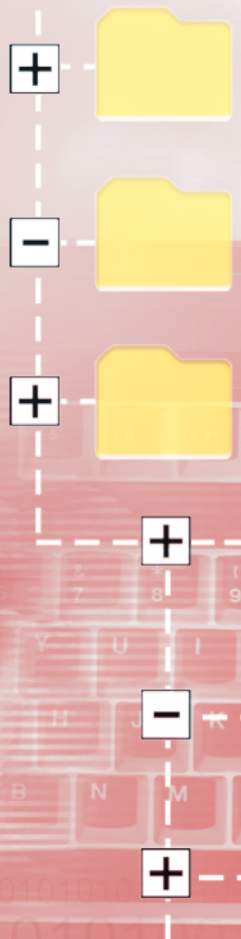
ANDREA M. DE ROSA



True and false problems of electronic CRFs

This book is based on the experience gained in years of use of e-CRFs for clinical data collection. Its aim is to provide food for thought and provide professionals with a deeper understanding of the problems that might arise from the actual use of this electronic tool.

Andrea M. De Rosa has worked in the medical IT field since 1981, when he presented a dissertation for his Medical degree by writing a software on one of the first "Midis" for the conversion of HPLC figures. This program was subsequently used by the Pharmacology Institute of Milan for over 15 years.

After working for several years as a medical executive in the pharmaceuticals division of two multinationals and as president of a healthcare advertising company, in the mid-90's he began to explore how the Internet could improve clinical trials data collection.

In 1998 he opened an experimental web-division at the communications agency which he owned.

In 2000 he founded Airon Telematica, a software-house specializing in the collection and management of clinical data in GCP (e-CRF), working in close contact with CRO and/or Sponsors.

In recent years he collaborated with several scientific boards in the Design and implementation of multicenter observational studies (retrospectives and perspectives), as well as with the Istituto Superiore di Sanità in the collection of epidemiologic data in specific sub-populations of patients.

Since 2008 he's been giving a yearly 'lectio magistralis' on e-CRFs (design, implementation, regulatory and practical aspects) at the 'Post graduate Master in Clinical Research' hold by the University Milano Bicocca (Faculty of Medicine).

He also teaches this subject at several Advanced Courses in Clinical Trial organized by scientific organizations and associations (e.g. SFFA – Society for Applied Pharmacological Sciences).

Dr. De Rosa also writes on specialized journals and contributes to scientific associations.

Andrea M. De Rosa

Real and false problems of electronic CRFs

INDEX

INTRODUCTION

There is no doubt that the world is facing an epoch-making change, (possibly greater than the industrial revolution). The widespread diffusion of new & powerful, instruments; is changing deep-rooted habits and creating new ones.

As with all changes, this too should be interpreted and understood, if we wish to "take the bull by the horns" rather than spending our lives running away from it.

However, in light of its world-scale dimensions, speed of diffusion and the all-pervasiveness, the Internet has already claimed a lot of victims, damping initial enthusiasm and strongly affecting modern society in the place where it hurts most: i.e. one's pocket (and Wall Street it bubble is there to remind us of this).

Speculation on revolutionary forecasts created the premises for one of the biggest global bluffs: the birth, the rapid growth and the undignified crash of numerous Internet start-ups which, though accurately predicting that things would be changing, nevertheless got their timing (and their investment plans) wrong: the 1990's ended in broken dreams and in a veritable bloodbath on the new information superhighway (2,3).

But the Internet is here to stay: this is apparent today, only a few years after the great flame-out, and is likely to become even more apparent during the next decade.

Although almost every production area has been affected by the telematic revolution, its ensuing effects and advantages are difficult to grasp: the intensity of change is such as to make traditional parameters inadequate, thus requiring the formulation, fine-tuning and, most difficult of all, the on-site verification, of others.

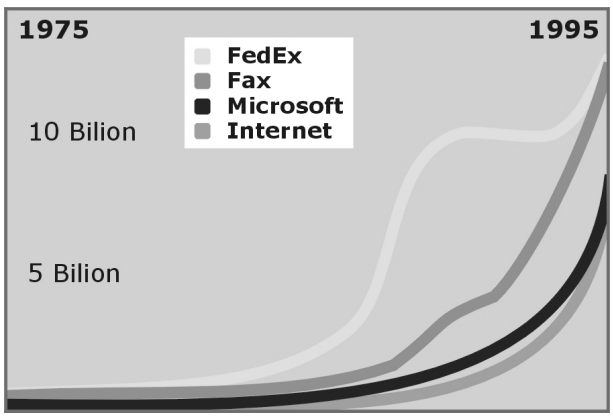


Figure 1.
Net based developments show a long incubation followed by rapid exponential growth: this is linked to the need to reach a 'critical mass' which then begins to self-fuels.

E-CRF OPERATING DEFINITION

Throughout this book, the term e-Clinical Report Form (e-CRF or, web-based-CRF) refers to a clinical data collection system that uses the Internet and its standard protocols.

The clinical data collection system can be divided into 3 main sections:

1. A centralized section which includes the server(s) holding the database, the application program and its interface.
2. A remote section, consisting of an unspecified (more often, unknown) number of PCs (potentially, any PC connected to the Internet); from each of these, authorized users can interact with the central database feeding system via the aforementioned interface. No specific software needs to be installed on these PCs: a standard browser (e.g. Explorer or Netscape) is all that is needed and, nowadays, it is always pre-installed on all of them. Nor is the operating system used by the PC itself relevant: any version of MS Windows, Apple, Linux or Unix will do the job perfectly.
3. A connection between all of the above elements: this is done by the Internet and three of its standard protocols/languages (TCP, HTTP/HTTPs, HTML).

In this type of data collection system, starting from the very beginning and for the entire duration of the study, the data are housed in a relational database specially designed to contain them (a study-specific database) and managed by data experts, in a secure environment.

Clinical investigators are assigned the task of feeding the database in real time, using a study-specific interface, capable of providing a guided fill-in procedure and preventing the most banal errors (obvious errors); at the same time, descriptive statistics allow study leaders to monitor, in real time, all data collection progress.

In order to use this type of e-CRF, an authorization is needed (similar to that given by banks to clients requesting a Home Banking service) as well as personal access codes.

As can be seen, in an e-CRF of this type, there is no need to periodically download data from the PCs of the Hospital Centre to central systems, nor to periodically 'synchronize' data originating from different Centres.

On the other hand, the CRO responsible for monitoring the study can avail itself, in real time, of invaluable tools that check the clinical data collected for internal consistency (coherence controls) established study by study; as a result, all "potentially inaccurate or erroneous information" will make itself visible, even emerging from an enormous mass of data.

With just one "click", an integrated query generator system is capable of generating pre-filled-out mails, addressed to the investigator who has entered "dubious" data.

On the contrary, the term e-CRF does not include all IT tools based on data collection performed locally, either on Investigators' PCs (e.g. Excell or Access files distributed with a fill-in request), or dedicated systems based on non-standard protocols.

In short

In short what distinguishes web based e-CRFs from other electronic data collection tools is the possibility of working from any PC, connected to the Internet, without having to install any specific program.

All of this takes time, i.e. the very element that proved to be lacking in internet start-ups and in their 2-3 years R.O.I. plans.

But time passes and with it comes experience.

From this point of view, in the highly specific field in which we operate (e-CRFs, see Box 1 for definitions), the last few years have proved to be extremely helpful in terms of acquiring experience.

The first "experiments" enabled us to understand that some problems, connected to the use of paper-CRF, ceased to be problems in the electronic world.

On the contrary, other types of problems (often totally unexpected) arose, thus requiring the preparation of specific strategies to counteract them.

In the hope that the experience acquired would also prove useful to others, we decided to give our findings a concrete form as a means of contributing to the debate on this issue.

DELOCALIZATION

While the possibility of sharing scientific documents in real time was the basis of the civil development of the Internet, the concept of delocalization is only now becoming evident and has not yet been fully grasped by the majority of users (4).

And yet, we are dealing with an extremely powerful concept which will affect the way in which we use the Internet (not only professionally) over the next few years.

By "delocalization", in this context, we mean:

> "It no longer matters WHERE a piece of information is kept as long as it can be accessed at any time from anywhere"

This is absolutely innovative: for centuries, localization of an 'item' and the ability to access it were almost synonymous; from the very beginning, if you wanted to access/use a document you had to have it physically in front of you.

