GUIDE TO eDISCLOSURE

Version 0.1
1 November 2013

Document Reference :
Guide to eDisclosure V1_0 01_11_2013.docx

Issue : 1.0 Dated 01 November 2013

COPYRIGHT AND CONFIDENTIALITY

DISCLAIMER

TRADEMARKS / LOGOS

DOCUMENT AUTHORISATION AND VERSION CONTROL

DOCUMENT TITLE	Guide to eDisclosure

REFERENCE	ISSUE	DATE
eDisclosure Guide	Issue 1.0	Nov 13

Date	**Version**	**Comment**
01/11/13	1.0	Issued Version.

CONTENTS

1. OVERVIEW

One of the primary objectives of the joint initiative undertaken by TeCSA, TECBAR and the SCL in March 2013 was to address the increasing disparity in terms of knowledge and experience of the eDisclosure process amongst practitioners which continues to affect the efficiency and cost-effectiveness of the litigation process much to the detriment of all concerned in its operation and for clients seeking to use that process as a means of resolving their disputes.

It is with that objective very much in mind that the Working Group asked Andrew Haslam of Allvision to prepare this Guide to eDisclosure which is principally targeted at providing a comprehensive introduction on the subject to TeCSA members and others involved in the litigation process who have limited experience of the topic. That draft formed the basis of this document and the Working Group are particularly appreciative of Andrew's efforts in this connection.

The intention is that this Guide should be used in conjunction with the eDisclosure Protocol and the other key documents which practitioners will have to master in this connection such as the Disclosure Statement and the Electronic Data Questionnaire. But, hopefully, it will materially assist the practitioner in all aspects of managing the process in practice, including the vital interfaces with the client and his IT personnel, the many different types of service provider, the other side and, last but not least, the Court itself.

This Guide is in two distinct parts. The first provides a description of what eDisclosure really means, rather than what others might try to persuade you it means, and then uses the industry standard EDRM model as a mechanism for taking readers through the eDisclosure process at a high level and explaining how it might impact upon their cases.

Having set the scene in terms of how eDisclosure fits into the litigation process, the second part of the Guide then provides an approach to procurement, and gives information on the eDisclosure marketplace, so that informed decisions on selecting software and vendors can be made. These aspects of the process are often the most difficult for the legal practitioner because of lack of familiarity with the procurement process but, with the introduction of the new procedure rules and the Jackson reforms, they have become increasingly important and need to be addressed much sooner in the process than many might expect.

A table overleaf provides a quick reference to the various parts of the Guide and how they might be used.

There is a glossary of terms used in eDisclosure at the end of the document.

A more comprehensive listing of the service providers and their software products can be obtained by downloading "The UK Buyer's Guide to Litigation Support Systems" from here.

2. GUIDE STRUCTURE

The remainder of the Guide gives more information on eDisclosure, issues you might encounter and how to procure the right systems for your needs from the correct type of vendor.

Throughout the Guide the following boxes will be used to draw your attention to specific points.

NOTE: Brief Description

Used to draw attention to specific elements of the text.

BEST PRACTICE: Brief Description

Details a point of best practice that users are advised to follow.

WARNING: Brief Description

Warns users of any potential issues.

3. EDISCLOSURE – AN INTRODUCTION

The purpose of this Chapter is to take the reader through the process of eDisclosure, starting with a definition of what it means, and what it does not mean, and then progressing (by means of an industry standard model) through the various stages of the procedure. At the end, there is a more advanced section on the specific area of Computer Assisted Review or CAR, but this can be skipped until required.

3.1 Definition

A simplistic definition is that eDisclosure is all about the disclosure of electronic material. However we need to dig a little deeper into that statement of the obvious.

There are two parts to the definition; the words "disclosure" and "electronic material". Let's explore the second one in a little more detail. Electronic information refers to the "stuff" that is within emails, Word documents, Excel spreadsheets and PowerPoint slide shows. This is the level that most readers will need to interact with for eDisclosure. The term also includes databases, social media (Facebook, LinkedIn, Twitter), digital audio such as recorded conversations in deal rooms, support centres et al, images, mobile phones and a host of other increasingly more exotic types/locations. All of this is known as Electronically Stored Information or ESI.

The definition of eDisclosure then becomes the process of identifying, collecting, processing, analysing and reviewing ESI for legal proceedings.

Because so much of the software in this area comes from the United States, it is as well to recognise the synonym eDiscovery, which is the American term for eDisclosure.

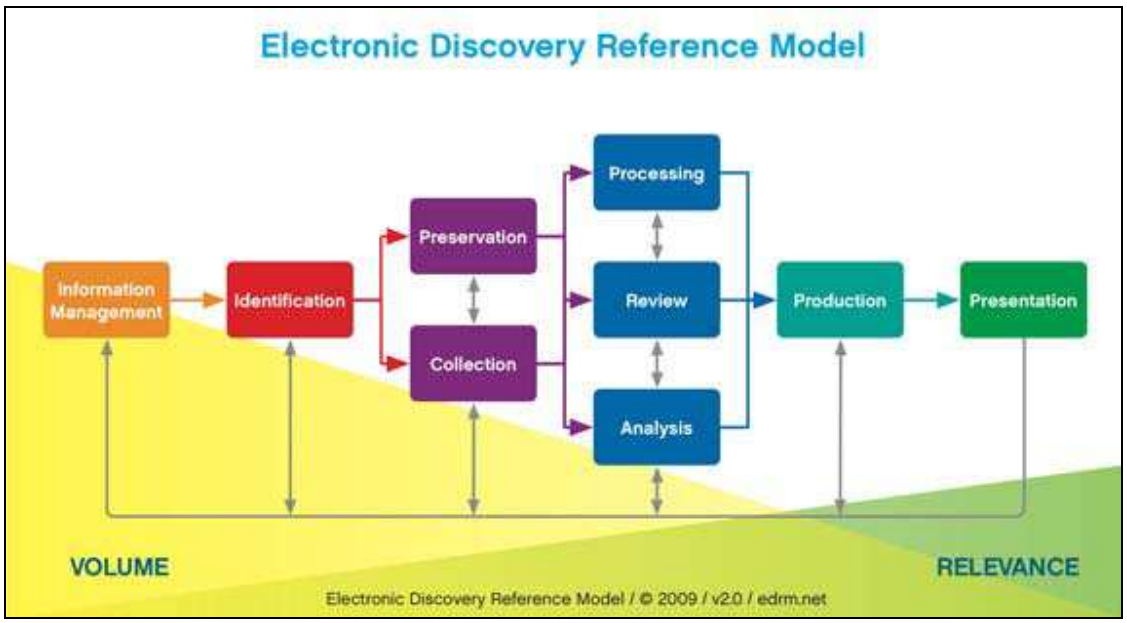
For the sake of completeness, eDisclosure is NOT the process of agreeing the electronic media by which information will be transferred. When disclosure was all about transferring images of pieces of paper between legal entities, there used to be discussions on format

might be used for the images, and which kinds of physical media could be used to hold the images and their data, be it "floppy disks", tapes or a memory stick. This discussion used to be incorrectly labelled as "eDisclosure", and is mentioned here to dispel any legacy misunderstandings.

3.2 EDRM Model

As you would expect for a technical process, there is an official model showing the various steps involved in the whole procedure. This is known as the Electronic Discovery Reference Model (EDRM). The model is discussed in great detail at www.edrm.net, but for the purposes of this report the standard graphic used to give an overview of the approach is shown below.

If you want to see how the various stages of the EDRM process relate to the procedural requirements contained in the CPR and the related Practice Direction, and what this means in terms of the action that you should be taking in the context of eDisclosure and when, you should refer to the eDisclosure Timeline in the rest of the protocol pack.



NOTE: Circular Processes

Some of the lines used to link the boxes have arrowheads at each end. This implies (and very often means) that the workflow "loops" around and that having done one box, you need to go back and repeat some previous processes.

The background in the graphic shows the volume of data decreasing as the various stages are completed, with a corresponding rise in the identification of relevant material. The various processes involved in eDisclosure are shown as discrete boxes with workflows between them. The main use of the model for this review is to provide a "shorthand" to explain the functionality that different software products provide.

For example a company specialising in area of Records Information Management, email archiving and the ability to "freeze" documents within a client environment in order to meet the requirements of disclosure (the US "Legal Hold" concept), might state that they "work in the left hand side of the EDRM model". A forensic firm might focus on the purple Preservation and Collection areas, whereas a litigation support software firm might have literature showing them focused on the blue boxes of Processing, Review and Analysis.

The following sections go through each of the "boxes" of the model and look at three things:

(a) **Description**

The official description of the individual process/procedure.

(b) **Legal Implications**

What you as a lawyer might become involved in within this particular process.

(c) **Who can help**

A brief overview of the types of services and/or software products you might need to support you in this process. Wherever possible software products are named, though these are non-exclusive lists.

3.3 **Information Management**

(a) **Description**

Getting your electronic house in order to mitigate risk & expenses should eDiscovery become an issue, from initial creation of electronically stored information through its final disposition.

(b) **Legal Implications**

This is an area also known as Litigation Readiness and is normally conducted by the client. You might be called upon to assist in terms of providing advice on data retention requirements for a specific industry.

(c) **Who can help**

Normally the preserve of the larger consultancy firms working with their clients.

> **NOTE: There's a BD opportunity here**
>
> Why not approach your clients (possibly with a technology partner by your side) and engage them in conversations about becoming litigation ready. You supply the detailed legal and business specific knowledge and the client is better prepared for the "evil day" of litigation.

3.4 **Identification**

(a) **Description**

Locating potential sources of ESI & determining its scope, breadth & depth.

(b) **Legal Implications**

This is the initial stage where the client comes to you and explains their problem. Using a combination of your legal knowledge and their understanding of the organisation they work for, you should start to get an idea of where the ESI might be located. You might want to incorporate a rough outline of the scope in your initial client care letter, you might be

happy with going on what the client tells you (it is their data after all), or you might want to delve a little deeper into what data silo's exist.

This is very much a "horses for courses" area. A good rule of thumb is; do you know enough about the client and their technology to run a "sanity check" over what they are telling you?" If not, you might want to get some professional help to try and uncover the "nasties", before they bite you downstream.

(c) **Who can help**

Some consultancy/forensic organisations specialise in producing something called a "data map". This is NOT a technical document showing all the servers and other bits and pieces that makes up the client's IT infrastructure. Rather it is a written description (possibly with a diagram or two) of where the various data sources are. For example; "*Most of the information is stored on the email servers, but some is on the back-up tapes, and there is a company the main firm took over last year that has got its own IT infrastructure which will need to be examined."*

BEST PRACTICE: Build a "Data Map" as soon as possible

Should be a single piece of A4 that describes where the data is stored and any issues surrounding it. Print it out and have it on the front of the Matter file.

3.5 **Preservation**

(a) **Description**

Ensuring that ESI is protected against inappropriate alteration or destruction.

(b) **Legal Implications**

Once you have determined the possible scope of the areas you might (or definitely will) be collecting data from, you need to ensure that the client doesn't delete or damage the ESI in those locations. Again this might be something for your initial engagement letter and you might need technical help. You might cover things like; stopping the re-use of back-up tapes (it can be cheaper to buy a whole new sets of tapes that over-write important evidence), or removing the limits on email in-boxes that cause emails over 60 days old to be deleted, or putting a hold on the re-use of the PC used by the employee that is suing your client.

(c) **Who can help**

There are very expensive, mainly US based tools that will enforce the American concept of "legal hold". If your client has got this kind of software in place, they are probably involved in serial litigation and you won't be reading this kind of Guide.

The key area that causes problems is when there is a "disconnect" between the client and their IT department. You might want to make sure that someone from the client's IT department is involved in the initial

meetings/conversations so that they can understand what you are asking the client to do. If you are not comfortable with your level of technical knowledge you might want to take along support from a vendor so they can talk “Geek to Geek”.

 **WARNING: Forget the Client’s IT staff at your peril**

Make sure the Client’s IT people understand what data you are preserving, so that they don’t inadvertently destroy it as part of their normal business practice.

3.6 Collection

(a) Description

Gathering ESI for further use in the e-discovery process (processing, review, etc.).

(b) Legal Implications

As a rough guide, there are two kinds of data collection, those that require a forensic process (complete with chain of evidence documentation) which is normally in cases of fraud, and the rest, where you just need to collect the data in a competent manner. The first group is a specialist area, and if you are involved in this kind of proceedings, you probably have got a "tame" forensic investigator that you can use, if not you need to find one. The forensic data capture might also involve things like retrieving data from mobile phones, making forensic copies of PC's or other computer equipment, and all other kinds of highly specific activities.

For the more general data collection requirement, there are issues here, more fully explored in subsequent chapters. Suffice it to say, that it is best to have a qualified individual or service provider collect the information and there are pitfalls involved in letting the client do it for themselves. However, in some cases this might be OK.

There are two schools of thought on the scope of data collection. One is that you collect very broadly (so you only disturb the client once) and use the downstream processing to winnow out what you need. The other is that you do a focused collection and run the risk of having to come back and widen the scope. Each is valid, and they are non-exclusive, in that you can start focused and (if the case progresses/warrants it) come back later and do a wider collection for downstream culling.

This is where an understanding of the case, and where the information is stored is invaluable, as you can then make informed decisions. With, of course the price of the different options and how it affects the downstream processing very much to the forefront of people's minds.

 **WARNING: Badly collected data can lose you the case**

Leave data collection to the experts. If the client offers to do to save money, be very careful and make sure they understand the implications if they get it wrong. It is NOT just a case of copying things from one place to another.

(c) Who can help

The forensic data collection organisations are a specific grouping of service providers. Make sure that you get a forensics company and not

just a litigation support service provider that has sent someone on a data collection course. They tend to use products such as Guidance EnCase or Access Data's FTK toolkit to carry out the collection process, indeed this product is often used by the other group as well, just without the formalised chain of evidence documentation. The key thing is that people in this first group are used to appearing in court as an expert witness to explain how they obtained the specific piece of information.

The second type of data collection can be handled by a number of vendors, though in practice using the specialist organisations and dispensing with the formal side of things is a sound tactic.

3.7 Processing

(a) Description

Reducing the volume of ESI and converting it, if necessary, to forms more suitable for review & analysis.

