummy data. The feedback obtained often relates to the system's design – layout, labels and prompts, colours, ordering of items, etc. Although data item errors may be found and changes may be requested (e.g. in the ranges allowed for lab data), the type of systematic, detailed testing described in DM02.07 is neither appropriate nor realistic, and could not in any case be supervised.

End-user feedback should therefore occur near the end of the design and specification phase, after the CDMA system has been constructed, but before the final specification has been approved and formally validated. It should be suitably structured (e.g. by the use of proformas) rather than simply using informal emails.

If such feedback is sought instead *after* approval of the final specification, as it sometimes is, and it results in requests for changes, then these will need to be handled by the CDMA change management mechanism, which will be both more complex and less safe than handling such changes within the final phase of the specification phase.

Evidence that the standard had been met would come from

- explanation by centre staff of how the level of user feedback is decided for any particular CDMA;
- how the feedback is organised and gathered, plus
- inspection of actual user feedback.

DM02.06: Cross-disciplinary approval of the functional specification

The CDMA's design and functional specifications are signed off and dated by signatories representing a cross-disciplinary team.

Once the CDMA's final specification is assembled it will need to be formally approved and signed off by the key individuals involved with the trial. The initial version of the CDMA should have been constructed at this stage, partly because that allows end-users to also provide their input (see DM02.05), and partly because the final specification can be generated from 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), those analysing the data (i.e. the trial statistician), and those sponsoring the trial, (a sponsor representative and / or chief investigator) sign off the functional specification. Others that are usefully included in the sign off are those building the CDMA (i.e. the IT staff), and a quality assurance or operational manager.

The cross-disciplinary approval does not necessarily mean that all parties will check the specification for the same things. In most cases, for instance, it would be unreasonable to expect the chief investigator to look through every data item in detail, but they should at least be satisfied that the main outcome and safety measures are properly covered. As described in DM02.02, statisticians may be asked to check there is no unnecessary data being collected, as well as confirming that the obtained data will be fit for their analysis. A data manager will probably check the eCRFs in detail, and confirm that feedback has been confirmed from end-users, while the unit's quality manager may also check for adherence to unit policies on CDMA design, use of coding systems etc. Some of this assessment can be done by inspecting the system itself, but some of it will require checking of more detailed documents.

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.

DM02.07: CDMA validation against the functional specification

Systematic, detailed testing is carried out against the functional specification for each CDMA before deployment to the production environment.

The newly constructed version 1.0 of the system needs to be subject to a detailed validation exercise, against the approved functional specification, to check that it really does match that specification and is therefore fit for purpose.

As well as checking the more obvious features like data type, captions, tab order and code lists for each data item, this means going systematically through the system to check the skipping logic, derivation logic, and each of the data validation checks. Any problems that are found at this stage are bugs rather than design faults, because the design has been fixed – CDMA validation, like any other, has to have a fixed target!

This can, admittedly, be a rather mechanical exercise, and may therefore be carried out by relatively junior staff. That is OK as long as the specification and any additional instructions are clear and there is sufficient supervision. Validation should *not* be carried out by anyone that constructed the system, usually the IT staff, because any misinterpretations of the original requirements will simply be repeated. In some units the data managers for the study are asked to carry out the CDMA validation exercise. This has the advantage that if they were not very

familiar with the details of the system before they certainly will be after the exercise is completed, but it may not be the best use of skilled staff time.

In some cases checks can be skipped if they have already been covered earlier in the validation exercise. For example, a test that a date entered is not after the current day is a condition that may be applied to almost every date item. If the date questions have been copied from a common precursor that included this test it does not necessarily need to be checked on every occasion it is used. Similarly, an eCRF representing a questionnaire, that has been used previously and imported and tested in another study, may not need such a detailed checking as a completely novel eCRF.

Some CDMAs may require additional testing for particular functionality (like coding, or message triggering) that is not found in other study applications. The testing should also be included in the validation because these features should be described within the specification.

CDMA Validation should be based on a risk assessment and identification of the elements that need to be tested. One would expect the 'default' position, for instance that all elements and all logic in the system should be tested, other than eCRFs drawn from a validated library, and / or that only the first two instances of each check are tested, to be described in the SOP dealing with the validation process. Study specific decisions about validation need to be described and justified to the extent that they vary from this default position.

An alternative approach to CDMA validation involves completing dummy paper CRFs, 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 also means that the extraction / reporting functions are tested as well. Unfortunately, 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 described above.

The bugs discovered and their resolution should be documented. At the end of the exercise all issues should be resolved, so that it can be shown that the system meets its functional specification. If the process generates requests for *design* changes, i.e. the specification itself needs to change, a full re-approval process is not required but the change management process needs to be used to assess the risks associated with the change (usually less at this stage because there is no real data in the system) and thus the additional validation required (see DM02).

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. Evidence that the standard had been met would come from

- examples of completed test documentation, (for at least two trials, dated and with the staff involved identified)
- discussion with staff to clarify how the validation process was organised in practice.