Even though everybody now knows that, with the Internet, localization of information and ability to access it are two separate things, this notion has not yet been fully grasped by users: after all, the burning desire to "download" musical files, videos and images from the Internet expresses a sort of immaturity in the system (and of users) rather than one of its long-term characteristics.

Think about it: why would we download, for example, a piece of music from the Internet to our PC?

Let's try to find some reasonable answers.

1. The fact that downloading it might require a fair amount of time; time that we might not have when we actually feel like listening to the piece. Therefore, we download it so that it's there (localized on our pc) when we need it.
2. The desire to copy it onto supports other than our PC (or the Hi Fi system connected to it); in fact, after downloading it, we can copy the piece onto a CD and listen to it in the car; or on a small mp3 player so that we will be able to listen to it outdoors or in any other place that takes our fancy.
3. The fear of not finding the same piece in the same place at a future date.
4. Last but not least, simply to satisfy a collector's desire (e.g. to possess all the pieces recorded by a certain singer or group)

The first three answers are directly linked to the immaturity of the system as a whole: in fact, the time needed to download tends to be drastically reduced as use of broadband becomes more and more popular, thus eliminating the reason given in point 1).

As regards point 2), CDs are likely to become obsolete in the future (do any of you remember vinyl disks or tape recorders?), possibly being replaced by multi-media players connected to their own broadband and capable of playing anything available on the net at any time.

Point 3) and its "fear" will fade away because, in the future, numerous copies of the same piece will be available on numerous servers (redundant information).

Point 4) will, obviously remain because it is the only one that the Internet will never be able to influence. In the case of the collector, the localization of any object is, and will continue to be, essential: she/he must actually own it (i.e. possess it), in order to enrich his/her collection.

ADAPT OR REDESIGN?

Delocalization and information sharing are the cornerstones of the Net, which allow for impressive improvements in the collection of valid, verifiable clinical data. On the other hand, validity and verification are indispensable premises for the use of this type of new technology in clinical research.

The first practical approaches were,, of an adaptive nature. By this I mean the attempt to merely translate the 'paper' procedures (used for the last forty years) into electronic procedures. The first results obtained, satisfactory though not particularly exciting, enabled us to understand that the process of adaptation was not optimal, since it also involved "adapting" a series of procedures, which would, simply, no longer have any reason to exist in an electronic format; these were procedures that provided specific solutions to the specific problems that arose when using the paper format.

Bad adaptations

The worst example of the adaptive use of a PC that I have ever come across, dates back to 1984: the first PC, complete with a spreadsheet system (Symphony, the forerunner of today's Excel and Lotus), was introduced into an office where 4 typewriters were already in use.

By switching on the PC and launching the program, you could print an empty spreadsheet: the printer, after a lot of bits and other noise, returned an A4 page, which only contained the standard grid of the spreadsheet itself (in short, a grid consisting of empty cells).

When, after some time, I returned to that office and asked whether they were pleased with their new "monster", they gave me a fairly disconcerting answer: "Quite, although we had to buy a new, expensive typewriter with a reduced pitch because, with our standard ones, there wasn't enough space to fill in the cells that we had just printed".

It took me a few seconds to understand what they were referring to and then I finally got it.

The secretary turned on the PC, the monitor and the printer: she launched the Symphony program and gave the "print" command without typing in any data. Then she took the printed page, with all its empty cells, and placed this sheet in the typewriter in order to fill in the data in the cells.

The problem was that the characters in their 'old' typewriters were too big, which made it almost impossible to write anything in the small empty grid cells printed by the PC.

This was the solution that they came up with: buy a new typewriter with a smaller typeface, so that the grid sheet could finally be completed with the necessary data, and possibly photocopied and distributed according to the method that had been used until then.

No-one in that office had ever considered that :

- it might be possible to change the grid to the desired size
- that the real advantage lay in the fact that the program would have accepted the data and automatically made all the necessary calculations
- all of the data could have been entered once and once only and then saved on the disk provided with the PC, both for storage and possible reuse at a later date

Another, much more recent example (1999) concerns e-mails: in an attempt to solve some kind of technical problem, I heard a secretary being given the following instruction:

"Please write X an e-mail but don't send it. Just print it and fax it to the following number…"

An important lesson

It is easy to make fun of "errors" such as this. But it is more difficult (and much more useful) to try and understand how intelligent individuals, equipped with all the necessary skills to carry out tasks that they have always carried out, suddenly find themselves incapacitated by silly problems such as these.
When looking for the root cause, we often discover that it is related to bad adaptations of old habits to new technologies.
There can be no doubt that this is a losing strategy, because it does not take one basic fact into consideration: old habits have been developed and consolidated because they were capable of offering a valid solution to real requirements.
In other words, old habits represent a means, whilst the achievement of objectives constitutes an end.
Of course, there are changes in context that enable one to adapt old habits to new circumstances: it is a question of changes that do not alter the foundations of previous operating methods but only require a few new touches.
The technological revolution is so widespread that, even in Clinical Research, it is certainly advisable to re-examine objectives and ask oneself what strategy will give the best results in the new context.

CONSIDERABLE ADVANTAGES

The aim of this book is not to outline the truly considerable advantages of using e-CRFs for multi-centric studies rather than traditional data collection systems. With regard to this, please refer to the short article published on AboutPharma, from which the table is taken (5).

Name	e-CRF	Traditional CRF
Information transfer speed	Seconds	Days or months
Phisical CRF size	Null	Often hudge
Data availability	Immediate	Days of months
Phisical data security	Granted	Site-dependent
Data tracking	Unavoidable	Required
Monitoring Trial progression	Real-Time	Delayed (site visits)
Data coherence check	Immediate	Delayed
On-line data sharing	YES	NO
Compliance with the study protocol	Granted	Required
Warnings linked to particular events	Automatic	Manual
Ability to convert and normalize different Units for lab exams	Automatic/Immediate	After completion
Selective information supply	YES	NO
Protocol variations (amendments)	Immediate and general	Requires new CRF distribution
Queries generation	Automatic	Manual
Descriptive statistics	Immediate (real-time)	Days of months
Drug accountability	Built-in	Manual recording
Site Visits main activities	Problem-solving	Problem-finding+problem-solving

Table 1.
Comparison between web-based e-CRFs and traditional (paper) CRFs. As can be easily seen, an e-CRF allows for a far better quality and coherence of the data collected.

TRUE AND FALSE PROBLEMS

As you have probably realized, many of the false problems related to e-CRFs are the result of somewhat illogical actions: i.e. the attempt to 'translate' habits and uses from paper to electronic media.

The good news is that, in almost all cases, such a 'translation' is not needed at all: in fact it is impractical in terms of achieving objectives.

However, when reorganizing strategies, several new and often unexpected problems may arise – problems that are linked to the specific aspects of the technological medium.

If not correctly handled, grasped and solved, these problems can make it difficult (or even impossible) to achieve the Study objectives.

So, in some respects, it is easy to understand why, over the last years, several researchers have opted to reject the use of new technologies in favour of the "good old systems".

But time goes by and new problems are dealt with and solved as they crop up.

And the time comes when the new "tried and tested" tool is finally able to show its true colours (in terms of a much more efficient achievement of end results).

In Clinical Research, this is finally becoming a reality of the new millennium, even though the wholly electronic implementation of all Clinical Research will take at least another ten years.

IT SECURITY

As shown by the huge amount of money spent in Internet purchases over the last few years, it is apparent that the Internet is capable of offering both banks and financial dealers the necessary level of security. The question here is: is this level of security also suitable for the transmission of confidential clinical data?

In my opinion, the answer is yes, for at least 3 reasons.