(b) Legal Implications

The "shorthand" term for this stage is Early Case Assessment (ECA), or more accurately Early Data Assessment. This is where the range of options open to you increases quite dramatically. In the main the processing goes through two stages; first the data is "cleansed" in that unwanted types of information are automatically removed, this process can also involve the identification of duplicate versions of emails etc. Second, the data is loaded into a tool that allows the user to identify the information they want to take forward into the next stage.

The capabilities of the tools at this stage are quite bewildering, but in the main, you are trying to identify the information that you will want for your case, so an understanding of the key people involved in the matter (normally called data "Custodians" in techie speak) is good, as well as the date range that covers when key events happened. You might have an idea of the kinds of terms or Key Words that would be useful to search on, but there are other much more powerful technologies at your fingertips that will surpass the results of key words.

The key to getting maximum benefit from this stage is for you to team up with a sympathetic vendor and let them drive the technical process, whilst you supply the background and legal requirements of the matter.

BEST PRACTICE: Build a partnership with a service provider

Let the service provider drive the processing software; you should focus on giving them the salient points of the case issues, so that you can identify and cull out unwanted data.

(c) Who can help

The choice of software here is split into two groups. On the one hand there are products specifically designed to work in this area. These range

from software that third party suppliers use such as Clearwell, LAW or Nuix, through to tools specifically designed by suppliers for this area, like the ICE product from Palmer Legal Technologies or the snappily named MM/PC tool from eMag. There are a number of these specialist tools, and the mention of those above is not to elevate them above the rest, purely to show examples of the genre.

Secondly, there are emerging products from the next stages of the process (Review & Analysis), that incorporate functionality for this step as well. Products such as Lateral Data's Viewpoint come to mind as does the Access Data range, and Recommind's offerings. Again, these are mentioned as examples and not as an exclusive list.

Until recently the Early Data Assessment modules of the "All in One" products did not match the functionality provided by dedicated tools. This is no longer the case, which is good news in one way as it adds to the choices available to you, and bad news in another, as it adds complexity to the selection process in this area.

NOTE: Controlling costs starts here

Getting a good deal on processing costs is as important as having the latest technology. There is no escape from this, you have to put in some effort in building a relationship with a service provider in order to understand your option.

3.8 Review

(a) Description

Evaluating ESI for relevance & privilege.

(b) Legal Implications

For many, this and the next stage form the hub of the eDisclosure process. The products mentioned here will be the environment in which you and your legal team will conduct most (if not all) of your on-line interaction with the data. The first half of the equation is the ability to review ESI and assign values for: Relevance, Privilege, Trade Secret, Personal Data, and as many case specific topics as you can shake a stick at. By the time you are in this stage, you will know what the issues are that you will be fighting the case on, and the various criteria you will apply to determine relevance et al. Your role will probably be to oversee the team that is carrying out this review work, though in smaller cases you might be doing the work yourself.

For large scale review exercises; you might have to recruit contract legal staff to do the work under supervision, you might have a "near shore" option of a cheaper office outside of London, you might be involved in an off-shore Legal Processing Operation such as Integreon (India), Exigent (South Africa) or Capita (Poland). In essence you will direct the team (whatever the size and geographical location) and provide overall Quality Assurance back to the Client.

(c) **Who can help**

The short answer to this one, is a lot of service providers. There is a lot more analysis of the firms and their products in the Chapters after this, but the broad split is between organisations that have their own software, and specialist software products that are supplied by different types of third party vendors be they consultancies (of different sizes) or more generic companies (that come from different backgrounds). There is a bewildering mix of software functionality and supplier personnel that combine to give you a multitude of options.

The key is that you should go through a procurement exercise before you are deep in the middle of a case, and thus make the decision in a rational cost effective manner, and not as a result of a frantic phone call to the first service provider you can find on a Friday afternoon.

That is what the rest of this Guide is about.

NOTE: What works for one firm, doesn't for another.

The "right" software for law firms, varies according to the people on the legal staff, the kind of law they specialise in, the functionality of the software and the "chemistry" with the people from the service provider.

3.9 **Analysis**

(a) **Description**

Evaluating ESI for content & context, including key patterns, topics, people & discussion.

(b) **Legal Implications**

This stage is so interwoven with the previous one, that though they are separated for technical reasons, in practical terms they will often take place within the same piece of software (albeit in some cases with the assistance of specialist plug-in modules).

The trick here is to understand what you need to do in order to meet the legal requirements of the case, and then how the technology can help you. By legal requirements I mean the issues of the matter as bounded by the court, cost and time. There is a scale of software tools available, and which ones you use are defined by the case, not by the product.

I find a useful analogy is the way in which it is possible to capture TV programs so they can be watched when we like. At the bottom end of the scale are VHS / Betamax video recorders, which are good for looking at something from start to end, but that's about it. So if you want to do a linear review in which you look at virtually every bit of ESI from "document" one to one million, then there are VHS litigation support equivalents that will let you do so.

If, however, you have a more complex viewing requirement and want the equivalent of Sky HD+ box that allows you to record three shows at once,

pause live TV, access the past 7 days of shows and download movies from on-line services, you want the litigation support products with more functionality.

The elephant in this particular room that everyone is avoiding is the concept of Computer Assisted Review or CAR. At this stage all we will do is acknowledge its existence, note that it has a formal definition in a Section or to, and hold that thought until we get to the following Chapters which give you an idea of all the different technologies that are available.

BEST PRACTICE: Understand the "what", not the "how".

No-one (particularly the Judiciary) will expect you to be able to explain the workings of the software. If you build an approach that samples the output to show it is working, then you have the basis for agreeing/arguing with your opponent at the first CMC.

(c) **Who can help**

Same as before. Lots of service providers can help. The trick is to select one who can become a trusted partner, before it all starts getting too hectic.

3.10 **Production**

(a) **Description**

Delivering ESI to others in appropriate forms & using appropriate delivery mechanisms.

(b) **Legal Implications**

You will want (have) to agree with the other side the scope of what you are delivering. The technical details of what is being handed over can be left to the service providers with the following exception.

It is assumed that the majority of the cases that readers of this Guide are involved in will focus on emails and their attachments, which will mainly be MS Office outputs in terms of Word, Excel and PowerPoint, with some PDF's thrown in. The battleground here can be the format of those ESI items. Word, Excel, PowerPoint exist in what is called "Native" mode inside their programs, that is you can modify and change the contents of the document, spreadsheet or presentation. Those Native documents have something called metadata associated with them, things like the original creator of the document, the date it was last printed etc. For Excel and PowerPoint, unless the ESI is handed over in Native mode, it isn't much use to the other side, so they will (quite rightly) insist upon it being handed over in that form. Word can cause all kinds of headaches, in that it can contain "Track Changes" which in turn may or may not hold privilege information etc. An option (reached by agreement with the other side) might be to create text searchable PDF's and only hand those over.

If the opposition suggests only giving you "TIFF" images of the ESI, you should resist very strongly.

It is sufficient for this level of explanation to know that there are issues in the production of ESI and be aware of them and your preferred approach long before you start engaging with the other side.

WARNING: Don't "degrade" information

Neither you nor the other side should be taking 100% searchable ESI and turning it into images of printed pages. Those kind of games are contrary to Paragraph 34 of PD 31B and could result in an adverse order for costs.

(c) **Who can help**

Whoever is supporting your litigation software should take care of the technical aspects of exchanging information. You might need to work with them in understanding the implications of the options that the opposition give to you.

3.11 Presentation

NOTE: Only for the specialist few

It's rare that we get to use presentation systems. Of more likelihood is a request for a set of linked PDF's and that is something the service providers will happily do for you.

(a) **Description**

Displaying ESI before audiences (at depositions, hearings, trials, etc.), especially in native & near-native forms, to elicit further information, validate existing facts or positions, or persuade an audience.

(b) **Legal Implications**

This is a very US centric part of the model, and for 99% of the readers of this Guide will never apply. For those who are using a court room presentation system, you should be well versed in the functionality of hat area.

(c) **Who can help**

In England and Wales this is a very, very specialist area and readers will be able to find the suppliers very easily.

3.12 Summary

At the start of the process it is about you establishing the scope of the case, identifying the real issues, and then designing and driving the data collection and culling in an informed manner. You should NOT have to get involved in actual processing of data.

Once you are into the Review and Analysis stages, there is a very large range of options. The best advice is to carry out a procurement exercise, were you try as much as possible to compare Apples with Apples. That is what the rest of this Guide is about.

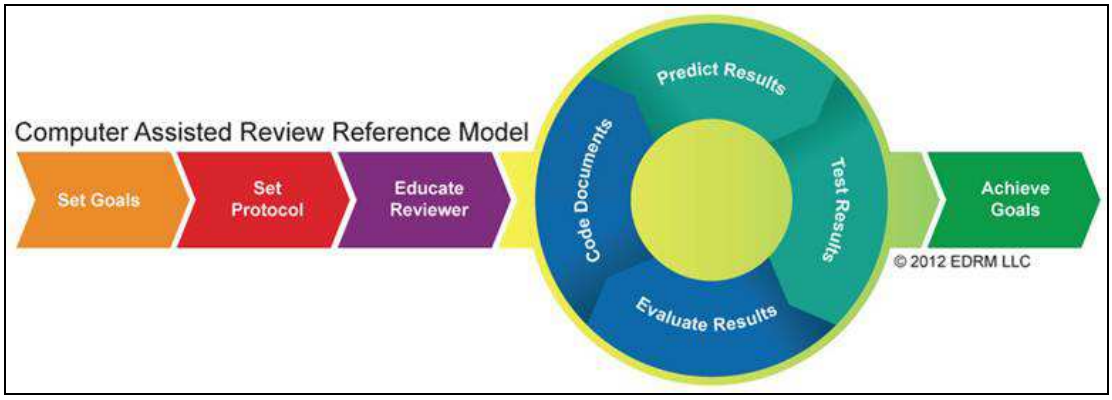
3.13 CARRM Model

NOTE: For Advanced Readers only

This Section is only for those individuals who are contemplating the use of Computer Assisted Review or CAR. First time readers can probably skip this bit and come back to it when they need to.

In December 2012, the EDRM team published a draft model and definitions for the area of Predictive Coding, otherwise known as Technology Assisted Review (TAR) or, as they (and a lot of other people) prefer to call it, Computer Assisted Review (CAR).

The model shown below is followed by the text from the EDRM site.



(a) **EDRM's Computer Assisted Review Reference Model**

Computer Assisted Review (CAR) is a process of having computer software electronically classify documents based on input from expert reviewers, in an effort to expedite the organization and prioritization of the document collection. The computer classification may include broad topics pertaining to discovery responsiveness, privilege, and other designated issues. CAR may dramatically reduce the time and cost of reviewing ESI, by reducing the amount of human review needed on documents classified as potentially non-material.

The EDRM Search team has prepared a draft Computer Assisted Review Reference Model (CARRM) to document the steps of the process. This model represents joint efforts of the best known providers in Computer Assisted Review – Autonomy, an HP Company; Daegis; Exterro; Falcon Discovery; FTI Consulting; kCura; KPMG LLP; Kroll Ontrack; NightOwl Discovery; and Recommind – as well as leaders from Bowman & Brooke LLP; DLA Piper LLP (US); Littler Mendelson, PC; and Quarles & Brady LLP.

The Major Steps in the CARRM Process are described below

(b) **Set Goals**

The process of deciding the outcome of the Computer Assisted Review process for a specific case. Some of the outcomes may be:

- Reduction and culling of not-relevant documents;
- Prioritization of the most substantive documents; and
- Quality control of the human reviewers.

(c) **Set Protocol**

The process of building the human coding rules that take into account the use of CAR technology. CAR technology must be taught about the document collection by having the human reviewers submit documents to be used as examples of a particular category, e.g. Relevant documents. Creating a coding protocol that can properly incorporate the fact pattern of the case and the training requirements of the CAR system takes place at

this stage. An example of a protocol determination is to decide how to treat the coding of family documents during the CAR training process.

(d) **Educate Reviewer**

The process of transferring the review protocol information to the human reviewers prior to the start of the CAR Review.

(e) **Code Documents**

The process of human reviewers applying subjective coding decisions to documents in an effort to adequately train the CAR system to "understand" the boundaries of a category, e.g. Relevancy.

(f) **Predict Results**

The process of the CAR system applying the information "learned" from the human reviewers and classifying a selected document corpus with pre-determined labels.

(g) **Test Results**

The process of human reviewers using a validation process, typically statistical sampling, in an effort to create a meaningful metric of CAR performance. The metrics can take many forms, they may include estimates in defect counts in the classified population, or use information retrieval metrics like Precision, Recall and F1.

(h) **Evaluate Results**

The process of the review team deciding if the CAR system has achieved the goals of anticipated by the review team.

(i) **Achieve Goals**

The process of ending the CAR workflow and moving to the next phase in the review lifecycle, e.g. Privilege Review.

 **WARNING: Driving a CAR needs skill.**

CAR technology is very powerful, but needs to be understood. A degree of mathematical skills is required to both use the tool fully and explain the methodology to the other side. If you don't have those skills, consider adding them to your legal team, possibly from the supplier of the product you are using.

3.14 Cooperation in England and Wales

Though not part of the EDRM model, this part of the Guide would not be complete without emphasising the focus on cooperation for the eDisclosure process within England and Wales. Practice Direction 31B requires that "t*he parties and their legal representatives must, before the first case management conference, discuss the use of technology in the management of Electronic Documents and the conduct of proceedings*". Consider a meeting with the other side where both parties have; their legal representation, the client(s), the client's IT representative(s) and the litigation support providers. Use this to agree the various processes you will undertake and how the information will be shared. There is still plenty of room for argument on all the other issues of the case, but in this area you are expected to present a united front to the Judge.

If you can't agree then you can apply for directions from the court, but this could be a risky business with no one liking the outcome. Far better to have discussed and agreed up front, and the earlier in the process the better. In some cases engaging an experienced neutral mediator to facilitate the parties in reaching a consensus may be a more satisfactory way of resolving disputes which may arise.

BEST PRACTICE: Cooperation is not collaboration.