DM02.08: CDMA final sign off into production

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

Once CDMA validation has been successfully completed a summary validation report needs to be created and signed. The signatures may represent a small cross-disciplinary team, but more often it is the trial or operations manager who supervised the validation process, even if it was actually done by members of their staff, who signs to say that they are satisfied that the system is validated.

In most cases a single sign off will cover the whole CDMA, but some centres may arrange to have each CRF signed off separately.

It is usual, and useful, to also include any intermediate validation results, i.e. the list of issues found on initial testing. The final summary report should confirm that all those issues have been resolved.

Evidence that the standard has been met will be appropriately signed and dated documents confirming that a CDMA meets its specification and can be used as a production system.

Standards and re-use of items and forms as 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* [16]*):* 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.

DM03: CDMAs – Change management

Once the specification for a CDMA has been approved, and thus fixed, any further changes to that specification will need to be considered with a formal change management process. If the process described in DM02 is followed, such changes should only occur once the system has been validated against its specification, and signed into production, and the change management process is designed to ensure that the system retains its validation status.

The change management required follows the general principles outlined in IT05 (standards IT05.06 and IT05.07 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 DM03.07).

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 revalidation 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: 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 DM04.04).

DM04.01: 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.

DM04.02: 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.

DM04.03: Isolation of training eCRFs

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

Users need to have the opportunity to train on CDMAs, generally using dummy or test data, but 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 CDMA instance and / or data store for training purposes, sometimes by setting up a dummy ‘training site’ within the production system. The latter is easier to manage and ensures that the training system will exactly match the production system’s definition, but it has the disadvantage that all data from the training ‘site’ must then be excluded from the analysis dataset (during or after the extraction process).

Whichever approach is used the access to the training system or site should be removed as soon as access to the real site is given, to remove the possibility of data being entered into the wrong system by mistake.

If a user, intentionally or accidentally, did have access to *both* the training *and* a real site for some reason then the possibility of inputting data into the wrong system can be reduced by clearly distinguishing the training and production environments. If a different CDMA instance is used for training this is much easier - distinct colours, banner labels and images can be used. If a dummy site is used within the production system than it should still be possible to use obviously false site names and codes, and participant identifiers, and set other variables to clearly show that the data is test data only. Some systems allow for site dependent visibility of on-screen features or labels, again allowing the distinction to be clearer.

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

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.
- Demonstration of the differences between production and training eCRFs.

DM04.04: 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 DM04.05), 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.

DM04.05: 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.

DM04.06: 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 DM03.06).

DM04.07: 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 copies of what are often known as 'delegate log's, covering the staff for each site in the trial. The Principal Investigator at the site has the responsibility for creating and maintaining the log and ensuring that staff are suitably qualified for their role, but the centre should have a copy of the resultant list, of named site staff and their roles within the study. How the logs are obtained and then maintained up-to-date 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.

DM04.08: 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.

DM04.09: 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.

DM05: Data Entry and Processing

The standards in this section deal with data entry into the CDMA. Most modern CDMS 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.

DM05.01: 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.

DM05.02: Production of interim CRFs

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 are paper representations of the data capture screens.

They 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.

DM05.03: 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.

DM05.04: 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.

DM05.05: 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.

DM05.06: 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 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." [17]

Modern CDMSs normally support such an audit trail.

The ECRIN audit trail requirements do not include an explicit 'reason for change' (RFC) as a mandatory data item. Although almost all CDMSs support this feature, and most data centres make use of it, some are less convinced of the utility and accuracy of the data recorded.

The GCP requirement (section 4.9.3) is that changes should be explained "if necessary". The argument can be made that if the associated query (or SDV) process is documented in the system than it is already clear why a value has been changed, and a reason will therefore only be required if a change occurs in some other context. Whatever the unit's decision about recording RFCs, it should be decided, documented and disseminated systematically, for example with a list of allowed RFCs included in the data management plan.

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

DM05.07: 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.

DM06: 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 preconfigured 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.

DM06.01: 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.

DM06.02: 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.

DM06.03: 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 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.

DM06.04: 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.

DM06.05: 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.

DM06.06: 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.

DM06.07: 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 07/11/208, 07/11/218 or 07/11/20018 for 07/11/2018, (albeit with an assumption that it could not be 07/11/2008) 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.
- The procedure should include a mechanism that allows the investigators at the site to endorse any SECs made, e.g. sending each site's final list of SECs back to the sites at the end of the study for inspection.

Note that the GCP requirement is for changes to be endorsed (i.e. approved) rather than checked. A pre-defined and pre-agreed set of pre-conditions for SECs, as described above, is therefore the most important component of SEC management.

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 and time consuming, but they are usually much quicker and cheaper to resolve in an eRDC study.

SECs are therefore discouraged in an eRDC context. An argument is sometimes made that before they can be coded, composite adverse event terms need to be split by applying SECs (e.g. 'vomiting and diarrhoea' turned into two distinct reports), and that this applies to eRDC systems as much as to paper based trials. Even here, however, good training of site staff, and prompt querying of problematic data entry, should be able to resolve the issue without recourse to SECs.