- From a hacker's point of view, although clinical data can be relevant, they are less desirable than cash. Generally speaking, those intent on committing a crime, do it for money. Anyone capable of violating a system obviously finds it preferable to violate a banking system. The fact that these violations rarely occur shows that the protocols and methods used are valid: this is particularly true following the liberalization, by the US government, of the "strong" encrypted (SSL3 – 128 bit) system, formerly considered a "strategic secret" and strictly reserved for use in the USA only.
- It is almost impossible to interact with a secure system without leaving a trace (see additional information on this subject further on in this chapter): although these traces are of little help to financial dealers (interested not only in identifying the thief but also and, above all, in retrieving the amount that has been stolen) they, hold enormous deterrent potential when clinical data are involved (unlike money, these data do not "disappear" when they are "stolen" and identification of the "thief" is all that is required to take action).
- Apart from inefficiency or fraudulent activities on the part of authorized users (see "Real problems: security and management of Passwords"), to all intents and purposes, data collected and safely stored can be considered locked away in a safe and only available to those with accessing rights. The same applies to any Clinical Study involving several Authorized Investigators, regardless of the medium (paper or electronics) used.

Data traceability

Good Clinical Practices (GCPs) strongly stress data traceability: by this term we mean the possibility of knowing whether data was originally entered as such or modified in any way: if it was modified, it is important to know exactly when which value was originally entered and, above all, who made the change and why.

The need for certainties in this sense is the underlying reason for the habit of printing the paper CRF on carbon paper. Once the copies have been separated, it is impossible to make amendments to either one of them without this being apparent during comparisons.

THE VIOLATION OF THE CIA WEB SITE

On September 18, 1996, at about 4.45 a.m., N.Y.C. time, some Swedish hackers broke into the Central Intelligence Agency's home page (http://www.odci.gov/cia/): they altered it, proclaiming that the Agency was the "Central Stupidity Agency".
At about 7.30 a.m. the hacked page was taken down by the CIA, which declared that they had no idea when the service would be back on-line and that, in any event, there had been no security breach of any confidential files or documents.

It would appear that the designated victim was Bo Skarinder, to whom an appeal was made to stop lying. Skarinder was one of the key Swedish prosecutors in a case against various telecommunications companies (including the Swedish conglomerate Telia).
During that same period, articles on similar types of vandalization began to appear:
"Internet break-ins have become an increasing concern for U.S. defence and intelligence agencies. Last month, hackers broke into the Justice Department's Web site, adding swastikas, obscenities and a picture of Adolf Hitler to the page. The Department pulled the plug on the vandalized page and assured that the hackers did not gain access to criminal files". (From: Simson Garfinkel - 'Web Security, Privacy & Commerce', O'Reilly)

Standards traceability

From this point of view, on account of its very nature, use of an IT system is always traceable, down to the very last detail. During a recent presentation, I told my audience that data traceability is "inevitable" when using e-CRFs.

In other words, while with studies on paper it is necessary to take specific precautions (i.e. carbon paper) to guarantee the traceability of the data, in electronic studies specific measures need only be taken if one wishes to eliminate said traceability. What's more, if the data collection procedures adopted are standard and oriented towards storage of the information, it then becomes impossible, even for the system analysts who manage the databases, to hide or eliminate elements of traceability without leaving some sort of sign. For those who are interested, the example below showa the traces left following each user and e-CRF interaction.

As can be noted, these traces are multiple and clearly evident. In fact, their removal entails a number of operations which, in turn, leave traces that arouse suspicion about their absence.

1. *Firewall log files*. The first element ensuring the security of a Clinical Research central network is undoubtedly the firewall. A firewall is a device that limits access between networks in accordance with local security policies. When a request is accepted, the firewall registers the identifying elements of the operation itself and well as those of the session to which this operation belongs (Fig. 2).

```
15:01:27 INFO/ACCT: INET: 08.06.2004
15:01:03 0 6 82.49.85.121:1394/18001 ->
151.99.182.84:80/1000 7 1357 6 2351
15:01:27 INFO/ACCT: INET: 08.06.2004
15:01:04 0 6 65.54.164.109:38019/18001 ->
151.99.182.79:80/1000 6 469 5 4274
15:01:27 INFO/ACCT: INET: 08.06.2004
15:01:08 3 6 64.140.49.68:47801/18001 ->
151.99.182.84:80/1000 5 465 3 825
15:01:27 INFO/ACCT: INET: 08.06.2004
15:01:03 0 17 80.21.163.156:1028/18001 ->
151.99.182.78:53/1000 1 71 1 87
15:01:38 INFO/ACCT: INET: 08.06.2004
15:01:04 17 6 82.49.85.121:1395/18001 ->
151.99.182.84:80/1000 15 1679 20 22019
15:01:38 INFO/ACCT: INET: 08.06.2004
15:01:17 0 17 212.95.252.16:43089/18001 ->
151.99.182.78:53/1000 1 62 1 139
15:01:49 INFO/ACCT: INET: 08.06.2004
15:01:24 0 6 82.49.85.121:1397/18001 ->
151.99.182.84:80/1000 5 563 4 502
```

Figure 2.

Example (a small portion) of a log file generated by the firewall: for each connection session, you can view, the IP address of the calling client's pc, that of the called server and the overall values of the time and traffic of the session itself.

2. *Web server log files:* an e-CRF shares several aspects of a standard Internet site, including the fact that the web server creates highly detailed log files, from which it is possible to identify each interaction between the user and the application, regardless of whether the user makes modifications. This means that the log file of the web server also registers simple page viewings, the time lapse between viewings, as well as numerous other interesting parameters. (Figure 3).
3. *Transaction log of the database management program (Oracle, MSSql server, etc.):* every interaction between the user and the database is

```
2005-03-16 16:08:42 148.177.129.212 -
GET /img/mediolanum.gif 200 5608 Mozilla/4.0+
(compatible;+MSIE+6.0;+Windows+NT+5.0)
http://www.eti-endovascular.org/
```

Figure 3.
Example (fragment) of the webserver log-file (generated by MS IIS 6.0). You can see the date and exact time of the request; the request type (in this case a GET request); the requested page or object (in this case the image Mediolanum.gif, from the folder img/); the response code (200 = success), page or object size (5608b), the type of browser and the relevant page.

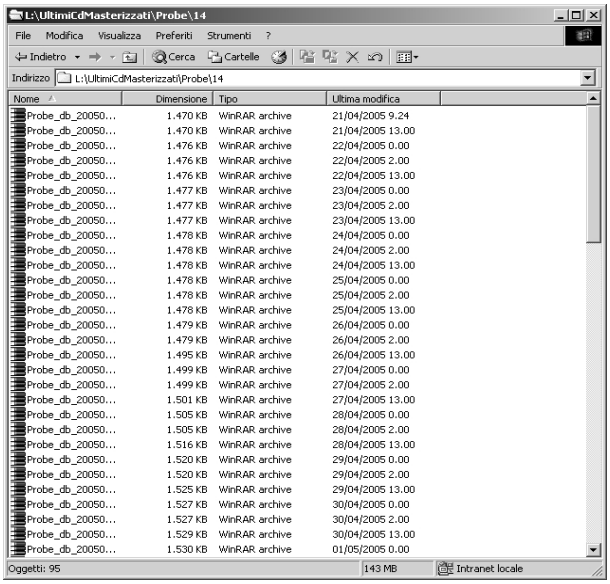


Figure 4
10 days of full backups recorded on a non rewritable CD. Every backup sequence is recorded onto two identical CD supports and represent an optimal 'long term storage' of this part of the Study data.

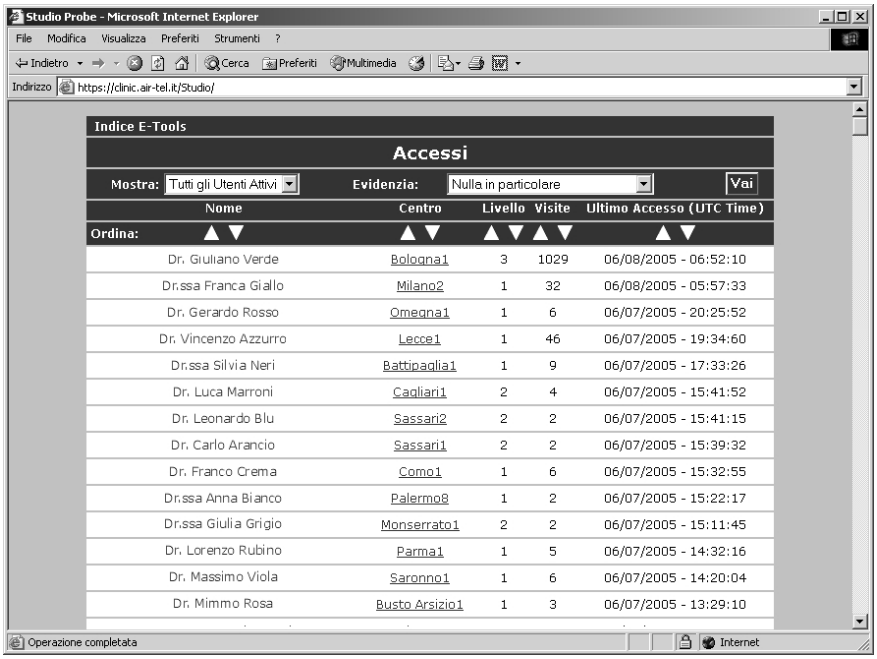


Figure 5.

traced and recorded. Even though the main aim of the transaction log is to allow for an accurate reconstruction of the database in the event of breakdown (the failure or malfunctioning of one or more components), there is no doubt that this constitutes a complete source of information in relation to interactions that have followed one another over time.