Working with the other side to smooth the path of eDisclosure is essential. You can still put your arguments and fight your case, just don't waste time and effort being obstructive. It will almost certainly add to the costs.

3.15 Summary

By this stage a reader should be comfortable with the definition of eDisclosure and the various stages it encompasses. They should also be familiar (at a high level) of what involvement they as a lawyer might have with each part of the process and what tools and service providers are available to help them.

To summarise the current position, most of the significant "players" in the litigation software world have similar capabilities, albeit they might be grouped as ECA on one hand, or litigation support on the other. The main products are truly Unicode compliant, have near duplicate facilities, "cluster" data into concepts without intervention from users, as well delivering a rich search environment and the ability to easily manipulate the results of enquiries.

The differences are evident in which area of the EDRM the product addresses. The ECA tools are far more focused on processing large volumes of emails and their attachments, with emphasis on various techniques to try and identify the potentially relevant data. Litigation support software has more focus on the review of documents for relevance and privilege, and the preparation of a case around identified themes, leading to a disclosure exchange and downstream courtroom production. Confusion arises because the various products are continuing to mature by absorbing functionality from competitors. Thus ECA tools drift into the right of the EDRM and litigation support products to the left.

Now we add to this complex mix the whole concept of Computer Assisted Review (CAR) which can be presented as a "Black box technology that supplants lawyers, so be afraid, be very afraid", when nothing is further from the truth.

Where this leaves readers trying to assess which product they should choose, is that they have to evaluate what piece of software works best for them, and their circumstances. Unlike scanning, coding and (to some extent) forensic support services, it is not possible to select a supplier on price and functionality alone. Firms need to evaluate the software by means of demonstrations (preferably with their own data) and then (optionally) trialling rival products against each other to gain an understanding of what suits their individual unique requirements and work mix.

The remainder of this Guide aims to provide information to enable readers to achieve those aims.

NOTE: What's Next?

The rest of the Guide take you through all the things you need to know in order to procure Litigation Support services and software. If you are not at that stage yet, then you can stop now, though there is some good detail on pitfalls and technical issues in Chapter 5 you might want to skim through.

4. TECHNOLOGY AREAS

The aim of this Chapter is to provide a brief outline of the various areas of technology in order to provide context for the rest of the Guide. It is not intended to provide a comprehensive description of each area, more a taste of the requirements and an overview of the main issues to be aware of.

The Chapter first looks at those areas that are outside the main scope of the Guide and then goes through the remainder in a roughly chronological sequence within the timeframe of a project.

Though the Guide mainly focuses on eDisclosure, there is mention of scanning as many cases still involve paper alongside the ESI material, so the topic is covered for completeness.

> **NOTE: Do I really need to read this?**
>
> This gives more context and detail on the different tools that you have at your disposal. If you think you have a good enough understanding already, skip to Chapter 5.

4.1 Litigation Readiness / RIM / Email Archiving

This Section examines those technology areas which are outside of the main scope of this Guide. It is included to provide background information on these areas.

(a) Litigation Readiness

A common analogy is that, if litigation support and eDisclosure are the ambulance at the bottom of the cliff, then litigation readiness is the fence at the top that stops you falling over. Litigation readiness is inexorably bound up with a sound Records Management policy, and this in itself is a good efficient business practice. Indeed for some firms in the financial sectors, the demands of Sarbanes Oxley, the Financial Services Act, Basel 2 and MiFID, make a sound records management strategy an essential element of their business. Increasingly, the way in which a business handles electronic data has a value in its own right and that value is being assessed in Merger and Acquisition situations. The demands of the Bribery Act only increase the pressure for organisations to have a good grasp on the control and management of their ESI.

As well as the regulatory and compliance drivers, there are increased risks for firms involved in global transactions and a poor or badly prepared response to litigation can result in significant brand tarnishment, or even the destruction of the company. On a more pragmatic level, there is a good business case for controlling the spiralling costs of eDisclosure, and let us not forget, a key element of the rules changes in both the US and England, was the requirement for lawyers to specifically address the challenges of eDisclosure.

For more information on this topic see the whitepaper stored here.

(b) RIM / Document Retention Policy

As just mentioned, one of the other main elements in this area is that of policy as encapsulated by Records Information Management as a topic

title, and Document Retention Policy as a specific concept. Of course what is really meant by a retention policy, is when can you delete or destroy records. That being said, a rational policy that at least removes all the superfluous duplicate copies of emails and other ESI items, does pay dividends once litigation is underway and you have to collect and process all of those individual files.

The main priority in this area is to ensure you are doing the best you can to conform to the relevant legislation.

> **NOTE: Another BD opportunity**
>
> Did I just hear someone say they wanted proactive legal advice on industry specific areas so I can show how much I know about your business and how I might help when it comes to litigation?.

(c) **Email Archiving**

In a way, this is a Ronseal title, in that it does what it says on the tin, that is, these are systems that archive email. At a very high level the choices on offer mainly fall between having an in-house system, or using an external "cloud" based provider. Again at a high level, the first generation of offerings seem to be based around the in-house option, with more sophisticated functions and pricing coming for the second generation of cloud based products.

4.2 Collection – Forensic & Generic

Once the data you need has been identified, the next step is to collect the information. This is normally the preserve of specialist organisations who provide specific services in this area. Alternatives that normally should be avoided, are either using your own IT department to carry out the work, or letting the client's IT personnel do the collection. See Section 5.5 (b) for more on the issues surrounding these choices.

The skills required to provide forensic services are very specialised, and take time to acquire. Therefore the firms operating in this area tend to be relatively small, even when the team itself is part of a large organisation such as the forensic groups of the big consultancy firms. It has been known for clients to form working relationships with specific individuals in the forensic world and remain "loyal" to that individual even when they change the organisation they work for.

The requirements tend to split into two areas, the provision of forensic services and the collection of ESI in a competent manner, which might have to conform to evidentiary standards. As a rule of thumb, forensic services are normally needed when fraud is suspected and information has to be retrieved from mainstream computer systems and/or more obscure places such as mobile phones, tablet devices and on-line social network sites.

See Section 6.5 for draft requirements within this area.

4.3 Scanning

By and large, scanning is a commodity item with most suppliers providing similar services. Prices are normally based on scanning black and white A4 pages that have printing on one side, and creating TIFF images at 200 or 300 Dots Per Inch (DPI). Exceptions to either the colour or size of the document cause the price to increase. It is normal to get the scanning company to use a process called Optical Character Recognition (OCR) to derive a text

version of the images. Even the best OCR process will contain spelling mistakes, with a 99.9% accuracy rate equating to 4-5 miss-spellings on a single sheet of A4.

The main issue to watch out for in scanning is that of unitisation. In this specific area we are focused on how the images that result from a document are then "bound" together to form the set of "pages" representing the document. An issue can arise whereby the supplier carrying out the scanning obtains the images, but these are passed to a second supplier for coding purposes. The individuals in the second supplier look at the images and derive the unitisation of the documents from their interpretation of where document breaks occur, without being able to reference the original documents and thus seeing how they were stapled or attached via paperclips etc. together.

Scanning suppliers have two approaches to this issue. The majority use their operators to "manually" unitise the images of the pages in the files. The more sophisticated minority have pre-prepared slip sheets that they insert within pages to show things like document start and end, paper clips, attached post-it notes etc. Their scanning software recognises all these coding forms (and strips out the images of the slip sheets) and is able to re-constitute the electronic version of the file so that it exactly matches the unitisation of the paper version.

For a relatively small scanning requirement, this issue is immaterial. However for larger projects (say over 20 or so lever arch files) it can cause significant problems unless addressed.

Finally it is normal practice to uniquely identify each images in some manner so that it can be referred to. The usual methodology is to apply a BATES number to the document following a pre-arranged format.

See Section 6.3 for draft requirements within this area.

4.4 Objective & Subjective Coding

It is normal process to annotate the images of paper documents with both Objective and Subjective coding. Objective coding is that information that can be derived from the document, usually this is seven fields:

- Document Date (With an additional field to show if the date had been estimated in any way).
- Document Title.
- Document Type.
- From.
- To.
- Copyee.

Subjective coding is information that is added by a suitably qualified individual on matters such as relevance, Privilege, Trade Secret, Personal data, etc.

There is debate about the need to add Objective coding to ESI. Email does not require any additional coding as it contains the date, title and addressees within itself. Sometimes coding is added to Word documents as it can be very difficult to identify the objective information from the metadata of a Word file. However, given the volumes of ESI, such coding (even when carried out overseas by an LPO) can add significant amounts to the cost.

The issue to be aware of is that there are many different dates held within the metadata of ESI items, such as date created, date last printed, date last saved, etc. Normal practice is to use the "date last saved" as the date of the document / spreadsheet / presentation / PDF, but this might well NOT be the date shown within the item. There is no easy answer here, just a requirement to be aware of the limitations in terms of reliance on metadata for dating items.

See Section 6.4 for draft requirements within this area.

4.5 Litigation Support Tools

This Section looks at the software in this area across three broad areas. A more detailed market analysis in shown in Section 5.2 (b). whilst Section 6.6 has draft requirements within this area.

(a) Early Case/Data Assessment

Once material has been collected, it is normally passed through some form of initial assessment/culling. The tools in this area allow users to "slice and dice" the information in many ways, from removing unwanted file types, grouping by date and custodian of the information, through to more sophisticated clustering of the data using semantic analysis. This last option is where the software groups "similar" items together. The computer determines what is "similar" by conducting a linguistic analysis on the content of the ESI items as well as comparing date, time and other objective data. The power of this approach is twofold. First, it groups superfluous material so that can easily be bulk coded as irrelevant. Second, the computer had no preconceptions about issues or keywords and can sometimes produce unexpected results in its grouping that a human reviewer would not have spotted.

(b) Litigation Support Systems

The systems in this area are focused on the review of "documents" in preparation for initial disclosure to the other side, and (if the case progresses that far) a trial bundle. They have facilities for tagging documents for user defined issues, as well as powerful search tools and auditable production processes. The more recent tools have built in workflow that enables large scale reviews to be controlled in a much easier manner and "logic" checks that improve the Quality Control side of things. There are a range of tools on offer within this sector of the market, from reasonably "cheap and cheerful" up to top end products. That being said, most (if not all) now share common functionality and will enable users to deliver their end product, the key here is to match user requirement with the offered functionality and only buy enough to get the job done rather than over-specify the need.

(c) Multi-Purpose Tools

There has been an increasing trend over the past few years for both the ECA and litigation support products to "drift" into each other's areas in terms of their functionality. Which is to be expected in a maturing marketplace. However there are also starting to appear offerings that span the whole of the two areas combined. These come in two flavours. First a single product that from the start has been designed to cover the entire spectrum, and second products that started as discrete offerings (sometimes from different companies) but by a process of acquisition and re-development now share a code base and have seamless integration between each part.

The issues here are about how well such an overarching aim can be met by a single set of software developers. The requirements of an ECA tool are quite different from that of a litigation support system and the

interfaces and functionality for each area are very specific to those specific requirements. The danger here is the "Jack of all trades and Master of none" syndrome. However, as products encroach on each other's areas and as these multi-purpose tools evolve, it is probable that the successful tools of the future will span all the areas, the trick is to make sure you don't buy too soon.

4.6 Presentation Systems

The main focus in this area is the production of some form of court room bundle, rather than in-court presentation systems such as those deployed in major inquiries like the "Bloody Sunday", "Diana Inquest" or "Leveson" public events.

The main choice here is between the more traditional approach of a set of PDF files linked together into some form of eBible and more generic "cloud based" offerings. The suppliers of the eBibles tend to be those suppliers with roots in scanning, whilst the preference in the cloud based solutions seems to be a choice between products from Merrill Corporation and Opus 2.

During 2013 the Magnum product from Opus 2 gained a significant amount of "traction" in the marketplace and should be explored by anyone who ends up taking a case into the courtroom.

5. MARKET SURVEY

This Chapter looks at the overall marketplace. It starts with a quick review on the changing UK legal environment and how that might impact upon the choice and use of technology. There then is a review of the various areas that make up the sector, followed by an analysis of the different types of service providers, a look at the current "hot" topics for users of the guide to be aware of, and culminates in a detailed listing of the UK organisations grouped by type.

5.1 Changing Legal Environment

© Alex Williams. More Queen's Counsel cartoons here.

It is assumed that users of this Guide are well acquainted with the Civil Procedure Rules (CPR) that are the formal rules for this area within the UK. Of main interest in conjunction with this this guide is Practice Direction 31B - Disclosure of Electronic Documents, which took effect on the 1st October 2010 and comes with an Electronic Documents questionnaire. The point of mentioning these changes to the rules is that they prompted an increasing awareness amongst practitioners of law that electronic information existed and had to be disclosed (the real meaning of "electronic disclosure", as opposed to arguing about the media used to hand over material).

The other key driver in this area has been Lord Justice Jackson's review of Litigation Costs that was published in its final form in January 2010. The suggestions on how costs might be managed have mainly been implemented with judicial training on costs management starting in January 2013, and the required legislation enacted in April 2013. For the purpose of this Guide the key factor is that parties should arrive at the first Case Management Conference with both a plan for eDisclosure and a properly derived budget for the costs of the eDisclosure options in the case, that they have agreed with the other side. Fundamental to the changes is that Judges will have been trained to explore how the disclosure exercise will be carried out, and if the proposed approach uses technology in an appropriate manner. So the days of printing everything, or conducting a linear review with an army of legal staff, or arguing over arcane lists of broad keywords are numbered.

Lawyers need to "tool up" with technology and understand how to fully use the tools they have at their disposal. It is hoped the rest of this Guide will help them achieve this goal.

5.2 Market review

This Section of the Guide looks provides a historical review of the marketplace in two main areas. First a review of the overall evolution of the technology and second a more UK specific analysis of the changes in specific technological areas.

> **NOTE: What's this Section about?**
>
> Section 5.2 gives the background to the differences between the types of software providers. If you are happy accepting that some are "pears" and some are "apples" then go to Section 5.3. If you just want a list of the main software products go to Section 5.7.

(a) Historical Market Review - Overall

The litigation support market is a maturing one, both in terms of the software products and the firms that provide them. The evolution of the software is best tracked by looking at a technology conference / exhibition called LegalTech that takes place in New York in January/February of each year. The author has been attending the exhibition for many years, and since 2008 produces an annual review of the proceedings. The following summary is based on knowledge gained during those years as well as practical experience of different products.