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.

DM06.08: 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 may vary from one study to another, because they should be part of a risk-based approach to overall quality management, as determined by the sponsor (in accordance with GCP 5.0), usually in conjunction with the data centre, as the

sponsor's main data management 'contractor'. The strategy is therefore likely to be described in a study specific data management plan rather than in a generic controlled document such as an SOP.

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.

DM06.09: 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.

DM06.10: 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.

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.

DM06.11: Supporting central statistical monitoring

The centre can generate reports to support central statistical monitoring

One of the key components 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. Centres should therefore be able to generate reports detailing query rates, missing or late data, and where appropriate additional study specific indicators of problems during data entry, on a site by site basis.

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.

DM06.12: 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.

DM07: Managing Data Transfers

This section deals with moving data files into or out of the data centre, normally to or from a different organisation. Transferring data into the centre is referred to here as *data import*, while transferring data out is described here as *data export*.

Data import occurs when centres receive bulk data, for instance from 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). As depicted in Figure 3, It consists of two or three processes:

i_1: Import of the data from the source organisation to the data centre.

i_2: (Possibly) further processing of the data to enable merger with existing data and / or centre systems.

i_3: Merger with other study data, usually by aggregating it with the analysis data set, though it may be by uploaded to the CDMS, followed by extraction of the combined data.

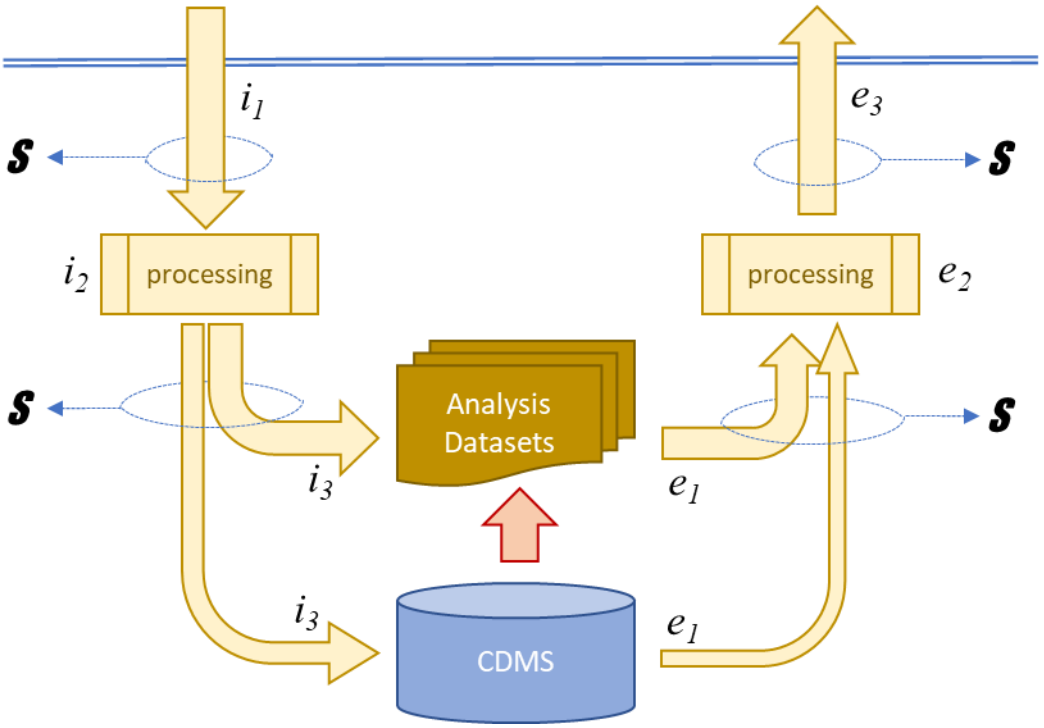


Figure 3: Data import and export. The processing stages, i_2 and e_2, may not be required. The **S** denotes a point at which files should be stored for audit purposes.

Data export can occur in the context of a collaboration or meta-analysis, or sending data to a statistician or investigator based elsewhere for analysis or review. It includes the process of sending data to an external sponsor. It also consists of two, or more often three processes:

e_1: An initial extraction of data, usually specific to a single study, from the CDMS or from the already extracted analysis dataset.

e_2: (Possibly) further processing /formatting of the data to match the recipient's requirements.

e_3: Export of the data from the data centre to the recipient.

Data import and export are therefore mirror images of each other, and their requirements are very similar.

DM07.01: Data Transfer Procedures

Controlled documents dealing with the transfer of data from and to 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, in either direction, including the documentation required.

In practice a transfer process, especially if repeated, will often need more detailed procedural guidance if consistency is to be maintained. This is especially the case if the data needs transforming in some way, either before merger, or after extraction, i.e. the processing steps i_2 or e_2 in the description above are used. Generic procedures will therefore very often be supplemented by study specific procedures, which should be described or referenced within the study's data management plan.