4. *Incremental copies of the database:* at fixed intervals (approximately 2-3 hours), the entire database is duplicated (backup copy) and "frozen" on two (for security reasons) different Write-Once-Read-Many supports (typically non re-writeable CDs). At a later date, this allows one to re-trace the entire history of the modifications made to the database itself as well as to retrieve any values present prior to correction (Fig. 4).
5. *Registration of accesses to the e-CRF:* all user access is also registered in the database of the application itself (the e-CRF). This makes it much easier to know who connected and when, thus drastically reducing the need to analyse the log files described in the previous points (Fig. 5).

Quality of traceability

Having ascertained that the use of a correctly designed e-CRF guarantees complete, comprehensive traceability of the interactions between users and the clinical database, one consideration of paramount practical importance remains: the ease of retrieving data history.

As a matter of fact, saying that something is possible is different from saying that it is easy to implement (sending a man to the moon was obviously possible but certainly not easy).

The traceability discussed above can be seen as a "guarantee": in fact, despite preserving all the information, obtaining a definite response requires analyses that are often laborious and time-consuming. These analyses are a sort of "last resort" and are only used when absolutely necessary for control and confirmation purposes.

Conversely, the traceability required in a Clinical Research tool has additional characteristics and can be referred to as "evident traceability": the Sponsors (and the Authorities interested in the data resulting from the Study) insist that not only is each modification registered, but also done instantly and easily accessible (e.g. by clicking on a link located at the side of any data that has been changed).

In e-CRFs generated in strict compliance with GCP (we call it Strict GCP), this is achieved by checking, each time data is transmitted, the value of new data against any data already, present in the database. In the event of differences being found, the old value is registered as a historic value (together with the name of the individual who entered it and time of entry) and the new value takes its place in the e-CRF (Fig. 4).

According to the old saying "There is nothing new under the sun", in actual fact, several of the requirements related to a clinical study are similar to those

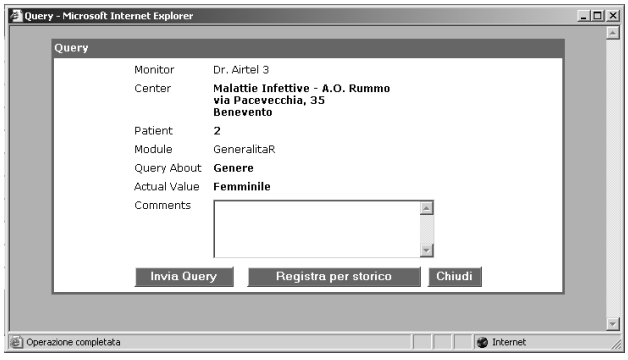


Figure 6.

of a classic accounting program; while it is impossible to delete even a single transaction, it is, of course, possible to perform a write-off which maintains the historic information of the transaction itself.

Therefore, in e-CRFs produced in Strict GCP, there is a table dedicated to "historic" values, while in database tables, only data corrected and considered valid (clean data) are stored.

In each form (and for each patient) all corrections are highlighted in a specific area and are instantly available to the study Monitors.

In view of the fact that the traceability available in log files is always guaranteed, not all studies require explicit traceability. For this reason, the e-CRFs most widely used in observational studies or in registers for compassionate use, allow the Study Monitor to decide whether to trace, or not to trace, a change, meaning that she/he can decide field by field and data by data.

ELECTRONIC SIGNATURE

Few subjects have been as widely debated and as little understood as those regarding digital signatures (or electronic signatures), especially when applied to on-line Clinical Research. When consulting literature, one discovers that everyone has something to say, offering rules and instructions which, at times, are so complex and obscure as to make the e-CRF that should 'adopt' them almost unusable.

Here is an excellent example of the need to, once again, return to objectives.

A method used by the FDA, which redefines the "electronic signature" as:

"a sequence of symbols or series of symbols, carried out, adopted or authorized by an individual as an equivalent of the obligations generally associated with a written signature (13)"

The objectives that appear to be most strongly linked to the term "electronic signature" are security objectives and can be divided into three broad categories (6).

- **Identification**: one wishes to be certain of the identification of the users in order to assign the responsibility of information accurately.
- **Integrity**: it is important to guarantee that the information transmitted coincides exactly with the information received (i.e. that no manipulations or changes of any type whatsoever have been made during transit).
- **Confidentiality**: only authorized users can read the information which must be totally inaccessible to all others.

It goes without saying that these objectives are absolutely essential for all types of on-line transactions (as always, the example of banking software applies) and the solutions actually adopted are often simpler and more effective than one might imagine.

On-line identification. Options and limitations

In Garfinkel's (6) by now classic publication on Web Security, he points out that the methods of identification, either currently available, or likely to be available in the future can be divided into 4 groups:

- Methods using something that the subject knows: e.g. a password, a PIN or a code (Fig. 7).
- Methods using something that the subject owns: e.g. a cash card (ATM) or a personal digital certificate (Fig. 8).
- Methods using something that the subject 'is': the world of biometry, by means of the detection of finger prints, the iris, etc. (Fig. 9)
- Methods using the place where the subject is located: workstations inside a network, from which critical operations can be performed (Fig. 10).

When the user enters his/her own UserID and Password, the system compares them with those contained in a Central Database related to authorized users.

If the UserID is recognized and the Password matches, then the user is given access to the application and attributed with the powers specified in the same Database.

Figure 7.
Something the user knows.

Figure 8.
Something the user owns.

Figure 9.
Something the user is

Figure 10.
Where the user is

According to the system, anyone using a valid UserID and Password figures as the legitimate owner and all of his/her actions are traced as if they had been performed by the owner himself. For this reason, it is absolutely crucial that particular attention and confidentiality be paid to one's own personal identification codes.

When the user introduces his/here card (e.g. ATM), although this acts as a UserID, it is also capable of transmitting and receiving a large number of additional information. Once the card has been inserted, it is necessary to digit a code, which acts as a password and allows the system to decide whether to enable each single operation.

Biometric terminals are capable of detecting and comparing, in real time, several biometric data with those contained in a reference database.

These systems are currently used in closed environments (e.g. banks and offices). They are fairly powerful as regards identification processes, even though, as shown in some science fiction films (Minority Report) biometric detection systems can be outwitted: one need only remove and use the biometric "piece" read by the system (fingertips, the iris, the entire eyeball, etc.).

On our safe net, the majority of administrative operations can only be performed from some unequivocally identified internal stations (by means of a unique IP address).

If used together with a UserID and Password method, this system proves to be the most secure of those mentioned until now. In order to get past them not only would a hacker have to steal the identification codes but also manage to physically sit at one of the enabled workstations.

It goes without saying that the methods mentioned above do not have either the same power nor similar characteristics as regards adaptability to this or that task; so, when passing from theory to practice, we realize that:

Biometry requires special machinery, not yet available to the average user, and, in fact, often only available at company head offices (nowadays, almost all banks have fingerprint detectors); furthermore, the same applies to card detectors (take, for example, personnel management in a large company, based on the use of these devices which detect workplace assiduity).

Place control has always been the main watchdog in the management of complex and critical networks; generally speaking, only some machines (physically or functionally internal to the network) are enabled to access systems and interact with administrative powers. However, even this method is not applicable to the vast majority of e-CRF users.