2008

Products started to emerge which focused on what was called analytics software. There was evidence of the use of data sampling methodology, and emerging near dupe, data clustering and intelligent search technologies. All of which combined to increase precision and confidence in high volume document review of electronic data.

A tipping point was reached in terms of US developers adopting true Unicode compliance and delivering software that could store, search and display Russian and Chinese/Asian alphabets.

2009

The theme that permeated throughout the conference was much more emphasis on sorting out data and information processing, before you started feeding into the litigation support environment that charges on a per/GB basis. This could be as formal as a litigation readiness project, but also included more tools and emphasis on Early Case Assessment (ECA) alongside culling data on-site or via per/hour professional services.

2010

The main focus for product development appeared in two areas, either the automation of collections, via various "early" technologies, or the coding of collected items. Both of these elements of the EDRM model are ones where costs can mount almost exponentially, hence the development focus.

Within the first group of Early Case Assessment (ECA) tools, companies such as Incept, Nuix, Clearwell, Digital Reef and FTI's Quickcull emerged on the scene.

On the automated coding front, there were various offerings. The issue here seemed to be a reluctance of lawyers to accept that software can actually do an equivalent job of coding than humans, let alone an better one. There was the start of continued pressure from clients to cut costs, lending greater focus on the need to overcome such prejudice.

2011

There were to be two main themes that came through: the rise of predictive coding (in all its names and guises) and the creeping expansion of Early Case Assessment (ECA) tools, though some commentators were calling this Early Data Assessment in a pedantic, but correct understanding of the focus of the process. The significance was that, in the previous year, the majority of ECA seemed to start once the data had been collected from the corporate environment, in this year, the tools were very firmly inside the corporate set up and presumably have continued to grow into the overall infrastructure.

The two predictions for next year's "hot topics" were first, how far products will have progressed in supporting tablet based computing, as the ubiquitous iPad faces competition from service providers offering alternative hardware and software, and second will be the progress made in proving tools to gather ESI from the more difficult locations such as Facebook, Twitter, "the cloud" and large structured databases.

Finally, there was a strong trend in terms of people employing two screens on their desktops and the ability of software products to support this way of working.

2012

This was finally the event that (after several years of trying) Computer/Technology Assisted Review (CAR or TAR) became an overnight success. It seemed as almost every litigation seminar was trumpeting the virtues of the approach, with some (such as Recommind) parading clients who had used the products in anger. Since then the impetus has continued on both sides of the Atlantic with judgements being delivered in the US on employing the technology and a more quiet adoption permeating through the UK courts.

As predicted there was an explosive growth in the tools enabling content to be extracted from social media, as well as several offerings for pure tablet (well iPad) based computing.

Also there was an increasing influence from the records management stream being driven by in-house counsel using technology not only to be litigation ready, but also to drive compliance with internal RIM policies.

2013

If last year's "buzzword(s)" were Something Assisted Review, this year (from nowhere) it was "Big Data". More on this in a minute. The rise and rise of Computer Assisted Review (CAR) continued, with virtually every vendor offering a product, and (according to one supplier) the number of actual cases using the technology rising from 1-2 a month at the start of 2012, to 1-2 a week, if not a day, by the end of the year.

Nuix took a big leap into the world of Information Governance with the launch of its Luminate product and a slogan "There's no such thing as Big Data, only Small Tools". Will be interesting to see how this strategy pans out during the year.

The functionality of "All in One" tools reached the point where they could challenge a "Best of Breed" combination, but the feeling was the breakthrough would come on pricing rather than functionality, as a deep backlash against per gigabyte pricing gathered momentum.

(b) **Historical Market Review – Litigation Support Products**

To understand the differences between software products, it is necessary to look a little at the history of their development and they grew to meet different challenges at different time.

Initially litigation support tools were about dealing with scanned images of paper as this was the requirement in the 90's. The tools, such as Concordance and Summation, were basic search and review environments with additional products providing the ability to look at the images. The software evolved rapidly and a new generation of vendors appeared, with Ringtail and Steelpoint (which became IntroSpect) as the leaders in this area by the time you come into 2000 / 2001. They were still based on handling large volumes of images and struggled initially as the explosion of electronically stored information (ESI) hit them. Products emerged such as Kroll's Electronic Data Review (now Ontrack Inview), Epiq's DocuMatrix and a host of others, most of which withered, or were bought out over the years. In a reflection of the struggle going on now between more "traditional" litigation support tools and their ECA brethren, so the products initially based on images and those on ESI mimicked each other's abilities, bought out software upgrades and eventually became a more homogeneous and mature market place.

The more recent products coming to the market have learnt from existing offerings and taken the best of the functionality, but wrapped it in a far more accessible interface. This now is the main differentiator between products, on the one hand you have Ontrack Inview, DocuMatrix and Ringtail, on the other are newer tools such as Relativity and CaseLogistix that have a far more "Outlook" type look and feel. Iron Mountain's Stratify product also has a Relativity style interface.

(c) **Historical Market Review – ECA Products**

The genesis of ECA was the explosion in volumes of disclosable material brought about by electronically stored information (ESI), a shorthand

acronym for emails, Word, Excel, PowerPoint, PDF's and all the other data that organisations and individuals produce. The main concept underpinning ECA is that the software groups items of ESI together by virtue of data analytics of their contents and metadata. The initial product in this area was called Attenex (now an integrated component of FTI Consulting's Ringtail product), with the name itself meant to be "At ten times", an indication of the increase in review speed you could obtain by using the product. Over the years other products have appeared, elements of ECA have been grafted into the main litigation support products and the capabilities of the offerings have expanded, with the cost also dropping exponentially. When Attenex first appeared it had a charging model of £2,000 per GB, now ECA can be accomplished for £200-350 per GB.

Up until the Autumn of 2012, the other key player in this market was the UK based firm Autonomy. They (like Recommind) come to the arena of litigation support from a background of enterprise searching and knowledge management. Recommind realised they almost had a litigation support tool in their product and added functionality to make it work. Autonomy first bought IntroSpect to give themselves a litigation support product and re-built it around their IDOL search engine. Then they purchased iManage to give themselves a document management offering. In terms of market focus, Autonomy (more so than Recommind) was pushing for domination in the corporate environment where the three overlapping areas of Litigation Readiness (IntroSpect), Knowledge management (IDOL) and document management (iManage) exist. However the events of Autumn 2012 when HP accused Autonomy of false accounting during the takeover of Autonomy by HP means that this entire product range is now under a cloud.

The most significant current thing in the ECA world, is that it seems to have firmly breeched the corporate firewall. By this I mean that the software runs inside a corporate environment and, when required, can by issue legal holds and then incorporate these into a focused search on "held" material to identify the data that needed to be processed further. Though the whole Legal Hold concept didn't really apply in the UK, there will be some clients in litigious marketplaces that this would be of interest to. Also the UK Bribery Act is providing an impetus for organisations to look again as to how they manage their electronic information. In this area, products from Access Data, Zylab and Tunnel Vision are mentioned as ones to watch.

For the moment, the focus for the UK is on the ECA tools that mainly operate outside the end client's environment. In practice this has tended to mean one of three products, Clearwell, Nuix or Digital Reef, with (up to now) only the first two really having success. It seems that lawyers prefer Clearwell to Nuix because of its interface and seemingly better functionality, Technology departments chose the speed and performance of Nuix, over what they consider to be its slightly flashier competitor. The Recommind product range incorporates both ECA and standard litigation support tools, as does the Access Data product suite and Lateral Data’s Viewpoint.

(d) **Historical Market Review – Predictive Coding Products**

The phrase "Predictive Coding" is shorthand for any process that uses computing power and software algorithms to try and carry out coding of electronic documents. On one level the machine can carry out objective coding and scan the document for the data it can "recognise" to give you the From, To, Title, Date kind of material. So far so good. Next you get the programs that will "search" the document and highlight the terms that it thinks means the document should be relevant and even highly subjective calls such as Privilege.

However, there are caveats. The software will only really work on fully electronic material, so you cannot get the same results on the OCR of images of scanned documents. Second, most products require the user to "seed" the review work with appropriate documents that have been reviewed by a human (normally senior) lawyer, so it is not a silver bullet that will solve all your problems. Finally, no one is (yet) suggesting that the relevance and privilege reviews are totally done by the computer, the software puts forward documents that meet criteria and asks humans to validate its choices.

What is significant, is that the documents that are not selected are never even looked at. Yet this in itself ties into the UK approach to proportionality. To paraphrase the UK approach, there might indeed be a slight chance that a "smoking gun" exists in the far reaches of the potentially disclosable material. However, it can be far too expensive to review everything and so that faint possibility must not be allowed to drive the review strategy.

5.3 Vendor Analysis

The purpose of this section is to give a brief overview of the different types of service providers that users of this Guide will encounter in the marketplace. This is done so that the relationship between the software offerings and the people that supply them is understood. It is deliberately UK focused.

> **NOTE: And what's this Section about?**
>
> This gives the background and detail of the different types of suppliers so that you can understand the relative strengths and weaknesses of each type when you ask them to demonstrate their software. If you just want a list of the main UK organisations go to Section 5.6.

(a) **Background / Gartner Report**

The definitive guide to the eDisclosure marketplace is provided by the analyst firm Gartner. It has published three reviews in May of 2011, 2012 and 2013. Retailing at some $2,000, they can normally be obtained for free, via vendor's websites for the "cost" of an email address for marketing purposes. One of the key elements of a Gartner report is its use of a "Magic Quadrant" to categorise firms according to both their ability to deliver functionality and the completeness of their vision. It is a useful background document, but is very US centred and does not cover third party suppliers. That is the purpose of this Guide.

(b) **Consultancy Firms**

Up until 2010, this group was mainly the accountancy / consultancy firms in Deloitte, KPMG, E&Y, PwC, Navigant, etc. In 2010, however, two formally independent third party suppliers were bought out by firms eager to provide litigation support to their clients. Legal Inc became part of Grant Thornton and Trilantic was bought by Huron Consulting. In both cases the litigation support firm gained the financial backing and larger access to market of their "parent", though there was perceived to be a slight disruption to their ability to deliver services during the initial "bedding in" period. This dissipated during the year.

2012 saw further consolidation with 7Safe being acquired by PA Consulting at the start of the year, and in June, Palmer Legal Technology (PLT) became part of Proven, a litigation support and investigations firm.

This group can realistically be split into two "camps". In one there are the "supermodels" of the big accountancy-based firms, Deloitte, E&Y, KPMG, PA Consulting and PwC. This group are larger scale firms that tend to exist on a steady diet of work generated by the rest of the organisation. Some are interested in the bigger litigation projects, but users of this guide are more likely to encounter them when the client announces that they will be using their services and the law firm will just have to comply. In house counsel might have far more to do with this group, and will have their own opinion on the strengths and weaknesses they bring to any project. I have labelled these the "supermodels" in that the litigation projects normally have to be of a certain size to warrant them "getting out of bed". If your matter is large enough to be of interest to them they can provide a premium service at (for the main) a premium price. If it's Friday and you have 1 GB of data you need to get processed by Monday, you probably don't bother with these guys, as the conflict checks alone might take the weekend, if not longer.

In the second sub-group are the relatively smaller consultancy organisations such as FTI Consulting, Grant Thornton, Huron Legal, Navigant and Proven. I use the words "relatively smaller" as though they might not be as large as their "supermodel" brethren, they are significant organisations in their own right and much larger than the third party suppliers they compete against. These are much faster at processing their conflict checks taking hours rather than days, and are interested in smaller projects (using the "set a sprat to catch a mackerel" sales approach). It might be worthwhile approaching them on the ubiquitous Friday afternoon scenario, but do be aware they can take slightly longer to respond, as they come with their own QA process that does have an overhead. But if that is what you need, then they will be a Godsend.

In summary there are good people doing good work within the teams inside the large (and very large) organisations. There are pro's and con's with selecting a "larger" firm over a "smaller" one, these will emerge as each potential user evaluates their own requirements against the supplier's strengths in the procurement process.

(c) **Software Specific Organisations**

This group is the firms that own the software that they bring to market and thus have greater control over its development path. However, obviously, their consultancy and approach to litigation services is predicated upon their own software. In the UK, this group is split between two groups, the more numerous UK arms of US based firms, and a smaller group of UK headquartered organisations.

In the first group there are firms such as; (a non-exclusive, alphabetical list follows) Access Data, Autonomy, Epiq, Kroll, Merrill, Recommind, Stroz Friedberg and Zylab. In the second are Control Risks, PLT and eMag.

These firms are financially secure and (generally) have a depth of resources, though there is a perception that, for some of them, their project management teams always seem to be stretched. They are generally the more corporate offering as opposed the final grouping described next.

(d) **Solutions/Bureau Organisations**

The members of this group tend to be relatively smaller firms that thrive on offering a variety of solutions to end users and can match their products to the specific requirement of a case. There is an element of this grouping that came from a scanning and coding background, with others from the forensic services area. These firms are sometimes more "nimble" than their bigger brothers and can be more focused on client care, but carry a slighter greater risk in terms of their financial standing and long term viability.

(e) **Outsourcing Organisations / Document Review Firms**

There are a number of Legal Processing Organisations (LPO), with the majority of offerings based in either India or South Africa, though Integreon is alone in also having a UK base and supplying services to a number of larger law firms. These organisations tend to be software agnostic and mainly (in this marketplace) provide a cheaper option for carrying out larger scale review work. Using these organisations is normally something that is done by the experienced ligation support client, so there is some coverage of LPO's in this guide, but not a significant amount.

There are also a small number of UK organisations that provide review teams of contract paralegals and lawyers. These teams can either work off site, or come to your premises to conduct the review tasks.

5.4 Current Issues / "What's Hot"

This Section of the guide looks as those issues that are "hot" within the marketplace.

> **NOTE: Why do I need to read this**
>
> Some of these are the tools above and beyond key words, that can enable you to complete eDisclosure within a tight budget. Even if you aren't using them, the opposition might well be, so you need to know what they are talking about.