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. Similarly, the decisions about which laboratories and other facilities to use, and thus receive data from, will also be the sponsor's. In both cases, however, the centre needs to have procedures in place to ensure that it can transfer the data securely and accurately, and record the entire process, ensuring that it fully discharges its operational responsibility for the transfer process.

The evidence that the standard had been met would come from the controlled documents themselves and the documentation associated with specific transfers.

DM07.02: Records of Transfers

Details of any specific data transfer should be logged, to maintain a complete record of how, when and why data has been transferred to and from the centre.

Once the transfer takes place it needs to be recorded. Each data transfer log record should contain, as a minimum,

- Details of the recipient (for data export) or sender (for data import)
- Reason(s) for the transfer
- A listing of the files received or sent (paths to current storage locations, see DM07.03)
- A summary description of the data, if not obvious from the above
- The transfer method(s) used
- The nature of any encryption used
- The date(s) of transfer and the personnel involved

For data exports, it is also useful to request and then record confirmation from the recipient that the data has arrived safely and meets their requirements.

If the data is processed in some way as part of the transfer process, then:

- Any scripts used for processing should also be retained (or referenced).
- The details of the initial extraction (for exports), or final aggregation or upload (for imports), should also be recorded, e.g. dates, files involved, personnel involved.

This data, coupled with the retained datasets described in DM07.03, should provide a complete record of all data transfers. It can be maintained separately for each study, but is probably more easily organised as a centre-wide record. Each study's TMF should include a copy of the relevant transfer records, or a reference to their location.

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

DM07.03: Retention of intermediate and transferred files

Copies of all files received or created in any transfer process should be retained within a read only environment and be available for audit / reconstruction purposes.

In an import process the datasets originally received, plus those datasets after any required processing, (i.e. the datasets as merged with the existing data), need to be retained. In an export process the datasets as originally extracted, plus those datasets after any required processing, (i.e. the datasets as transferred out of the unit), will also need to be retained. The 'retention points' are marked with an ***S*** in Figure 3. In that way the centre retains a clear audit trail of the data at each stage of any transfer process, which can be checked if necessary, and which complements the logging of the transfer and the description of any processing (see DM07.02).

To prevent any possibility, accidentally or otherwise, of modification of the data, the datasets should be kept in a 'read only environment'. In practice this usually means within a folder where only a few staff (usually IT staff, because they are seen as having no direct stake in the study or its results) can insert or modify files. For all other staff the folder and the files within it are set as read only, and are therefore protected from amendment.

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

DM07.04: Encryption of Individual Data

Any file(s) transferred 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). Because there is often difficulty in distinguishing patient identifying data from other data relating to individuals (see IT02.07) the requirement is that all individual participant data that is transferred is encrypted, and not just that which contains direct identifiers (i.e. fields that can uniquely identify someone, alone or in combination with other data).

When data is exported the centre has control over the data and can ensure that encryption is in place. With data that is imported the centre must rely on the data exporter to encrypt the data, and the encryption cannot therefore be guaranteed. The expectation would be that the centre would work with the data sender to try and ensure suitable encryption was used.

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. It is therefore often necessary to develop transfer methods that can work around this problem, for instance physically sending USB sticks holding the encrypted data. It may be safer and easier to routinely encrypt *all* data transfers, so that such methods become standard.

Encryption will usually occur before transfer, e.g. before transfer of the data to a USB stick or CD. In some cases, however, secure access systems maintained by industrial sponsors may allow encryption to take place *during* direct electronic transfer.

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.

DM07.05: Requests to amend previously transferred data

Procedures should exist to deal with requests for direct changes of previously transferred data.

This standard deals with a relatively unusual situation, and one that some centres may never experience. It involves the need to directly change a few values in data that has been previously transferred, and either bulk uploaded to the CDMS or aggregated with the rest of the analysis dataset.

It does *not* deal, or in any way suggest that it is acceptable, under any circumstances, to change data directly when that data should and can be altered via the normal user interface, by the normal data originators, i.e. the clinical sites. But such a situation can arise if data is imported in bulk (so there is no eCRF corresponding to it in the system) and it then becomes apparent that it needs correcting.

The easiest and recommended way of dealing with this situation is simply to repeat the data import with a corrected dataset. That provides a record of the data transfer process and the source files would be retained as an audit record. But it may be that the sponsor requests that the amendments are 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 at the very end of the study because one or two subjects were found to have received the wrong treatment.

A centre should be either be prepared for such a situation, or prohibit it entirely and insist on another method of editing the data (e.g. by repeating the bulk upload). 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, probably as a file note in the trial's master file.

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

DM08: 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.

DM08.01: 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.

DM08.02: 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.

DM08.03: 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.

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.

DM08.04: 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 DM08.07, DM08.08) 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.

DM08.05: 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.

DM08.06: 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.

DM08.07: 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.

DM08.08: 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.

DM09: 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. Whatever the electronic format used for the data itself, associated metadata and other project documents (the protocol, TMF, analysis plans, etc.) must also be included in long term storage, to provide the necessary context for full understanding of the data.