PGP SIGNATURE

The public keys used by the PGP (Pretty Good Privacy) method are short text files. Once installed on your PC, they are read and interpreted thus becoming more explicit and easier to use. Below on the left, is shown the Philip Zimmermann's public key, in text form; on the right, you will notice the appearance of a key (in this case belonging to Simson Garfinkel) following "interpretation" and storage by a PGP management program in a Windows environment (from 6. modif).
It is, however, necessary to point out that being capable of generating a pair of keys (public/private) is not enough to identify oneself. In other words, the person generating the keys can still pretend to be a different person.
This explains why, on the Internet, it is possible to retrieve public keys referring to Batman or Superman. The step from simply generating keys to actually guaranteeing the identity of the owner requires the involvement of Certification Authorities.

```
- - - Philip Zimmermann

5 June 2001
Burlingame, California
http://www.philzimmermann.com

——-BEGIN PGP SIGNATURE——-
Version: PGP 7.0.3
iQA/AwUBOx0vPsdGNjmy13leEQJ4qQCgoLgA
AZJfe2ORgoplAv9s39/JtP8AoOhu
nnhGSufR7jjAGj4tM8djwrcm
=MeBD
——-END PGP SIGNATURE——-
```

So, what's left? Methods based on codes or passwords!
In fact, they are the ones that are still most widely used today. To tell the truth, the distribution of personal digital certificates is a strategy for which all the necessary devices already exist (any PC can manage them). However, there is one drawback which prevented the first banks (which had initially adopted them) from using them: i.e. that a personal digital certificate must be installed on every machine which one wishes to use for connection purposes (the principle of being able to use any workstation connected to the Internet for operating purposes does not hold true).
Furthermore, both correct storage of the certificate by the user and the above mentioned installation are operations that are more complex than they might seem and are, however, a far cry from "sit down, connect and work" which is the underlying basis for the success of distributed services. You will be pleased to know that the method that uses UserID and Password can prove to be an extremely powerful and secure method, especially if combined with a valid system that guarantees the integrity and privacy of information, starting from the password itself (Fig. 11).

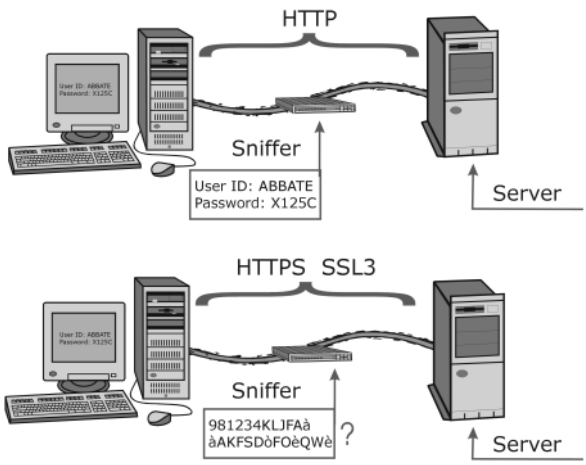


Figure 11.
Secure, protected access using UserID and Password together with encrypting: in case A, a "sniffer" (a device designed to "capture and read" everything that transits on a given network node), could identify the UserID and Password entered by the user, giving it to a hacker for subsequent use. The transit of a UserID and Password, even only once but in clear text, is sufficient to suggest a change of the Password itself, considering it "burnt". Conversely, in case B, first and foremost, the encrypting system protects the information entered for access and, consequently, the UserID and Password themselves.

Even the Internet has no protection against traitors

Since time immemorial, confidential information has been leaked or fraudulently spread for profit by unscrupulons authorized users.

When one uncovers identification by means of a UserID and password (or digital certificates or, once again, cards) one realizes that there is no protection against those individuals who give the access codes in their possession to others in order to cheat the system.

If I give someone my cash card (charge card) and access codes, this person could easily cheat the bank's IT system by pretending to be me; if he withdraws money, that money will subsequently be charged to my current account.

Similarly, if I give my UserID and my Password to someone for a particular application (e.g. e-mail), this person could trick the system and read (or cancel or send) all my e-mails.

In other words, UserID and Password systems are based on the collaboration of the user and holder of the codes who, to all intents and purposes, becomes a part of the security system.

From this point of view, the world of electronics does not differ from the real world: the security of documents is guaranteed, first and foremost, by those who have access and who commit to not divulging them illicitly (7,8).

Integrity and security guaranteed by cryptography

To encrypt means to transform the way in which a piece of information is "written", making it indecipherable. Only those who know the key (i.e. the rule used to transform the information) are capable of deciphering it and understanding it. Nowadays, the situation is a little more complex (for example, the key used to encrypt a piece of information is not the same as that used to decipher it) but the principle remains the same.

In any case, the real strength of the system lies in the difficulty of identifying the key and, consequently, of being able to "read" the information without having the necessary authorization: if one uses a banal key (e.g. the next letter in alphabetical order, so that ANDREA becomes BMESFB) we can expect a good cryptographer to identify it in a few seconds. Computers have increased the possibility of creating and managing complex keys. Unfortunately, they have also increased the possibility of breaking these keys (e.g. by trying out all the possible combinations in just a few seconds or minutes.)

The good new is that, lengthening the encrypting key exponentially increases the number of attempts needed to break it.

This is why the increased power of the processor plays in favour of security (9,10) (see insert on page 24).

Figure 12
The cryptographic capabilities of a browser are shown in the window "information about…"

As regards encrypted Internet transactions, the standard system currently in use is based on the HTTPS (11) protocol, a variant of that used by the web (HTTP), with a Cypher strength of 128 bit. This indication is provided by every browser (such as Internet Explorer or Netscape) generally in the help section, "Information about…" (Fig. 12).

As previously mentioned, this system is ideal for bank transactions and is thus also considered ideal for the transmission of clinically confidential information, in the sense that it guarantees both integrity and confidentiality.

While confidentiality is based on the fact that, during the transit of information on the Internet, this is encrypted and remains encrypted until it reaches the central Server or the user's PC (meaning that the possible theft of information in transit would be unusable by the "thief"), the guarantee of integrity derives from a second characteristic of the encrypting method used.

A "thief", though incapable of understanding the information in transit on the Internet, could, however, "steal" it and change it by introducing pieces at random or removing others. He could then send it back to the original address (although not an easy task, this possibility cannot be excluded) simply to reduce the operating efficiency of the system.

HOW MUCH TIME IS NEEDED TO "BREAK" A SYSTEM?

Thanks to the advances in processing power, it has become possible to use longer and more complex encrypting keys. This means that it is now virtually impossible to violate these keys using "brute force". This table shows that the increase of processing power is to the advantage of security due to the exponential nature of possible combinations resulting from the increase in the size of the key used (6).

Key length	Speed (keys/sec)	Instrument used	Time required
40 bits	10	10 year old desktop computer	3,484 years
40 bits	1,000	Typical desktop computer today	35 years
40 bits	1 million	Small network of desktops	13 days
40 bits	1 billion	Medium-sized corporate network	18 minutes
56 bits	1 million	Desktop computer a few years from now	2,283 years
56 bits	1 billion	Medium-sized corporate network	2,3 years
56 bits	100 billion	DES-cracking machine	8 days
64 bits	1 billion	Medium-sized corporate network	585 years
80 bits	1 million	Small network of desktops	38 billion years
80 bits	1 billion	Medium-sized corporate network	38 billion years
128 bits	1 billion	Medium-sized corporate network	10^{22} years
128 bits	1 billion billion (1×10^{18})	Large scale Internet project in the year 2005	10,783 billion years
128 bits	1×10^{23}	Special-purpose quantum computer, year 2015?	108 million years
192 bits	1 billion	Medium-sized corporate network	2×10^{41} years
192 bits	1 billion billion	Large scale Internet project in the year 2005	2×10^{32} years
192 bits	1×10^{23}	Special-purpose quantum computer, year 2015?	2×10^{27} years
256 bits	1×10^{23}	Special-purpose quantum computer, year 2015?	3.7×10^{46} years
256 bits	1×10^{32}	Special-purpose quantum computer, year 2015?	3.7×10^{37} years

a. Computing speeds assume that a typical desktop computer in the year 2001 was capable of performing approximately 500 million instructions per second. This is roughly the speed of a 500 Mhz, Pentium III computer.
b. In 1997 a 40 bit key was cracked in only 3.5 hours.
c. In 2000, cracking a 56 bit key took almost 4 days.