(a) Clustering

Clustering is the ability to automatically group together documents with similar content. It was pioneered by Attenex and their "petri dish" visualisation of the documents, with clusters running off a spine of a shared set of keywords. Other software companies have followed suit in terms of technology, if not the way they display it. Most useful in the context of "find me all documents similar to this one", which can enable bulk actions such as making the document set relevant, or eliminating it from the review process.

(b) Email Threading

Threading is the ability to display all the emails within a chain of correspondence as a single "thread". In its more sophisticated versions, any missing emails can be "inferred" by their presence in subsequent iterations of the chain, which might influence the collection decisions. The way in which a chain can branch out can also be captured, so that only a small number of emails have to be read in order to gain an understanding of the entire thread.

(c) Automatic Language Translation

Some programs have the ability to automatically translate a number of common languages. Most can spot that the text is in a foreign language, but the ability to translate is a little less common. No one is suggesting that the translation is of evidential quality, but normally it is enough to enable an initial view to be taken in terms of relevance etc. Practical experience has thrown up one or two issues. The software does not cope well with the presence of two languages in a document, you can get a number of "false positives" with say an English email which has some French words in its address footer, being wrongly categorised as "French". Also sometimes the document is correctly identified as being in a different language, but (if the module for that language is not installed) then it is arbitrarily categorised as some other country.

(d) Audio Files

A real growth area. This is the ability of certain programs (Aurix and Nexidia being two market leaders) to index digital audio as if it was text and then provide functionality enabling you to search in a similar manner. So, if you have an hour of a recording, the software will take you to the 30 second slot, some 45 minutes in, that contains the words "inside dealing". A specialist tool for specialist projects, but an absolute Godsend if you

have thousands of hours of digital material to listen to. Further improvements allow the production of text, so that you can read the conversations rather than listening to them.

(e) **Computer/Technology Assisted Review**

This is a topic that has caused a tremendous amount of debate. In essence the computer is trained to recognise relevant/Privileged documents and once fully trained is set loose on the rest of the document set. The fears were seemingly based on a misconception that the computer would replace lawyers, when the truth is the opposite. The computer might replace the "grunt" reviewer be they UK paralegals or overseas lawyers, but is totally dependent upon the experienced lawyer(s) being involved in the training and monitoring process. In a way, this is a moot point as service providers are reporting that the process is being actively embraced, with the number of cases using the technology rising from 1 or 2 a month at the start of 2012, to 1 or 2 a day by the end of it. If you aren't using this technology, then your opponent most probably is.

See Section 3.13 above for more detail on this.

(f) **Collection of data from Social Media environments**

As electronically stored information proliferates into different areas, so the ability to collect it from within those environments becomes more urgent. Various suppliers are developing tools (or buying up companies that have done the development) to enable them to hook into Facebook, Twitter, Yammer, et al and extract information in a meaningful way. Again, the need for this functionality will depend on the area the matter is within, but increasingly data stored in social media is becoming important in more and more cases.

(g) **Small Quantities of ESI**

See Section 6.7 below. This is a constant requirement that comes through every sized procurement and articulates the real need for users to "just read the emails". The key problem is that information is passed to lawyers in electronic format, yet (for very good reasons as far as the IT department is concerned) they are not allowed to use the firm's environment to review it. So, they are sat there with a PST of a small number of email, an email with 50 or so Word attachments, or a thumb drive with a couple of thousand items and they "just want to read them". Providing a quick and easy solution to this requirement will be a real game changer for the various suppliers.

(h) **Charging Model**

Just as lawyers are coming under increasing pressure on prices, so service providers are being stressed by their clients. The default model is that people will charge you by volume, so much per GB at various stages of the process. Increasingly users are looking for a fixed price solution so they have clarity of costs to pass on to their clients. In response to this, some suppliers are offering a "managed solution" option that guarantees fixed pricing for users, irrespective of individual case volumes.

See Section 6.1 for more discussion on this.

(i) **Redaction tools for "Native" Formats**

A bit of a specialist requirement, but one that could be significant if you really, really need it. What we are talking about here is the ability to redact (that is blank out the offending text, and remove it from all search capabilities) areas within things such as Word, Excel and PowerPoint documents. Normally this involves a cumbersome process of turning the "native" item into a PDF version and then redacting the PDF, but for things such as Excel spreadsheets this is not very workable. A handful of service providers are now starting to supply toolkits that let you redact within the "Native" mode. However, if you have Privileged information within a note on an Excel spreadsheet and the opposition has convinced the judge that you must supply the document in its original Native mode, this could be a lifesaver. My normal rule of thumb is that some 0.2% of documents in a collection end up being redacted, and they are Word files to start with, just how crucial the ability to redact Excel spreadsheets really is remains to be seen.

5.5 Potential Problems

This sub-Section details some of the more common issues that can be encountered in using litigation support systems. It is not to say that a software package with one of these problems is automatically excluded from your procurement, there are very often workarounds, but you need to know the issues exist so you can factor them in to your evaluation criteria.

(a) **Assigning Privilege to a single attachment in an email Group**

WARNING: This can cause serious amounts of delay and cost

If there is one single issue you need to be aware of when selecting software, this is it.

In the United States it is possible to claim Privilege over an entire email family (that is an email with one or more attachments). This is not the case in the UK and other jurisdictions. However, some of the software packages treat the email family as a single entity and do not allow you to split out attachments because they are privileged. This can cause significant overheads at production time and should be an issue you are well aware of when selecting software, and engaging service providers.

(b) **Re-unitisation of Images of Paper Documents**

Most of the software on offer comes from a background of handling electronic information, emails, Word documents and the like. The one thing a piece of Electronically Stored Information (ESI) never does, is change its boundaries; it is what it is. Compare this with scanning, storing and coding paper based images. With the best will in the world, there will be time when the images that make up a paper document need to be re-unitised, that is the coding that encompasses say 6 pages, needs now to be split into two sets of coding, one for the first three pages and one for a second document of the last three pages. Why is this an issue? Surely the answer is just to split up the images in the software and change the coding as needed. This is where you hit the mind-set of the R&D team for

ESI based software. They have no concept of the boundaries changing and so have little, or no functionality for re-unitising paper-based records. Paradoxically the "ancient" software of Concordance and original Summation could do this with no problems as they came from a paper-based background, it is the newer "kids on the block" that have problems.

This will not affect you, unless you have significant amounts of paper to process for your disclosure exercise, but if you are in that situation, explore with your service provider how they will deal with this.

(c) **High level allocation of alias for Names Normalisation**

The issue here is the variety of names that appear during collection of emails. Not only do you get people who have different email hosts, so;

andrew.haslam@allvision.co.uk, andrew.haslam@gmail.com, andrew.haslam@etc

Also in Outlook you have the option for a "Display as:", where you can edit the text in the "Display as" field. Many people tend to differentiate between people's personal and work email addresses, so they change the text in the "Display as:" field to reflect this, so the entry with an email address of;

andrew.haslam@allvision.co.uk, could be displayed as Andrew Haslam (Work)

Now when an email is collected, the email address shown is Andrew Haslam (Work) not andrew.haslam@allvision.co.uk.

Also if you are collecting email from within an organisation, you can get the SMTP version of this that has all kinds of letters, brackets and punctuation.

Most Early Data Assessment tools are equipped to deal with this issue and will allow you to pick a set of names to search on. So if I was trying to get all emails sent by Andrew Haslam, I could tick the boxes to get the all the variants of my name. After a while this gets really boring, particularly when you want to start doing searches of email sent to and from a group of people, each with 4 or more versions of their email address.

What is required is a facility to have a single alias to which all the variants could be assigned and then you can far more easily conduct complex searches. Alternatively you ensure that the service provider covers all the bases and your role is simply to specify what is required and identify the people you wish to search on.

(d) **Data Collection by Client or Law Firm's IT Department**

It is strongly recommended that clients are encouraged NOT to go down this route.

Data collection is not simply a matter of copying an item of ESI. If you don't know what you are doing when you copy something, you can

unintentionally change all the metadata associated with a document. What does this mean in the real world?

Take this real life case as a warning of what can go wrong in these circumstances. One set of clients used to present monthly reports to their board using a PowerPoint slide deck that had Excel spreadsheets underpinning all the graphs. The dispute revolved around actions that had taken place in 2006, so copies had been made by someone (client's IT department, incompetent supplier, some gremlin along the way) of the 2006 PowerPoint shows some time in 2010. Except they hadn't been forensically copied, and all of the shows now had a date displayed on the first slide of sometime in 2010, not the original correct 2006 date. So there we were in 2012, coming late to the case, relying on other people's efforts and evidence, and the other side kept demanding we give them the 2006 documents and all we had were "tainted" versions with no way of now collecting the originals.

The proposal to self-collect data normally comes from a client desperate to keep their costs down. Make sure they are well aware of the potential dangers before you let them do this. In close second, comes the lawyer, also keen to cut costs who volunteers their in-house IT team to get the information. In most cases, a law firm's IT department does not have the expertise, the time nor the professional indemnity insurance to be going anywhere near a data collection. Avoid it and get a professional to do the job, then, if it does all go wrong, their insurance can take the hit, not your reputation.

(e) **Issues of working in "Native" formats**

Most litigation support platform have viewing tools that let you look at Word, Excel and PowerPoint documents without firing up the original software. This is fine for a quick glance, but of no use at all for real review. In a number of the real life cases I've been involved in, the text that makes a document Privileged has been contained in the Track Changes comments in a Word Document. (There's a whole Section's worth here on organisations that hand over Native documents without scrubbing this kind of data, but that's for another day). Similarly unless you look at the formula's and workings of Excel, how can you begin to understand the purpose of the spreadsheet.

The answer to this used to be that people would offer up PDF versions of the ESI. Nowadays that won't cut it and will be resisted (very strongly) by any half-awake opponent. You need to be aware of the "iceberg" of issues that collecting and review Native data brings, and (at the very least) have protocols built into your review platform so you can see reviewers have downloaded the native document to review it. Plus, that the people doing the review have the technical skills to do things like look in Word Track Changes, or know how to remove the "hide" command in Excel.

5.6 Vendor list

The following table shows the main organisations in the UK, grouped under the headings used in the previous Section. It is shown to give a start point for readers to decide whom they wish to reach out to, for demonstrations and/or quotations. A more comprehensive listing of the service providers and their software products can be obtained by downloading The UK Buyer's Guide to Litigation Support Systems from here.

The assignment of organisations has been a relatively arbitrary process as some suppliers could lay reasonable claim to being under more than one heading. It has been done purely to try and group like with like, and is no reflection upon competence or suitability to task. It will be up to the Reader to decide what their requirements are, and then to approach the firms that best meet those needs.

A number of organisations provide services across the EDRM model, so just because their forte is in Forensics does not de-bar them from providing a review platform, and vice versa. You need to read the rest of the Guide to work out what your requirements are, and which kind of organisation best meets your needs, or indeed, if you should interview across the range of service providers to get a feel for their relative strengths and weaknesses.

Consultancy Firms (Large)		
7Safe, (PA Consulting)	Deloitte	Ernst & Young
KPMG	PwC	
Consultancy Firms (Medium)		
AlixPartners	Consilio	FTI Consulting
Grant Thornton UK	Huron Legal	Navigant Consulting
Palmer Legal Technologies	The Oliver Group	
Software Specific Organisations		
AccessData Group	Control Risks	Epiq Systems
Integreon	Kroll Ontrack	Merrill Corporation
Recommind	Stroz Friedberg	Symantec
ZyLAB		
Solutions / Bureau Organisations		
Altlaw	Hobs Legal Docs	LDM Global
Legastat	Millnet	TransPerfect
Unified	Xerox Litigation Services	
Forensic Services		
CCL Group	CY4OR	eMag
Guidance Software	MD5	
Outsourcing Organisations / Document Review Firms		
Capita	Exigent	i-Lit
Integreon		

5.7 Software list

Following on from the principles used in the previous Section, shown overleaf is a list of the major software products available within the UK and the organisations that supply them. Again the headings used to group the products are an outline rather than hard and fast differentiations. The aim is to allow readers to identify what products are in a similar category so they are able (as much as possible) to compare like with like.

No endorsement is made or implied for any of these products and their capabilities do evolve relatively quickly, so contact the supplier(s) to find out more.

Product	UK Supplier(s)
Collection	
FTK	Access Data
Guidance EnCase	CCL Group, Integreon
Processing and Early Case/Data Assessment	
Clearwell	AlixPartners, CCL Group, Deloitte, Epiq, Ernst & Young, Grant Thornton, Hobs Legal Docs, Integreon, KPMG, Navigant, Symantec, Unified.
Digital Reef	TransPerfect Legal Solutions.
ICE™	Palmer Legal Technology (part of Proven group)
In Control	Navigant.
Index Engines	Deloitte, Integreon, Millnet.
IPRO (eCapture)	Hobs Legal Docs, Integreon, Legastat.
KOFAX Capture	Legastat.
Law PreDiscovery	AlixPartners, Legastat, Navigant.
MM/PC	eMag Solutions.
Nuix	7Safe, AlixPartners, Altlaw, Deloitte, Epiq, Ernst & Young, Grant Thornton, Huron Legal, KPMG, Legastat, Millnet, PwC, Unified.
Review & Analytic Tools	
Case Logistix	Hobbs Legal Docs, Legastat.
Concordance FYI	AlixPartners, Hobs Legal Docs, Huron Legal, Integreon, Legastat.
Equivio Relevance	AlixPartners, Epiq, Legastat, Millnet. Equivio is often embedded into other litigation support products (particularly Relativity) to provide "Predictive Coding" functionality.
iCONECT	AlixPartners, Huron Legal, PwC.
Merrill Lextranet	Merrill Corporation.
Relativity	7Safe, AlixPartners, Altlaw, Deloitte, eMag, Epiq, Ernst & Young, Hobs Legal Docs, Huron Legal, Integreon, KPMG, Millnet, Navigant, PLT, PwC, TransPerfect Legal Solutions, Unified.
Ringtail	FTI Consulting, Grant Thornton.
Xera	Integreon
"All in One Products" (Both ECA and Review)	
AccessData (Summation)	Access Data
Consilio Product Suite	Consilio
DocuMatrix	Epiq Systems
eTrium™	Control Risks
Integreon Tools	Integreon
Ontrack Advanceview & Inview	Kroll Ontrack
Recommind Suite	AlixPartners, Recommind.
Stroz Software	Stroz Friedberg
Viewpoint	LDM Global, Xerox Litigation Services (XLS)
ZyLAB Systems	ZyLAB
Digital Audio Review Tools	
Aurix	Aurix
Nexidia	7Safe, Deloitte, KPMG.