Anonymising the data can simplify long term storage requirements because the data becomes less 'risky' if it is accidently made accessible. Anonymisation, and other de-identification techniques, also allows data to be shared with others, because sharing pseudonymous data (and almost all trial data is pseudonymised) is normally seen as requiring explicit consent, though regulations governing data sharing may change in the future.

At the moment the standards do *not* include the need for data transformations as part of long-term storage, or the need to prepare data for possible sharing on request, though they may in the future, especially as sharing individual participant data becomes a more prominent issue and the techniques required become more common.

DM09.01: Determining long term storage

Controlled documents are in place that ensure that long term storage arrangements, of both trial documents and electronic data, are agreed with the sponsor.

The final decisions about *what* should be stored, *where* and for *how long* will be taken by the sponsor, acting in the context of national regulations. It is important, however, that the centre's procedures include mechanisms to explicitly agree with the sponsor these three things, as well as the 'final fate' of both electronic and paper forms of data and trial documents – usually either destruction or archiving. The agreement with the Sponsor should be in place at

the start of the trial, and be part of the Data Management Plan, rather than being negotiated at the end of the trial when the data is ready to be archived.

This is a rapidly changing area, especially with the recent interest in using data repositories for individual level trial data, which may impact how trial data is stored in the future. Nevertheless, it is important that there are procedures in existence now that make sure these issues are addressed for each trial, and that the centre does not simply end up with an electronic copy of the data (as collected and / or as analysed) 'by default', without, in some cases, the sponsor even being aware that this is the case.

Evidence that the standard had been met would be the controlled documents themselves, with discussion of the types of long term storage typically managed by the centre, and examples of agreements or contracts between the trials unit and the sponsors that covered this area.

DM09.02: Long term storage of documents

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

Some centres return all paper records to the sponsor on completion of all trials, because the responsibility of keeping records available for inspection is usually retained by the sponsor, and also because they may not have the room or resources to arrange for long term storage of documents themselves.

If a centre does provide long term storage for paper based records, however, typically when the sponsor is its own parent organisation, then the storage facilities should be secure and include environmental protection (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.

DM09.03: 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.

DM09.04: Ensuring deletion or de-identification of data

Measures are in place, or are being developed, to ensure that if and when data is required to be destroyed or de-identified, then the destruction or de-identification process applies to all copies of the data.

This standard applies especially to those using a SaaS based CDMS, where multiple copies of data may exist in infrastructures that are not directly controlled or even easily identified by the trials unit. The same principles apply, however, to all data management scenarios.

If it is decided that data in electronic and / or paper form needs to be destroyed the data centre should have procedures in place to ensure that the destruction is complete and recorded. This applies both to the data 'as collected', e.g. to the data stored by the CDMS, and to the analysis data sets generated from them. For data or paper under its direct control this should be straightforward. For data kept within an external infrastructure, perhaps accessed through a CDMS system, this means liaising with the data processors, to understand where copies of the data are located and to receive documented assurances that all copies have been destroyed.

The practicality of destroying data within backup sets will be dependent on the nature and number of those sets (and should be one of the things considered when organising and costing backups). The centre should, however, at least be aware of the situation and able to communicate this to the sponsor.

If data needs to be de-identified (e.g. prior to transfer to a data repository) then, usually, only one copy of the data will require de-identification and other copies will be destroyed. The additional information that was keeping the data pseudonymised (i.e. that held the key to the identity of trial participants) may be destroyed, for fully anonymised data, or retained separately (for data that remains pseudonymous).

If the destruction takes place because the *physical machine on which it is stored needs to be retired or disposed*, then again the data centre needs to know how access to their data is made impossible. Traditionally, when trials units controlled their own machines, internal procedures could guarantee the physical destruction of hard disks. Now, as data is increasingly stored in facilities controlled by others, ensuring that data wiping and disk destruction takes place requires liaison, transparency, and documented feedback from the data processor.

These are relatively new concerns and some younger centres will not yet have reached the stage of destroying or archiving data. Nevertheless, dealing with them is part of the legal responsibilities of data controllers and processors and needs to be considered.

The current standard therefore allows a centre to be still developing relevant processes and procedures. The key requirement is that the centre is addressing the issue of the 'final destruction or de-identification' of their trial data and examining how that will be managed. Evidence that the standard had been met would be by discussion of current and planned processes and procedures, and draft or current controlled documents.