Server Digital Certificate

In order to ensure secure transactions (using the HTTPS protocol), a server must have a "Digital Certificate" (Fig. 13). Not only does the digital certificate enable secure information exchange but it also enables the user to ascertain that the Internet connection is taking place with the desired server and not with an "impostor" (a sort of fake cash card dispenser, used to steal information rather than cards).

It is, in fact, the principle of inalterability that enables anyone to check whether a Certificate is authentic. In addition to all the information related to the Server and the company that manages it, it also contains the digital signature of a Certification Authority recognized worldwide (see, for example, Thawte – 12) and, most important of all, recognized also by your computer.

In fact, when correctly configured, a browser already contains a long list of valid Authorities. Furthermore, explicit warnings are also given in regard to the addition of new Authorities to this list.

A fraudulent system could induce you to add a false Certification Authority among those accepted as valid by your browser (or, if you prefer, by your PC).

As previously mentioned, the encrypted transmission of information protects first and foremost, the identity of the authorized user: the first information to be "stolen" would, in fact, be his UserID and password and, from that moment onwards, the user would have an IT clone capable of impersonating him at will.

Also for this reason, it is advisable to use at least two different passwords for daily work purposes: a medium level Password, that can be entered even when the connection is not using a secure protocol, and a high level password, to be used only for critical applications (requiring absolute privacy and security).

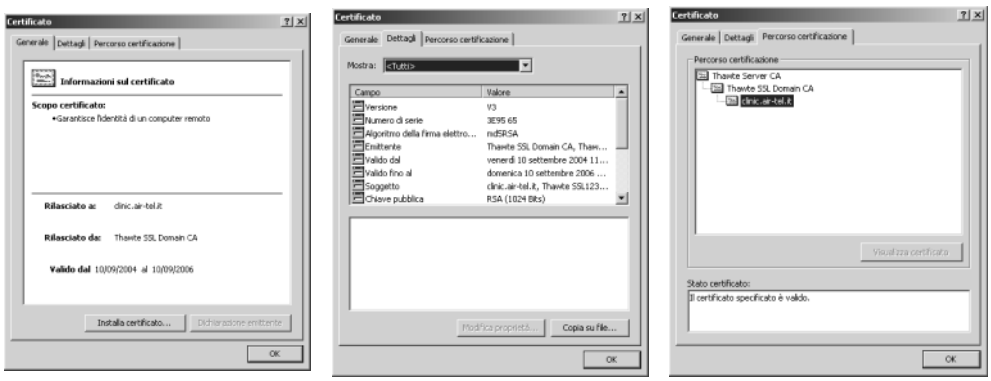


Figure 13.
Digital certification of one of our servers, as displayed in an MS Windows system. Please note that the Certification Authority (Thwate.inc) guarantees server identity.

And what if a critical application does not use secure transmission?
I will answer this question with another question: would you leave your valuables in a safe without a key? A Password is a code that protects our identity and should be treated with care. If a critical application does not use secure transmission then no application should be used (call, send a fax or use some other method).

TWO TYPES OF DATA TRANSMISSION

While on the subject of Clinical Studies, at least two types of data transmission exist:

- The transmission of data regarding individual patients from the Experimental Centre to the collection point (generally the CRO or, when using an e-CRF, the Central Database).
- The transmission of all data regarding the Clinical Study to the various Authorities involved (e.g. the Ministry of Health or the FDA).

This co-existence creates some confusion when both transmissions are done via an electronic medium (a real possibility with e-CRFs) (Figures 14 and 15).

Everything becomes simpler if we consider the two cases separately, using the necessary security strategies for each one of them (attempting to apply the same strategies to both of them, might create an unnecessarily complex and unmanageable situation).

The first case: from the Centre to the Central Database

As previously mentioned, the transactions that take place on the Internet using electronic forms can be protected by the use of the HTTPS protocol (Figure 14).

When combined with an automatic tracking system, it is possible to guarantee that the Data transmitted :

1. have been entered by an authorized user and no-one else (unless the user himself has "given" his identity to others by providing them with his UserID and Password);
2. have not been subjected to any changes along the way and are registered as such in the database.

As far as these transactions are concerned, there would not appear to be any need for a digital signature in that identity, integrity and confidentiality are already guaranteed by the system.

Or, even better, as defined by the FDA itself (13), the system already uses a valid digital signature, in the sense that the owner of the UserID and Password is aware of his actions when using his UserID and Password.

The important point here is that the acceptance of responsibility – requested in explicit, written form – is often the only means of making the user understand the importance of not revealing access codes to others.

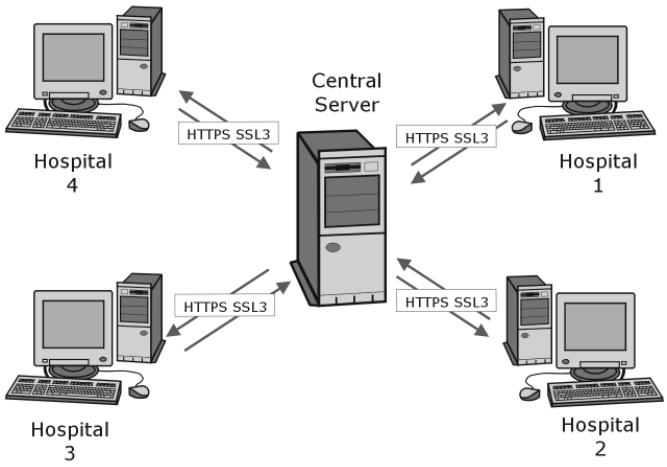


Figure 14.
Transmission of Clinical Data from Experimental Sites to the Central Database. Privacy and Security are granted by the HTTPs protocol (the same used for home-banking services).

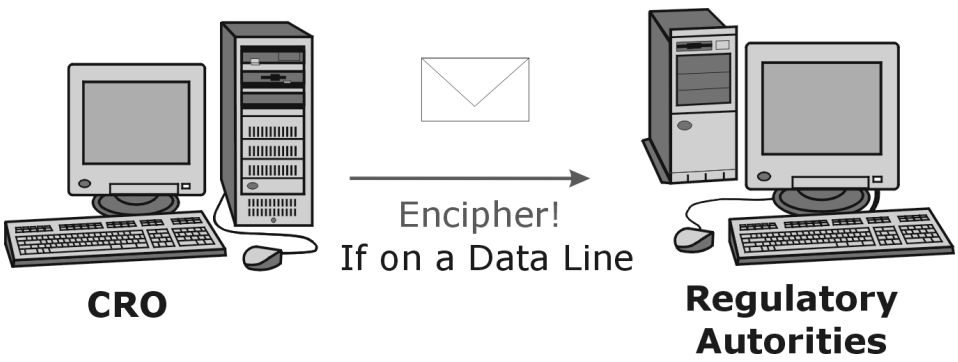


Figure 15
Final Study Data submission to Regulatory Authorities. All data should be actively encripted by the sender(s).

The second case: from the CRO to the Authorities

On the contrary, when one wishes to transmit documentation using an electronic medium (e.g. the entire dossier of a Clinical Study, with attachments, pictures, declarations and so forth), it is necessary:

- To digitize (type in or scan) and clearly order all information
- To send it in a encrypted electronic format (Figure 15).

Hence the reappearance (this time justified) of all the above mentioned problems related to security: how can the recipient be sure that the documents have not undergone any changes during transmission?

And what guarantee do both parties have that no one has "accessed" the information during transit?

As you might have guessed, the answer is always the same: encipher it!

But the method of implementation is different.

In this case, the encrypting operation must be performed by the person sending the information, using his own key (or a known key, protected by a Password).

In this way a "package" is created. A package that can only be opened by the person holding the key (or by someone who knows the Password) and which can, therefore, be sent over the net (e.g. as an attachment to an e-mail).

We are obviously talking about locking/sealing the entire document in a sort of container which will subsequently be encrypted to safeguard "transportation" to areas that are not secure.

Only the legitimate sender knows the key which, when given to the recipient, also acts as a guarantee of self-identification.