6. PROCUREMENT APPROACH

This Chapter proposes a procurement approach that can be used in selecting vendors. It is anticipated that users might wish to cut and paste text from the chapter into their own documents so the word [CLIENT] has been used to enable a "find and replace" with firm specific details/name.

It provides an analysis of the types of pricing models suppliers might adopt, an overview of the procurement process and then specific requirements for:

- Scanning.
- Unitisation and Coding Services.
- Data Collection.
- Litigation Support Services.
- Processing small volumes of ESI.

BEST PRACTICE: When should I involve a service provider?

As soon as possible in specific cases. Many law firms have been through procurement exercises and set up call off arrangements with a selected litigation support partner. They have an in-built advantage when it comes to eDisclosure in the Jackson era.

6.1 Supplier's Pricing / Client Tactics

The purpose of this Section is provide an overview of the various factors that affect service providers' pricing so that the reader is better equipped to compare like for like. It is included in the Guide because this is a very difficult area, so be prepared for some work when you obtain quotations.

If the three most important things about buying a house are location, location, location, then the three most important factors about costing eDisclosure projects are volume, volume, volume. That is to say, ESI is notorious for becoming a black hole of ever increasing volumes, and as such it can be difficult to price the work. That being said there are some words of advice that can help.

(a) **Involve the supplier as soon as possible**

The sooner a service provider knows what the project is about, the quicker they will be able to give you a price, and the more likely they are to be able to give you a more fixed cost. In an ideal world, readers will have selected a preferred supplier, (with a fall back in case of conflicts), familiarised themselves with the software and its capabilities, and have a call-off contract in place with a table of pricing. Failing that, you should at least have 2-3 known suppliers that you can call upon at short notice. If you telephone someone out the blue at 5 o'clock on a Friday afternoon and say you have 200 GB of data that you have to review by Monday, you will get a very different price than if you have pre-prepared the process.

(b) **Tell them as much as possible about the case**

A good supplier will want to come and talk to you to explore all the aspects of the case; where the data is stored, what technical "gotcha's" might be lurking in the shadows, any obscure software or products that have known issues (Lotus Notes for example, is a favourite of corporates

and has a lot of pitfalls for the unwary). The more they know about the task, the more they can advise you on the best way forwards and the more flexibility they will have on pricing. Treat them as you would any technical expert. They have a deep range of knowledge, that you need on your side.

(c) **Understand the supplier's constraints**

Service providers have three main components to their cost base, these are the costs of; storage, software and their professional services.

Storage in the UK is typically charged from between £50 - £100 per GB, per month. Some organisations rent their storage space, and thus are constrained on how much flexibility they have on pricing, others own their environment and thus have more room to manoeuvre. There are still some significant margins in this area, so be prepared to press hard on pricing here, but equally be prepared to divulge how much work you might be putting their way, what's the average size in ESI terms of your cases, how many litigation cases you do a year, and those kind of things. You do have that information at your fingertips don't you?

Suppliers that use third party software that they don't own, can have limited room in terms of pricing. For example people offering Relativity (the current market leader in the Review space) are constrained by kCura (the people that supply Relativity) on what they have to charge on a per GB basis. The more firms process through Relativity, the less the unit cost per GB is, so the larger vendors might have more "wiggle" room in this area.

Finally there will always be an element of professional services. The software tools need to be managed, and there is a need for some overall project management. Be aware that when a vendor provides project management, it will not include managing the project from your end (unless you are specifically paying for these services). That normally ends up the role of some hapless junior associate who spends more time being an ad hoc litigation support manager then they do practising law.

(d) **Ask for flexibility in pricing**

Once you have built a relationship with a supplier, then you can explore what flexibility they can provide in pricing. Just as the courts and clients are looking to pass risks on to the law firms, so you could be looking to pass some of that onto your litigation support partner. Note the deliberate use of the word partner. You cannot get a good price by ringing round the suppliers and seeing who will give you the cheapest quote for "processing 50 GB".

Service providers are well aware of the implications of the Jackson reforms and should be looking to build relationships with you and give as firm a pricing as they can. The "worst" case should be a straight per GB price, but this is becoming increasingly rare. Vendors have a range of pricing plans they have on offer. Some will not charge for the data going into the processing stage, only the material that emerges from the other end. Some will offer you a range of options for specific stages, and agree

only to charge you the cheapest once the work is done and they can individually price what the components cost them. Some will ask for a retainer and then provide a fixed amount of capacity and processing for that price, which (if you are pushing the limits of the deal) can be good value.

The shorter version is that there are many ways to "skin this cat", ask for them and then work out what suits you best.

(e) **Be prepared to do some work in comparing pricing**

If the good news is that, there are many options for pricing, then the bad news is that, there are many options for pricing. You need to allocate some time and effort to getting the information and then putting it into a form so that you can compare like with like. It is not a simple matter of ringing three firms and asking them for their price to process 50 GB. You will tend to get three different sets of questions and then three different pricing models. If you haven't set up a preferred vendor arrangement, then the best way to proceed, might be to have an example requirement, that mirrors one of your cases, complete with all the background information and assumptions, then pass that to the vendors so that you can get an idea of their pricing. Be prepared to build and maintain a spreadsheet for this.

You will have sensed by now the common theme running through this section of involving service providers as early and as fully as possible in the case. Try to build a partnership with one of the many competent firms out there and (in the main) you won't go wrong. You will certainly be in a better position than those who don't bother and leave it to the last possible minute.

The rest of this Chapter takes you through how to procure suppliers and software.

6.2 Overall Requirement/Approach

A generic procurement exercise will typically go through the following stages :

- Requirements scoping.
- Match potential vendors to requirement.
- Assess vendor capability.
- Short list.
- Demonstration with real data.
- Establish call off contracts.

> **NOTE: How much do I really need to do?**
>
> As little or as much as you want. The aim is to end up with a preferred supplier, whether for one case or for many. Some firms have followed a full procurement route taking weeks, others have asked 3 vendors in for demonstrations on one day and made a selection after that. The absolute minimum you must do is look at the software with the people who will support it, if you can use your own data for the demo, even better.

The main considerations for each of these stages are shown in the following sub sections.

(a) **Requirements scoping**

All "Stakeholders" in the procurement should be interviewed. This not only includes the lawyers/legal staff in the litigation department, but might also embrace the legal teams from other disciplines. It should definitely include the IT department and (if one exists) the print room function. You would be amazed (or perhaps not, if you are undertaking a procurement) as to how much stuff is still printed and photocopied.

The firm's accounting system might be examined to see which suppliers are already being paid for services, as well as a firm wide question on which suppliers people have heard about/used, and which they recommend/would never use again.

The aim of the exercise is to distil down the requirements of the various parties into a coherent whole, so that competing products can be assessed in an "apples" for "apples" manner.

One thing that will be needed is an estimate of the volumes of litigation that the firm undertakes. Firms should extract details from the their Practice Management System on the number of new litigation matters created each month, and couple these with estimates on the number of "live" matters, a rough idea of how much each is worth, and how long the cases run for. In this way they will have some idea of the overall scope of their requirement, and also vendors will understand the approximate value of the work they might obtain, and can price their offering accordingly.

> **NOTE: How much work have you got?**
>
> If you don't have a good idea (or even a rough one) of the volume of work you might give a supplier, then you are at a distinct disadvantage when negotiating for the best deal.

(b) **Match potential suppliers to requirement**

The aim here is to select a pool of potential service providers that match the requirement, with the trick being the elimination of those suppliers that do not "fit" the needs or ethos of the firm. This is a two way process, the vendors themselves are selective in their sales qualification process and will only focus on those prospects they have a real chance of winning. If your approach is too wide, you run the risk of alienating the marketplace, and only getting the desperate and the naïve bidding for your work.

(c) **Assess supplier capability**

Assessing supplier capability can be done in a number of ways. More formal procurements go through a cycle of Request for Information (RFI) to gather market knowledge, followed by a much more detailed Request for Procurement (RFP), which would contain the types of requirements shown the following Sections.

Some firms prefer to go direct to a short list and straight into the formal RFP, whilst others will skip this stage entirely (with some advice) and invite a selected group in for demonstrations.

All approaches are valid, it depends upon the budget for the procurement exercise, the level of formality the firm requires and the scope of the requirements.

(d) **Short list**

If a more formal RFP (and RFI) has been issued then the response will need to be evaluated, given some form of scoring (with or without weighting, depending upon the firm) and a short list selected. In cases where the circumstances warrant it, firms might move straight to this stage, relying upon their ability to identify which suppliers might meet their requirements.

(e) **Demonstration with real data**

The next stage is key. Firms must evaluate the software they are considering using by means of a demonstration, preferably using the firm's own data, ideally from a real life case, so that the issues with it are known. The panel looking at the products should be consistent across all demonstrations, and feedback should be collated in a managed format, so that personal bias is reduced as much as possible.

References might also be taken up at this stage. Ultimately, however, it is the interaction between end users, the software and (most importantly) the vendor's staff, that is of greatest significance.

(f) **Establish call off contracts**

Once suppliers have been selected, the call-off contracts with discounted rates should be negotiated. It is better to use the law firms contract as a starting point, though the service provider T&C's should be examined. The trap to avoid here is the over eager in-house lawyer looking to make up their hours who sees this as an exercise to re-write everything from scratch.

6.3 Scanning

This Section explores the requirements for scanning services, which are mainly commodity items, with some room for added value in terms of the approach to project management and unitisation.

(a) **Overall Considerations**

It is expected that documents will mainly comprise single sided black and white A4 pages. There might be some colour pages, and document sizes other than A4. These are normally dealt with by an exception process.

The documents will normally be provided from [CLIENT] offices, though occasionally they might need to be collected from clients. The collection and transport of files/documents to and from the supplier will be the responsibility of the supplier.

It is proposed that the scanning work will be carried out at the supplier's site, however, it is the responsibility of the supplier to meet [CLIENT] security and document-tracking requirements in order for this option to be followed.

An optimal process needs to be identified which will ensure accurate document scanning including physical unitisation at a document level.

The documents to be scanned will normally be provided in banker's boxes. Within these boxes, the documents will be sub-divided into individual files that vary in size, type and binding.

(b) **Example Requirements**

The activities to be completed by the selected provider are:

- Document / file preparation.
- Physical unitisation.
- Electronic scanning.
- 100% Quality control.
- Reassembly of documents / files.
- Objective coding and logical unitisation.
- Electronic BATES numbering – post scanning and quality control processing.
- Delivery of single-page tiff images with unitisation and coded information for loading into the litigation support environment.

In order to carry out these activities the provider will need to:

- Enter into a call-off contract with [CLIENT].
- Dedicate the required facilities at their site.
- Resource the required operations.
- Provide an experienced project manager to manage individual projects.
- Provide periodic reporting on project status.
- Manage the project budget to ensure costs are monitored and controlled.

(c) **Possible Quotation Format**

Please provide pricing for the following items.

Scan, OCR and BATES number 100,000 pages. The pages are contained in 333 files, each file holding 300 pages, estimated 3 pages per document. If charged separately please shown the cost of preparing the pages and then re-constituting the files, and the provision of logical unitisation information.

Please provide the cost for the provision of each of the services shown below. [CLIENT] have "weighted" these costs as an overall percentage of processing 100,000 pages to reflect their real life experience of the frequency of these requirements being needed. For example, roughly

15% of the pages encountered by [CLIENT] also need glasswork copying before they can be scanned, whilst other services occur at even smaller percentages.

The services are:

- Glasswork A4 B/W Scan
- Large Format Scan (i.e. A3 and above)
- A4 Colour Scan
- Glasswork A4 Colour Scan
- Large Format Scan Colour
- Re-Binds (Documents that have to be un-bound before they can be scanned and then reconstituted in a bound state)

Suppliers should also quote for the cost of providing load files, DVD and any project management / professional services surrounding the scanning.

If there are any other costs associated with the provision of scanning services, suppliers should show them in their completed quotations.

A potential layout for a quotation spreadsheet is shown below:

All prices exclude VAT			Vendor :		Vendor Name
Bidding for provision of these services :					Yes / No
Item/Activity	**No**	**Unit**	**Unit Price**	**Total**	**Suppliers Comments**
A4 B/W Scan	100,000	Pages		£0	
OCR	100,000	Pages		£0	
BATES Numbering	100,000	Pages		£0	
Physical Unitisation	33,333	Docs		£0	
File/Doc Handling	33,333	Docs		£0	
Glasswork A4 B/W Scan	15,000	Pages		£0	
Large Format Scan	20	Pages		£0	
A4 Colour Scan	5,000	Pages		£0	
Glasswork A4 Colour Scan	200	Pages		£0	
Large Format Scan Colour	40	Pages		£0	
Re-Binds	100	Docs		£0	
Load File Creation	10	Per File		£0	
DVD/CD	10	Per DVD		£0	
Technical Labour/PM	2	Days		£0	
					Spare for supplier use
					Spare for supplier use
					Spare for supplier use
			TOTAL :	**£0**	
Assumptions					
Pages / Document	3				
Pages / File	300				
Docs / File	100				
Total Docs	33,333				
Total Files	333				

6.4 Unitisation and Coding Services

The key consideration within this Section is the decision on whether to objective code ESI items or not.

(a) Overall Considerations

There are advantages in using the firm that conducts the scanning to carry out the coding work, principally the overall control and cost saving that ensue.

The key to keeping prices down is to have the work carried out overseas. Most firms will offer a UK based service, but the cheaper rates come from

shipping the work overseas. The service offerings do need to be explored, as some firms conduct their coding work in locations where the speed/quality of the internet connection is so poor, that the data needs to be transferred to that country for the coding work to be carried out. Other firms use organisations in countries where this is not an issue, and the data remains in the UK.

(b) **Example Requirements**

[CLIENT] require a supplier who will provide objective coding for [both] paper documents [and ESI]. The supplier will provide the environment in which the coding takes place.