References

[1] Moher D, Hopewell S, Schulz K et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869 doi: 10.1136/bmj.c869. Available at www.consort-statement.org/Media/Default/Downloads/CONSORT%202010%20Explanation%20and%20Elaboration%20(BMJ).pdfb

[2] ICH. Integrated addendum to ICH E6(R1): Guideline for good clinical practice E6(R2). Available at http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2__Step_4_2016_1109.pdf

[3] C. Ohmann, W. Kuchinke, S. Canham, et al. Standard requirements for GCP-compliant data management in multinational clinical trials. Trials, 12 (2011). 85–10.1186/1745-6215-12-85. Available at: http://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-12-85]

[4] C. Ohmann, S. Canham, J. Dress, *et al.* Revising the ECRIN standard requirements for it and data management in clinical trials. Trials, 2013 (14) (2012), p. 97. Available at http://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-14-97]

[5] Cornu C, Donche A, Coffre C. Référentiel ECRIN pour la conformité aux bonnes pratiques de gestion des données des essais cliniques multinationaux. Thérapie. 2015, August 3. https://doi.org/10.2515/therapie/2015042

[6] ICH GCP E6(R2) (see reference 2). Section 5.2.2 (page 23)

[7] European Medicines Agency. Q&A: Good clinical practice (GCP). Qs 2 and 8 under 'GCP matters'. Available at http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/q_and_a/q_and_a_detail_000016.jsp&mid=WC0b01ac05800296c5

[8] Canham S, Crocombe W. Data on trial: what do DMOs require from their IT hosts? International Clinical Trials. May 2017, pages 38-41. Available (paywall) at http://www.samedanltd.com/magazine/13/issue/272/article/4589.

[9] EUR-LEX. General data protection regulation. Available at http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679

[10] ISO/IEC 27001:2013. Information technology -- Security techniques -- Information security management systems – Requirements. Available to purchase at https://www.iso.org/standard/54534.html

[11] Seltzer, L. POODLE not fixed? Some TLS systems vulnerable. ZDNet. December 9th 2014. Available at http://www.zdnet.com/article/poodle-not-fixed-some-tls-systems-vulnerable/

[12] FDA. Guidance for Industry - Process Validation: General Principles and Practices. 1987. Available at https://www.fda.gov/downloads/Drugs/.../Guidances/UCM070336.pdf

[13] McDowell R. Understanding and Interpreting the New GAMP 5 Software Categories. Spectroscopy. Volume 24, Issue 6. Jun 01, 2009. Available at http://www.spectroscopyonline.com/understanding-and-interpreting-new-gamp-5-software-categories

[14] Plagiannos C. What is GAMP5 and how do I use it effectively? Montrium. Available at https://blog.montrium.com/experts/what-is-gamp5-and-how-do-i-use-it-effectively

[15] Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. BMJ Open 2017;7:e018647. doi:10.1136/ bmjopen-2017-018647. Available at http://bmjopen.bmj.com/content/bmjopen/7/12/e018647.full.pdf

[16] CDISC. Clinical Data Acquisition Standards Harmonization (CDASH). Available at https://www.cdisc.org/standards/foundational/cdash

[17] FDA CFR 21. Part 11 – Electronic records; electronic signatures. Section 11.10 (e). Available at https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=11.10

Appendix A: Treatment Allocation standards (optional)

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.

ST01.01: 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.

ST01.02: 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.

ST01.03: 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.

ST01.04: 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 ST01.05).

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.

ST01.05: 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 inhouse 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.

ST01.06: 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.

ST01.07: 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.

ST01.08: 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.

ST01.09: 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.

Appendix B. 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 described in the Introduction to section DM02.

Clinical Data Management System (CDMS): Within centres, the system (or collection of systems) that holds the clinical data gathered during trials. CDMSs are often commercial software systems 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: See *Clinical data*

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 research staff cannot access eRDC systems while interviewing patients and / or collating information, or prefer not to, 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: See *Clinical data*

IT host organisation: 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: 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) because unique 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: 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: 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: 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.