But, unfortunately this is not enough: one might wish to have the guarantee that each, individual document contained in the "container" is actually original.

How can one be sure that it has not been "manipulated" somewhere along the way? This makes things a little more complex: it is not, in fact, possible to repeat the same operation twice (i.e. transform each individual document into a sort of small sealed container) or, at least, it is not possible to do this without running up against the problem of who gives the keys to who.

Imagine that three experts A, B and C, complete their reports and then send them to Mr. X at Y company for subsequent transmission to the Authorities. Each of these experts has obviously taken steps to encrypt their work (thus making it illegible and unchangeable) but now have a small problem: to whom should they give the key?

To Mr. X or to the end receiver (the Authorities)? And what happens if the documentation needs to be sent to another Authority? Should the experts also entrust their keys to the new Authority?

As things presently stand, there are different answers to this question, even though the current trend would seem to encourage more advanced users to use a Personal Identification Certificate issued by a recognized Certification Authority (i.e. the same Authorities that issue servers with the necessary Certificates, enabling them to use the HTTPS protocol).

The hugeadvantage of this type of Certificate is that there is no longer any need to transmit keys or Passwords in a secret form. In fact, if I happen to hold a certain certificate I will have not only one, but two keys. One, called a private key, known only to me, and one, called a public key, which can be revealed at will. Moreover, public keys are, in fact, available to anyone in a list stored by the same Certification Authority which issued it.

When I, (the expert, in our example) complete a document, I will have to encrypt it using my private key. In order to read it, the person in question will have to use my public key. If he manages to read it using my public key, he has the guarantee of two important factors:

- The document was actually drafted by me (only by codifying it with my private key will it be legible with my public key)
- The document has not been subjected to any alterations (in the event of it having been altered by someone else, he or she would not have been able to recode it using my private key).

It's easy to imagine a near future in which each of us will possess our own personal Certificate. However, as things presently stand in Europe, this is not the actual situation.

And then some of you might have noted that, although the above mentioned system guarantees the origin and authenticity of a message, it does not make it illegible to the world (if you think about it, a public key is, in fact, public).

What now? A simpler system (which uses the same encrypting technologies but which does not require users to own a Personal Certificate) is based on the possibility of encrypting a document using two different Passwords: one to read the document and the other to change it (Figure 16).

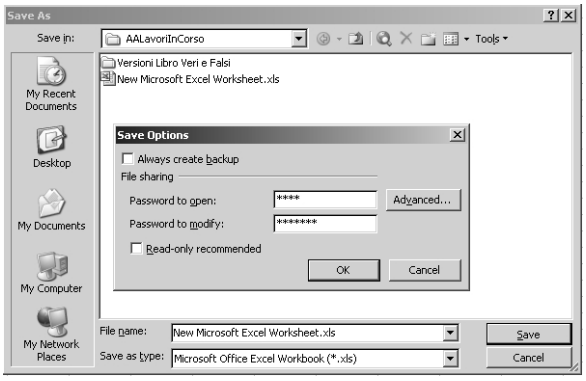


Figure 16.
An example of data encription based on an MS Access save option.
The sender can decide to protect his document with different passwords (one for reading permission, the other for editing permissions).

In this case, the experts mentioned in our previous example will have no concerns about giving their "read" Passwords to Mr. X. Furthermore, they will also be able to authorize him to give it to anyone interested in the document (i.e. various Authorities) while keeping the "change" Password for themselves in order to guarantee the authenticity of the documents.
This system is currently very popular but, because it is based on Passwords generated by users, it is also much less secure than those based on Personal Certificates.

Non re-writable supports

On the contrary, a widely used and accepted "bullet-proof" strategy exploits the physical properties of CD-Rs, i.e. the fact that it is only possible to write them once, and once only.
The contents can be read (with the usual restrictions linked to possible encrypting) or destroyed, but never changed. In this sense, CD-Rs come closest to the indelible ink system used in the past. It is impossible to tamper with them without leaving obvious traces.
This solution is ideal for long-term data storage (obviously in duplicate copy) and for tamper-free transmission.

CONCLUSIONS

The early years of this third millennium have shown us that it is possible:

- To collect clinical data safely, using an HTTPS protocol and suitably handling authorizations
- To transfer such data, at pre-set intervals, onto non-rewritable media.
- To guarantee, on conclusion of the collection period, the authenticity and originality of data (documenting all changes made to the data).

Other technologies (including removable media, such as CD-Rs), can be effectively used to integrate such data into the full study documentation and to ensure valid transmission to the Authorities.

The experience acquired over the last few years has helped us to understand that the real "problems" (intended as a means of identifying solutions) involved in the use of e-CRFs lie elsewhere.
Once identified, as with all problems, these too can be effectively solved: the main difficulty, nowadays, lies in correctly identifying them.
Here are those that we have encountered to-date.

AUDITING OF THE SYSTEM AND QUALITY RATING

Stringent auditing is an unalienable right of the Sponsor who, justifiably, wishes to ascertain that the instruments made available for the purposes of the Study (collecting clinical data in a secure, reliable way) are, in fact, suitable.

LACK OF STANDARDS

Unfortunately, it is not yet clear whose job this is or how the person involved should go about it. The skills generally involved in auditing do not necessarily include either skills or experience in the evaluation of distributed data collection systems, such as, for example, e-CRFs. So, they often tend to be based on parameters originally developed for other purposes: hence, one finds oneself having to fill in questionnaires designed to investigate aspects that should be taken for granted whilst neglecting other critical aspects to which no-one pays any attention.
During our initial start-up phase, we were invited to fill in a questionnaire that had obviously been devised for connectivity suppliers. It was, therefore, important to specify the number of modems available to users (none), the minimum band guaranteed for access of this type (no access, no band) and what connectivity back-up systems we were able to guarantee (none).
Only the questions on the band supplied for the applications (at that time, 2Mbit, about 60 times greater than the band used by enabled users) were slightly more realistic. But what struck us most strongly was the apparent desire to "safeguard" authorized users: from the point of view of the technology provider who believes in what he is doing, a user cannot simply be "safeguarded"…but must, literally, be "kowtowed to" and, at the first vague signs of a "congested" band, it is the provider's job to double it or increase it tenfold (if the provider is lucky enough to have this need).
Similarly, we were also asked about system autonomy in the event of power outages or faults on the line. The only possible fix to this problem are some Uninterruptible Power Supply Devices. They will 'hold the current' for the few minutes needed to complete a correct shut-down of the servers. In fact, in case of power outages or Telecom line failure, the line would be down in any case (it is centrally powered) and no local device would be able to reverse the situation: the only thing to do is to wait for Telecom (or the power supplier company) to fix the failure.
Unfortunately, if a parameter cannot be applied, it simply cannot be applied, and we had some problems explaining that the outstanding features of our

Total hours	35.072	100%
Total up-time	34.991	99,77%
Hours off-line	81	0,23%
Off-line Reason		
Power Shortage	18	0,052%
Line down	50	0,145%
Unplanned maintenance	13	0,037%
Planned ordinary maintenance does not require service interruption		

Table 2
Service interruptions registered on Air-Tel servers in the 2001-2004 period, and related reasons.

products lay elsewhere, not in this technical aspect.
Undoubtedly, table 2 helped give an idea of the quality of the service provided over the last few years, even though, as with financial investments, past prospects are no safeguard for the future.
In conclusion, there is still a considerable lack of homogeneity in the offer of instruments such as e-CRFs. At times, the term is referred to a database in MSAccess, enhanced by input masks and supplied to various Centres. Each Centre enters its data "locally" and, at fixed time intervals, the set of all access files is collected and centralized for analysis purposes (alignments often take place at night, either in an automatic or semi-automatic way).
At others, the term is used to identify a few, simple forms, created without particular concern for data security or data traceability, generated with no encrypting systems.
However, as time goes by, it is becoming clearer what should be meant and what one can expect of an e-CRF and, it is not difficult to imagine a time in which the minimum quality requisites, necessary to define an electronic and distributed data collection system as an e-CRF, will be clearly indicated.

AUDITING WHAT?