[CLIENT] will allow the data required for the objective coding process to reside in an overseas location.

For the paper documents, suppliers will receive physical unitisation information including attachment groups and will apply logical unitisation alongside objective coding

The following coding fields will be required:

Paper Documents

- Document Title.
- Document Type (taken from a list provided by [CLIENT]).
- Document Date (either taken from document or by deduction from context within the original file.).
- Estimated Date (Yes/No field, if Yes, supplier and [CLIENT] to agree convention used).
- Author.
- Recipient.
- Copyee.

Word efiles

- Document Title.
- Document Date (either taken from document or meta-data).
- Author.
- Recipient.
- Copyee.

Excel and PowerPoint efiles

- Document Title.
- Document Date (either taken from document or meta-data).

The Author / Recipient / Copyee fields for paper and efiles should show the name(s) and organisation(s) in brackets as shown below.

e.g. Andrew Haslam [Allvision Computing], Fred Smith [Client PLC].

A process of names normalisation will need to be carried out.

It is assumed that the document type data for all efiles will provided as a bulk update in the export process, i.e. all Word efiles will have a document type of “Word”.

In order to carry out these activities the provider will need to:

- Enter into a call-off contract with [CLIENT]
- Dedicate the required facilities at their site.
- Resource the required operations.
- Provide an experienced project manager to manage individual projects.
- Provide periodic reporting on project status.
- Manage the project budget to ensure costs are monitored and controlled.

6.5 Data Collection

The skills required to provide forensic services are very specialised, and take time to acquire. Therefore the firms operating in this area tend to be relatively small, even when the team itself is part of a large organisation such as the forensic groups of the big consultancy firms. It has been known for clients to form working relationships with specific individuals in the forensic world and remain "loyal" to that individual even when they change the organisation they work for.

(a) Overall Considerations

There are two main types of data collection:

- Forensic services, such as imaging hard drives, copying data in a forensically sound manner (for example in fraud cases) or retrieval of deleted information.
- Data collection in accordance with the guidelines embodied in Senior Master Whitaker's ESI collection questionnaire.

(b) Example Requirements

The services to be provided by the selected supplier should include:

- Forensically sound imaging of individual PC's, laptops, Apple Macs.
- Provision of a "data map" describing where the ESI is located throughout the client's IT infrastructure
- Collection of selected ESI from client's infrastructure.
- Remote collection of selected ESI for overseas clients either by hardware or software based solutions.
- Sound knowledge of the data protection requirements in various countries.
- Support to [CLIENT] in exploring the collection of ESI from external devices such as pen drives, mobile phones, external hard drives, etc.
- Support to [CLIENT] in exploring the possible collection of ESI from on-line applications such as email systems or social networking software.

- Processing of back-up tapes.

In order to carry out these activities the provider will need to:

- Enter into a call-off contract with [CLIENT]
- Dedicate the required facilities at their site.
- Resource the required operations.
- Provide an experienced project manager to manage individual projects.
- Provide periodic reporting on project status.
- Manage the project budget to ensure costs are monitored and controlled.

6.6 Litigation Support Services

This Section is the main focus of the Guide and contains a significant amount of detail.

(a) Overall Considerations

The following list of requirements is an extensive one. It is intended that users of this Guide will select those requirements that meet their needs and shape their procurement accordingly. The initial approach should be to determine what tools are available within the marketplace, and what their functionality might be. Service providers should not be disqualified from the selection process if they cannot supply some of the more advanced requirements, indeed users might not have need of that functionality for some time to come.

(b) Example Requirements

The services to be provided by the selected supplier include:

- Provision of processing facilities to enable de-duplication and filtering of data.
- Processing of Outlook, Lotus Notes and GroupWise email files.
- Any visualisation of the connections between data custodians.
- Processing, display and review of family groups of ESI files.
- Processing, display and review of scanned and coded documents.
- Ability to display information across dual screens.
- Provision of OCR from imported image files.
- Ability to store, display, search and review PDF files.
- Handling of foreign language based ESI.
- Near-duplication.
- Concept clustering / searching.
- Any other analytics functionality.
- Email threading.
- Predictive / Technology Assisted / Computer Assisted coding.
- Handling of digital audio files.

- Automatic translation of foreign language ESI.
- Provision of objective and subjective coding (including issue/hot topic) fields.
- Full search capabilities, including Boolean logic, saved searches and dynamic search update.
- Disclosure production capabilities, including BATES numbering, production of Privilege and Trade Secret lists, mass burning to external media.
- Production of a courtroom bundle.
- Strategy for meeting the demands of "tablet" based computing.

In order to carry out these activities the provider will need to:

- Enter into a call-off contract with [CLIENT].
- Dedicate the required facilities at their site.
- Resource the required operations.
- Provide an experienced project manager to manage individual projects.
- Provide periodic reporting on project status.
- Manage the project budget to ensure costs are monitored and controlled.

6.7 Processing small volumes of ESI

Most firms have a need to "just read the emails" and this Section attempts to articulate that requirement.

(a) Overall Considerations

A requirement has been identified to provide a mechanism whereby users in the firm can review relatively small amounts of electronic information. A classic scenario, is that a client will ask for an evaluation of a potential matter, often on the assumption, that they will either not be charged for this initial evaluation, or it will be a nominal fee. This need is not limited to the litigators in a firm, most lawyers/legal staff receive information in an electronic form and wish to quickly examine it. The issue is that it is difficult, if not impossible, to put this data into the firms in-house systems, as the IT department actively discourages ephemeral client data being added into the sanitised, backed-up, secure environment they provide for the lawyer's use.

The broad outline of the proposed approach is as follows:

- Vendors provide a solution with specific boundaries in terms of size of data, number of users and timeframe with a low price attached. For example, up to 500 MB of data, with one person looking at it, for one month at £50-100/month up to a maximum of 3 months.
- There is some form of easy to use mechanism whereby the Client information, PST file, email attachments, USB stick can be securely transferred to the vendor.

- The vendor loads the data into their litigation support environment without any additional processing and stores it under a client/matter reference.

(b) **Example Requirements**

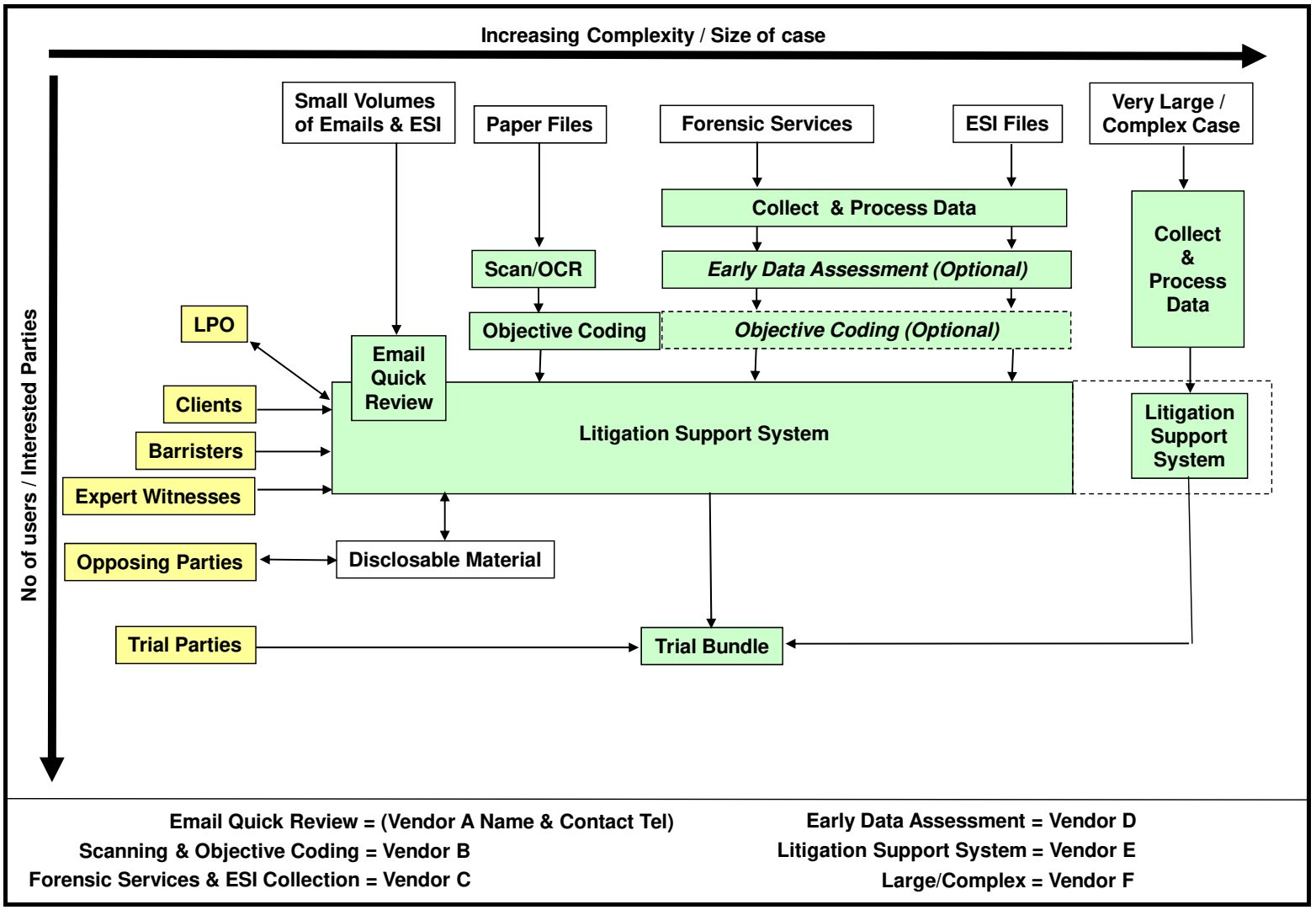
The services to be provided by the selected supplier will include:

- Provision of means to securely transfer the data between [CLIENT] and the supplier, probably using a secure file transfer mechanism known as "secure FTP".
- Loading of email and electronic files into the selected product in a speedy manner, (the aim will be an overnight service).
- Provision of tools to allow the initial reading and analysis of documents.
- Provision of review tools to meet the needs of a "quick peek" scenario.
- A well-defined route to export the data so that it can be processed and loaded into the "full blown" litigation support tool.
- Very competitive pricing.

In order to carry out these activities the provider will need to:

- Enter into a call-off contract with [CLIENT]
- Dedicate the required facilities at their site.
- Resource the required operations.
- Provide an experienced project manager to manage individual projects.
- Provide periodic reporting on project status.
- Manage the project budget to ensure costs are monitored and controlled.

6.8 Schematic of Generic Requirements



6.9 Summary

This Chapter has provided readers with the information for them to undertake a procurement exercise. As with all such exercises, the key to a successful project is a clear understanding of what the requirements are, and how best they might be met. Various potential requirements are shown, so that a selection can be made of the ones that best meet the user's needs. Once that framework is in place, and a shortlist of potential suppliers identified, then the proposed software solutions must be evaluated by means of a demonstration to a representative panel, preferably using the firm's own data.

TECHNICAL GLOSSARY

ACTIVE OR LIVE DATA: Information residing on a computer's hard drive or servers which is readily visible to users (e.g. a document, spreadsheet or an e-mail).

ALGORITHM: A detailed formula or set of steps for solving a particular problem (e.g. searching for relevant electronic documents, such as, MDS# or SHA-1#)).

APPLICATION: A collection of one or more related software programmes that allow a user to enter, store, view, change or extract information from files or databases (e.g. Word, Excel and Microsoft Office). Also referred to as "programmes" or "software".

ARCHITECTURE: Hardware and/or software comprising a computer system or network.

ARCHIVAL DATA: Information that is not directly accessible to the user of a computer system but is data that the organisation maintains for long term storage and record keeping purposes (e.g. backup data).

ATTACHMENT: A record or file associated with another record for the purposes of retention or transfer. The attachment is commonly referred to as the "child" with the record it is attached to as the "parent". If the attachment itself has an attachment this would be a "grandchild" and so on. A synonym is an **ATTACHED DOCUMENT,** which means a Document attached to, or embedded in, a **HOST DOCUMENT.**

AUDIT TRAIL: Information about where data has been, in whose possession and why, held in sufficient detail so as to allow the reconstruction of that activity.

AUTHOR: The person, office or designated person responsible for a document's creation or issuance. Also referred to as "originator".

BACKUP DATA: A copy of data created as a precaution against the loss or damage of the original data. Backup data is information that is not presently in use by an organisation and is routinely stored separately upon portable media, to free up space and permit data recovery in the event of disaster. Backup data can be incremental (where only new data is saved) or complete (where all data is saved).

BACKUP TAPE RECYCLING: The process whereby an organisation's backup tapes are overwritten with new backup data, usually on a fixed schedule (e.g. the use of nightly backup tapes for each day of the week with a daily backup tape for a particular day being overwritten on the same day the following week; weekly and monthly backups being stored offsite for a specified period of time before being placed back in rotation).

BATES NUMBERING: is used in the legal, medical, and business fields to place identifying numbers and/or date/time-marks on images and documents as they are scanned or processed, for example, during the discovery stage of preparations for trial or identifying business receipts. Bates stamping can be used to mark and identify images with copyrights by putting a company name, logo and/or legal copyright on them. This process provides identification, protection, and automatic consecutive numbering of the images. The process is named after the late 19th century inventor Edwin G. Bates of New York City.

BYTE: The basic measurement of most computer data.

CD-ROM (CD READ ONLY MEMORY): Data storage medium that uses compact discs to store about 1,500 floppy discs worth of data, that is, approximately 55,000 pages. Variations include CD-Rs (CD Recordable) and CD-RWs (CD Re-Writable).

CLUSTERING: Functionality whereby ESI containing similar content is grouped together by the software without human intervention. Results might be shown in a pictorial manner with items of ESI "clustered" together, or by folders of similar documents.

COMPRESSION: The reduction of the size of a file to save storage space. "Compression ratio" is the ratio of the size of an uncompressed file to a compressed file.