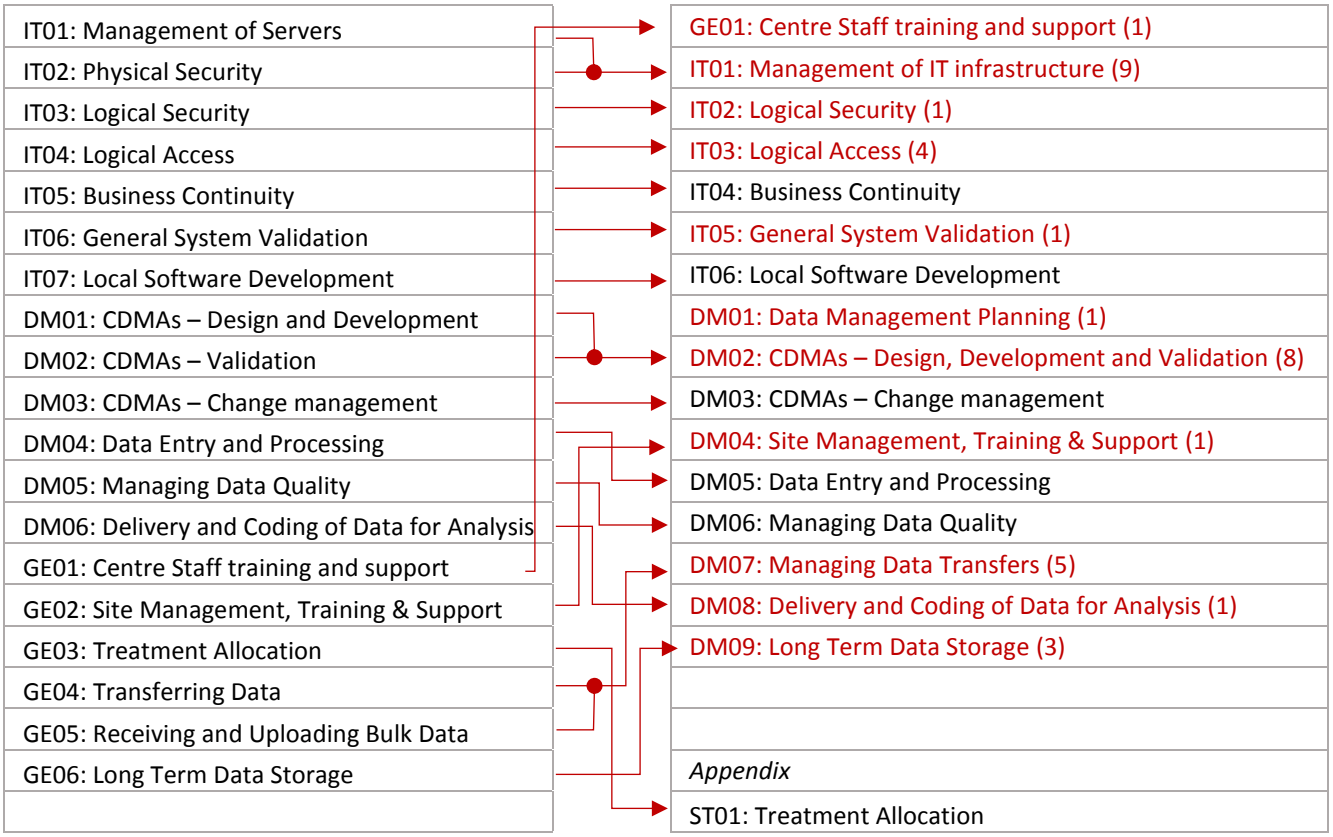
Systems directly supporting Clinical Trials: This phrase, and minor variations of it, refers to all systems that store or process trial clinical data or analyses, trial administration and financial data, or trial specific documents (e.g. protocols, agreements), i.e. all things that directly support trial 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, these 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*).

Appendix C: Differences between versions 4.0 and 3.1

Version 4.0 represents a substantial revision and re-ordering of the standards. The sections used in each version are tabulated below, showing the changes in ordering. These changes were introduced to try and simplify the arrangement of the standards, and – especially for the DM lists – put them in an order that more accurately represents the 'life cycle' of data within clinical trials.

The V4 lists that include changes to one or more of their standards are in red. The number of changed standards is indicated by the number in the brackets (in several cases list headers were also changed).



As can be seen from the table, in 3 cases pairs of lists were rationalised to a single list (IT01 and IT02 to the new IT01, DM01 and DM02 to a new DM02, and GE04 and GE05 to a new DM07). In addition GE03 on treatment allocation has been turned into an appendix. One new list has been added (DM01) though this only has 1 standard. There was therefore a net reduction of 3 lists from the certification standards. The number of standards is reduced from 129 to 106, partly because of the removal of the treatment allocation standards, mostly because of the simplification of standards within the merged lists.

The changes in each modified list are described in more detail below:

GE01: Centre Staff training and support: The standard GE01.04, Managing concerns – alternative pathways, has a new, clearer, standard statement and revised E&E material. The previous version was sometimes misinterpreted by units.

IT01: Management of IT infrastructure: This list has been created from the combination of the old lists on server management and server physical security. The constituent standards are:

IT01.01: Infrastructure location. A new standard that deals with the location of infrastructure in terms of the legal jurisdiction within which it falls. Especially relevant for units using a SaaS based CDMS.

IT01.02: Secured server room. A revised version of the previous standard IT02.01 Locked server room.

IT01.03: Secured power supply. A slightly revised version (E&E material only) of the previous IT02.02, with the same name.

IT01.04: Controlled temperature environment. A renamed and revised version of the old standard IT02.04, on the need for a controlled environment.

IT01.05: Fire and smoke alarms. A renamed, revised and clearer version of the old IT02.05.

IT01.06: Server failure and response. A revised version (E&E material only) of the previous IT02.03.

IT01.07: Server support and recovery from downtime. A renamed and revised version of the old IT01.03 standard. The standard statement itself remains the same and the E&E material has only been slightly revised.

IT01.08: Server configuration records. A slightly revised version (E&E material only) of the previous IT01.02, with the same name.

IT01.09: Server software maintenance. A renamed and revised version of the old IT01.05 standard. The standard statement itself remains the same has only been slightly revised.

The previous standards on server specification (old IT01.01) and server retirement (old IT01.04) were dropped, mostly because neither discriminated between units. Concerns about data destruction on machine retirement have been included in a new standard (DM09.04, see below). Most of the points made in the previous 'other aspects of environmental and system control' section have now been integrated into the E&E material of the relevant standards.

IT02: Logical Security: IT02.02 on the unit's commitment to data protection is a revised version of the old IT03.02, of the same name, to take into account the demands of the GDPR.