COMPUTER ASSISTED REVIEW (CAR): Also known as T**echnology Assisted Review (TAR)**. A process of having computer software electronically classify documents based on input from expert reviewers, in an effort to expedite the organization and prioritization of the document collection. The computer classification may include broad topics pertaining to discovery responsiveness, privilege, and other designated issues. Also see: **Predictive Coding**.

COMPUTER ASSISTED REVIEW REFERENCE MODEL (CARRM): Model used to show stages of process of Computer Assisted review (CAR).

COMPUTER FORENSICS: The use of specialised techniques for recovery, authentication, and analysis of electronic data.

CSV FILE: A computer file containing a list of values separated by a comma or other delimiter.

CUSTODIAN: Person having control of a network, computer or specific electronic folder.

DAT (DIGITAL AUDIO TAPE): A high capacity storage medium. Used in some backup systems.

DATA MAP: A written description (possibly with a diagram or two) of where the client's data sources are.

DATA SAMPLING: See **SAMPLING**.

DE-DUPLICATION: The process of identifying and removing duplicate Documents from a collection of Documents so that only one unique copy of each document remains. A cryptographic hash function such as the Message Digest algorithm 5 may be used to generate a digital fingerprint for an Electronic Document. The digital fingerprint of a Document can then be electronically compared against the digital fingerprint of any other Document to determine whether the Documents are exact duplicates. De-duplication may also be implemented by using a cryptographic hash function applied to a group of Documents.

DELETED DATA: Data that, in the past, existed on the computer as live data and which has been deleted by the computer system or end-user. Deleted data remains on storage media in whole or part until it is overwritten by on-going usage or "wiped" with a software program specifically designed to remove deleted data. Even after the data itself has been wiped, directory entries, pointers, or other metadata relating to the deleted data may remain on the computer.

DELETION: The process whereby data is removed from active files and other data storage structures on computers and rendered inaccessible except by using special data recovery tools designed to recover deleted data.

DISC (DISK): It may be a floppy disk, or it may be a hard disk. Either way, it is a magnetic storage medium on which data is digitally stored.

DISCLOSURE DATA: Data relating to disclosed Documents, including for example the type of document, the date of the document, the names of the author/sender and the recipient, and the party disclosing the document. See **OBJECTIVE** and **SUBJECTIVE CODING.** Normally only **OBJECTIVE CODING** is disclosed with documents.

DISTRIBUTED DATA: Information belonging to an organisation which resides on portable media and non-local devices such as remote offices, home computers, laptop computers, personal digital assistants (PDA's), wireless communication devices (e.g. Blackberry) and internet repositories (such as email hosted by internet service provider or portals and web sites).

DOCUMENT: Anything in which information of any description is recorded (see CPR Rule 31.4). It includes all **ESI.**

DOCUMENT CODING: The process of identifying and recording case-relevant information (e.g. author, date authored, date sent, recipient, date opened, etc.) from a document. Can be automated or manual. Also referred to as **INDEXING**. See also **OBJECTIVE CODING** and **SUBJECTIVE CODING**.

DOCUMENT MANAGEMENT: The manual and automated processes for the management of documents during the course of proceedings, including the identification, preservation, collection, processing, analysis, review, production and exchange of documents.

DVD (DIGITAL VIDEO DISC OR DIGITAL VERSATILE DISC): Data storage medium, like a compact disc, upon which data can be written and read. DVDs are faster, can hold more information, and can support more data formats than **CD**s.

EARLY CASE ASSESSMENT (ECA): Also known as "**EARLY DATA ASSESSMENT**". Initial process in the EDRM approach whereby a large volume of data (normally emails and attachments) goes through

various processes such as clustering, semantic analysis, and email threading to enable early decisions to be taken on the relevance of ESI.

ELECTRONIC DATA DISCLOSURE (EDD): Also known as **EDISCLOSURE.** Process of disclosing ESI. Not to be confused with using electronic means to carry out the disclosure of images of paper documents or printed out emails, Word documents etc.

ELECTRONIC DISCOVERY REFERENCE MODEL (EDRM): Model used to show stages of process of electronic discovery.

ELECTRONIC DOCUMENT: see **ELECTRONICALLY STORED INFORMATION (ESI).**

ELECTRONIC IMAGE: an electronic representation of a paper document or Electronically Stored Information. An Electronic Image may be a **SEARCHABLE IMAGE** or an **UNSEARCHABLE IMAGE**. Examples are image **PDF** files and **TIF (/TIFF)** files.

ELECTRONIC STORAGE SYSTEM: A system or medium for retaining Electronically Stored Information.

ELECTRONICALLY STORED INFORMATION (ESI): Electronic files on a computer such as emails, Word, Excel, PowerPoint, Adobe PDF documents. It includes (for example) e-mail and other electronic communications such as SMS and voicemail, word-processed documents and databases, and documents stored on portable devices such as memory sticks and mobile phones. In addition to documents that are readily accessible from computer systems and other electronic devices and media, it includes documents that are stored on servers and back-up systems and electronic documents that have been 'deleted'. It also includes **METADATA** and **EMBEDDED DATA**.

EMAIL THREADING: Software functionality that pulls together the various emails that make up a "thread of conversation" and display them in an easy to understand manner. The normal aim is have the final email in a chain readily identifiable so that all the secondary emails in the conversation can be read in one pass.

EMBEDDED DATA: Text or other information which is not typically visible to the user viewing the output display on screen or as a print-out. Examples of Embedded Data include spreadsheet formulae (which display as the result of the formula operation), hidden columns, externally or internally linked files (e.g., sound files in PowerPoint presentations), references to external files and content (e.g., hyperlinks to HTML files or URLs), references and fields (e.g., the field codes for an auto-numbered document), and certain database information if the data is part of a database (e.g. a date field in a database will display as a formatted date, but its actual value is typically a long integer).

ENCRYPTION: Procedure whereby the contents of a message or file are scrambled or made unintelligible to anyone not authorised to use it.

FIELD: A section of data in a database, for example a field containing the date of a document.

FILE SLACK SPACE: A form of residual data, slack space is the amount of on-disk file space from the end of their logical record information to the end of the physical disk record. Slack space can contain information soft-deleted from the record, information from prior records stored at the same physical location as current records, metadata fragments and other information useful for forensic analysis of computer systems.

FORENSIC COPY: An extract copy of an entire physical storage medium (hard drive, CD-ROM, DVD, tape etc.). Also referred to as "mirror imaged copies", "imaging" and "disc mirroring".

FORMAT: The way in which Electronic Images and other documents are stored and made accessible.

GIF (GRAPHIC INTERCHANGE FORMAT): A computer compression format for pictures.

GIGABYTE (GB): A measure of computer data storage capacity and equivalent to a billion (1,000,000,000) bytes. Also referred to as a "gig".

HARD DRIVE: The primary storage unit on PCs, consisting of one or more magnetic media platters on which digital data can be written and erased magnetically.

HOST DOCUMENT: A Document with one or more **ATTACHED DOCUMENTS**. For example, an e-mail is a Host Document and any Documents attached to the e-mail are its Attached Documents.

INDEXING: See **DOCUMENT CODING**.

INTERNET SERVICE PROVIDER (ISP): A business that provides access to the Internet.

JPEG (JOINT PHOTOGRAPHIC EXPERTS GROUP): An image compression standard for photographs.

KEYWORD SEARCH: A search for documents containing one or more words that are specified by a user. Normally conducted on **ELECTRONICALLY STORED INFORMATION**, but can also be carried out on **OCR TEXT**.

KILOBYTE (KB): A measure of computer data storage capacity and equivalent to a thousand (1,000) bytes.

LEGACY DATA: Information that has been created or stored by the use of software and/or hardware that has become obsolete or has been replaced ("Legacy Systems").

LEGACY SYSTEMS: Systems containing legacy data.

LITIGATION HOLD: An instruction issued as a result of current or anticipated litigation, audit investigation or other such matter that suspends the normal processing or disposal of records.

LITIGATION SUPPORT SOFTWARE/SYSTEM: Application that supports the process of litigation. In terms of the EDRM approach this stage occurs after the Early Case Assessment stage.

LOOSE DOCUMENT: An Electronic Document that is stored in its Native Form in a file system or directory system but not an email box. An email or document attached to an email, even if extracted from the email box in which it was originally stored, is not a Loose Document.

MEDIA FREE SPACE: Unused space on storage media that is available for storage.

MEGABYTE (MB): A measure of computer data storage capacity and equivalent to a million (1,000,000) bytes. Also referred to as a "meg".

METADATA: Commonly described as "data about data". It is information that may describe, for example, how, when and by whom it was received, created, accessed, modified and how it is formatted. Some metadata, such as file date and sizes, can easily be seen by users. Other metadata can be hidden or embedded and is unavailable to computer users who are not technically adept. Metadata is generally not reproduced in full form when a document is printed.

MIGRATED DATA: Information that has been moved from one database or format to another.

MIRROR IMAGE: Used in computer forensic investigations and some electronic disclosure investigations, a mirror image is an exact bit-by-bit copy of a computer hard drive that ensures the operating system is not altered during the forensic examination. May also be referred to as "disc mirroring", or as a "forensic copy" or "imaged copy".

MPEG (MOVING PICTURES EXPERT GROUP): An image compression standard for full motion video.

NATIVE FORMAT: An associated file structure for an electronic document defined by the original creating application. Viewing or searching documents in the native format may require the original application (for example, viewing a Microsoft Word document may require the Microsoft Word application).

NETWORK: A group of one or more computers and other devices connected together for the exchange and sharing of data and resources.

OBJECTIVE CODING: Coded information that can be derived from a document without any specific legal training. Normally comprises; Date, Estimated Date, Document Title, Document Type, From, To, Copyee. Objective coding is normally conducted by a vendor (often overseas to provide a cheaper service).

OFF-LINE DATA: The storage of electronic data outside the network in daily use that is only accessible through the off-line storage system.

Optical Character Recognition ('OCR'): means the computer-facilitated recognition of printed or written text characters in an Unsearchable Image

OFF-LINE DATA: The storage of electronic data outside the network in daily use that is only accessible through the off-line storage system.

ON LINE DATA: Electronic data stored on the network in daily use.

PDF (PORTABLE DOCUMENT FORMAT): A common format for images of documents which enables documents to be displayed or printed a manner which preserves the formatting originally used by the author. A PDF file may be either a Searchable Image file or an Unsearchable Image file.

PETABYTE (PB): A petabyte is a measure of computer data storage capacity and equivalent to one quadrillion (1,000,000,000,000,000) bytes.

PERSONAL DATA: Information of a personal nature that must not be disclosed, such as medical records, salary, home addresses, relationship discussions, social security numbers, etc. Personal data is normally **REDACTED**.

PREDICTIVE CODING: Functionality that automatically codes records by conducting analysis on the ESI. The coding can encompass **OBJECTIVE** and **SUBJECTIVE CODING**. Objective coding is usually a simpler process than the Subjective work which requires the software to be "seeded" with examples of relevant and/or Privilege documents. The application then "learns" what criteria it uses to arrive at the Subjective decisions and (once trained) will identify those documents and pass them to a user for confirmation on the coding calls. See also: **COMPUTER ASSISTED REVIEW REFERENCE MODEL (CARRM)**.

PST (PERSONAL STORE): The place where Microsoft Outlook stores its data (when Outlook is used without Microsoft Exchange Server). A PST file is created when a mail account is set up. Additional PST files can be created for backing up and archiving Outlook folders, messages, forms and files. The file extension given to PST files is .pst. Can be broken down into individual emails called .msg.

RETENTION PERIOD: The length of time a given records series must be kept, expressed as either a time period (e.g. four years), an event or action (e.g. audit), or a combination (e.g. six months after audit).

REDACTION: The process whereby sensitive information is hidden by rendering part of a document unreadable. It is sometimes referred to as 'Masking'. Redaction is typically used to render unreadable; confidential, privileged or personal data portions of an otherwise disclosable document.

RESIDUAL DATA: Data that is not active on a computer system (sometimes referred to as "Ambient Data"). Residual data includes (1) data found on media free space; (2) data found in file slack space; and (3) data within files that has functionally been deleted, in that it is not visible using the application with which the file was created, without use of undelete or special data recovery techniques.

RESTORE: To transfer data from a backup medium (such as tapes) to an on-line system, often for the purposes of recovery from a problem, failure, or disaster. Restoration of archival media is the transfer of data from an archival store to an on-line system for the purposes of processing (such as query, analysis, extraction or disposition of that data). Archival restoration of systems may require not only data restoration but also replication of the original hardware and software operating environment. Restoration of systems is often called "recovery".

SAMPLING: Usually (but not always) refers to the process of statistically testing a data set for the likelihood of relevant information. It can be a useful technique in addressing a number of issues relating to litigation, including decisions as to which repositories of data should be preserved and reviewed, and determinations of the validity and effectiveness of searches or other data extraction procedures. Sampling can be useful in providing information to the court about the relative cost burden versus benefit of requiring a party to review certain electronic records.

SEARCHABLE IMAGE: An **ELECTRONIC IMAGE** in which the text-based contents can be searched electronically.

SEMANTIC ANALYSIS : Method by which a number of products conduct clustering. Refers to the "automatic" identification of key words and concepts within a document so that there is a "spine" of a central concept, off which related groups of documents are clustered.

SCANNING: The process of converting a hard copy paper document into a digital image for use in a computer system. Often associated with the **OCR** process, as in "documents will be scanned and subject to an OCR process".

SUBJECTIVE CODING: Information held in a litigation support system about records (either paper or electronic). Subjective coding requires legal input and covers initial calls on Relevance, Privilege and Trade Secret as well as case specific issue and matter coding.

TECHNOLOGY ASSISTED REVIEW (TAR): See: **Computer Assisted Review (CAR)** and **Predictive Coding**.

TERABYTE (TB): A measure of computer data storage capacity and equivalent to one trillion (1,000,000,000,000) bytes.

TIF OR TIFF (TAGGED IMAGE FILE FORMAT): One of the most widely supported file formats for storing bit-mapped images. Files in TIFF format often end with a .tiff extension. Other file formats include JPG and BMP.

UNATTACHED DOCUMENT: An **ELECTRONIC DOCUMENT** without a **HOST DOCUMENT**.

UNSEARCHABLE IMAGE: An **ELECTRONIC IMAGE** in which the text-based contents cannot be searched electronically.