IT03: Logical Access: These standards were re-ordered, into a more logical order, and the following were changed, though not radically in any case.

IT03.02 Network log-in management. Slightly revised version of IT04.04. Only E&E material changed.

IT03.04: Remote access (not using a browser). Standard name changed. Otherwise the same as the previous IT04.05.

IT03.05: Access control management. Slightly revised standard statement and E&E material revised to incorporate the previous DM04.02, as well as the previous IT04.02.

IT03.07: Administration of access to clinical data. The same standard as the old IT04.07, but with revised, expanded E&E material.

IT05: General System Validation, and DM08: Delivery and Coding of Data for Analysis. The standard on validation of extracted data (the old DM06.07) has been dropped, now covered by an expanded IT05.08.

DM01: Data Management Planning: This is a completely new standard that requires the presence of a data management plan (DMP) for each study, and for the data centre to have a DMP template available. It has been introduced because data management is becoming more complex, with a greater variety of data sources, and with the long term management of data acquiring greater importance. DMPs are therefore seen as necessary.

DM02: CDMAs – Design, Development and Validation. This list represents the combination of the previously separate lists on CDMA development and validation. In fact these two processes are almost always discussed and demonstrated together, so that it makes sense to combine the lists and take the opportunity to rationalise some of the standards, because there was considerable overlap between them. As well as a much revised and extended header, which attempts to give an overview of the development process, the constituent standards were re-ordered and revised:

DM02.01: CDMA development and validation policies. A straightforward combination of the old DM01.01 and DM02.01, each dealing with the need for policies in the different areas of development and validation.

DM02.02: The CDMA and the protocol. A revision and substantial expansion of the previous DM01.02, 'Requirements specifications of CRFs'

DM02.03: Creating a full functional specification. A revision and substantial expansion of the previous DM01.03, 'Functional specifications of CRFs'

DM02.04: Isolation of CDMAs in development. A revision and expansion of the previous DM01.05, 'Isolation of development CDMAs'.

DM02.05: Input into CDMA development by end users. A revision and expansion of the previous DM02.04, 'Assessment of CRFs by users'.

DM02.06: Cross-disciplinary approval of the functional specification. A revision and expansion of the previous DM01.04, of the same name.

DM02.07: CDMA validation against the functional specification. A combination of the previous DM02.02 (CDMA specific test plan), DM02.03 (CDMA testing against functional specification), and DM02.06 (Validation detailed findings).

DM02.08: CDMA final sign off into production. Slightly revised version of DM02.05 CDMA approval.

The standard on isolation of training CRFs (was DM01.06) was merged into the very similar standard in the site support list, now DM04.03. The standard on the use of interim CRFs was transferred unchanged to the data entry section, where it is DM05.02

DM04: Site Management, Training & Support: The standard DM04.03, Isolation of training eCRFs, (originally GE02.03), has been revised and now also incorporates the old DM01.06.

DM07: Managing Data Transfers: This new list represents an amalgamation of the previous lists GE04, on transferring data out of the unit, and GE05, on transferring data in, to a new single list about data transfers in general. This was possible because the requirements in each of the previous lists were very similar. Along with a revised header, the standards are:

DM07.01: Data Transfer Procedures. A combination of the previous GE04.01 and GE05.01, dealing with procedures for both export and import of data.

DM07.02: Records of Transfers. A merger of GE04.04 and GE5.04, which dealt with this topic in the context of export and import respectively.

DM07.03: Retention of intermediate and transferred files. A merger of the old standards GE04.05, GE04.06, GE05.02 and GE5.03. These had split the retention of different types of file but were all essentially requiring the same thing.

DM07.04: Encryption of Individual Data. Modification of the old GE04.02 on the same topic, plus expansion to consider the receipt of encrypted data.

DM07.05: Requests to amend previously transferred data. A modification of GE05.06, with a changed name and standard statement, for greater clarity, as well as revised E&E material.

DM09: Long Term Data Storage. This section has been simplified, though now includes an additional standard. The standards affected are:

DM09.01: Determining long term storage. A merger of GE06.01, GE06.04 and GE06.05, as all three standards dealt with establishing the policies and parameters of long term storage. This revised version stresses the need to organise this with the sponsor at the beginning of a trial.

DM09.02: Long term storage of documents. Revision of the E&E material.

DM09.04: Ensuring deletion or de-identification of data. A new standard that deals with the need to ensure that data is destroyed (or de-identified) when this is required, either because the data's life has reached the end of its agreed retention period, or because the hardware on which it is sitting is retired. Though particularly relevant to those using SaaS facilities it applies to all data centres.

ST01: Treatment Allocation. As explained in the introduction, this entire section has been moved into an appendix, as an optional set of units that is no longer included in the data centre certification process.

In total 30 of the standards in v4.0 are revised versions of one, occasionally two or more, previous standards. Of the 129 standards in version 3.1, 9 were transferred to the appendix, whilst 17 were dropped or merged. 3 completely new standards were added, giving a new total of 106.