In the same way, it is not clear what must or should be reasonably subjected to an Audit: it goes without saying that the stability, over time, of a structure or system is an indispensable characteristic for Auditing to make sense. Why should one certify a PC if, after only ten minutes, this could be infected by one of the innumerable viruses present on the net or lose some of its properties following the installation of a new type of software by a fifteen-year-old son or daughter?
Initially, a typical Auditor will probably say "Certifying a PC also means verifying its adequacy in terms of protection against viruses. Furthermore, if a PC is used for Clinical Research, not only will access be denied to everyone except the Monitors, but it will also be impossible to install anything else on it, let alone useless games".
This means: checking a machine and applying severe restrictions in regard to

everyday use.
Therefore, without even wanting to or fully understanding what is happening, one of the revolutionary (and winning) characteristics of e-CRFS are eliminated at the start: i.e. the possibility to connect up and work from any part of the world, using any type of PC that is capable of navigating on the Internet.
However, should we wish to preserve this characteristic, we will then ask ourselves: how can an Auditor check all the PCs that might be used for data input? And, above all, how can he prevent others from using them and guarantee that their legitimate owners keep their anti-virus or anti-intrusion software updated? Put in this way, it is simply impossible.
According to the best traditions, when something is impossible one ignores it, only subjecting what can actually be verified to auditing (i.e. centralized data collection systems).
In actual fact, although this strategy is the only feasible one, it must however be backed up by some rules. These rules (or standard procedures) should be aimed at eliminating any damage stemming from a lack of control of the uncontrollable (the infinite number of PCs present on the net which are potentially suitable for data entry into an e-CRF Central Database).

Keeping responsibilities and tasks separate

The first main distinction should separate what it makes sense to audit from what it is impossible or useless to audit.

Immediately afterwards, it is important to guarantee that all critical tasks are performed or verified by "auditable" systems. On the contrary, all that one can ask of "non-auditable" systems is that they "do their best", ensuring that every piece of data originating from them (i.e. all data) will be verified and validated by Central (and auditable) systems.
The following are included in the auditable systems:

a. central servers (Database and Web application),
b. the local network that hosts these servers and the anti-intrusion systems implemented to protect them (certifications, firewalls),
c. backup servers (ideally housed in a different building with respect to the former, as an anti-disaster strategy) and the lines connecting them to the central servers.
d. Last but not least, the Standard Operating Procedures adopted by the technology provider to guarantee the integrity and continuity of the clinical data collection and storage.

On the contrary, as previously mentioned, PCs that can be used by Investigators for data entry cannot be subjected to Auditing. The Auditing of such PCs is impossible (because neither their number nor their location are detectable) and totally useless (the PCs could become infected or "break down" just a few seconds after Auditing has been completed).

Green codes and red codes

In keeping with the above, it is advisable to assign either an imaginary green code to each machine participating in an e-CRF (machine subjected to Auditing and thus, reliable) or a red code (PC not subjected to Auditing and, therefore, potentially unreliable) (Fig. 17).

Having done this, it is then only necessary to structure the e-CRF in order to ensure that:

- All processing performed by machines bearing green codes will be considered valid.
- All processing performed by machines bearing a red code will NOT be considered valid and will be re-checked by machines bearing a green code. Only after the successful completion of this last check, will data originating from machines with a red code be entered into the central database; otherwise, the data will be rejected and the integrity of the database preserved.

In order to implement the above, it is necessary to clarify, in greater detail, what is meant by "processing performed by".

As some people may know, the world wide web browsing (i.e., those using HTTP or HTTPS) is based on the fact that a client (usually a PC) asks a specific server for some information (usually a page) and, in response to the request, the server sends back the information to the client (or a message saying that the requested page was not found).

The information sent back to the client can be divided into:

- Real information (data, text, images, etc.)
- Codes needing to be executed (e.g. page formatting, alerts, etc)

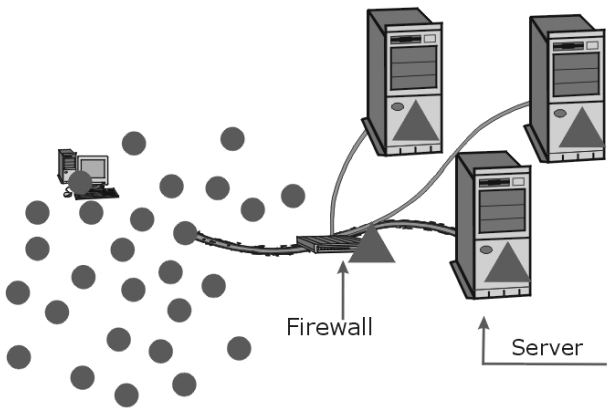


Figure 17.

It must be stressed that html page codes transmission to the users' PCs is unavoidable (i.e. instructions explaining page layout, text positioning and how to create fields that accept data).
Although it is impossible to do without this "scarcely reliable" minimum processing, the application should be designed in such a way, that, in the event of a PC not functioning correctly, the user will be the (only) one to suffer. He will not see a correct representation of the page (or of an e-CRF, or of any other web site) and, in more serious cases, he will not even be able to connect to an application, nor enter, nor send data to the central database.
A situation of this type entails "repairing" the PC in order to restore it to normal working order (not only for e-CRF use but also in order to use the Internet in the broad sense of the term).

UNRELIABLE EXECUTIONS

Nevertheless, more complex instructions using programming languages studied for this specific purpose (e.g. JScript), are often sent to the user's PCs via the Internet.
The interesting advantage of this strategy is that, since the instructions are executed directly on the user's PC, they are capable of intervening before data is transferred to the central servers (Figure 18).
Possible malfunctioning at this level could "cancel" the request for confirmation even if the code sent by the server contains the right instructions.
Unfortunately, the most that the programmer can do is send well written lines of code (instructions) to user PCs; there is no way that he can guarantee that his instructions (even if they are correct) will be correctly interpreted and exe-

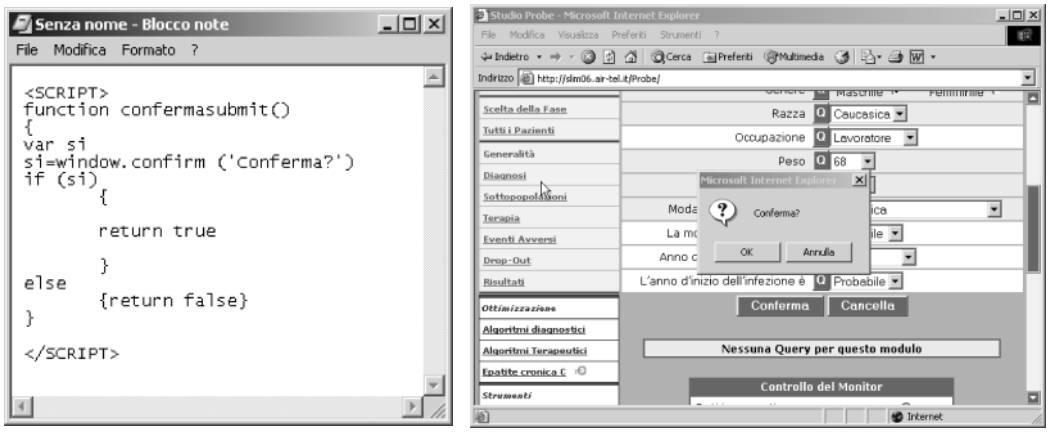


Figure 18.

cuted by the PC of the users, due to the uncertain reliability (red code) of the latter.
On the other hand, experience teaches us that the PCs scattered around the area will miscalculate the interpretation and execution of the codes in a percentage of between 5-8% of PCs.
This will result in behaviours (or output) different from that desired by the programmer.
This explains why, when performing a Clinical Study using an e-CRF, it will be impossible to assign the execution of codes to the PCs of users or assume that the result is always reliable. This does not mean that a tool like JScript should be totally avoided: on the contrary, it means that the instructions in JScript, cannot, under any circumstances whatsoever, be given tasks that have anything to do with the validity or integrity of the data transmitted. JScript will only be entrusted with instructions which, if correctly carried out, could help guide the user, without interfering with the quality of the data sent. In this way, if the remote PC "makes a mistake" or does not correctly execute these codes, the only person to suffer will be the user sitting at that PC: he will not have either those instructions or those concessions that the programmer has entrusted to JScript, while the data sent will, in any case, also be checked and eventually validated by a central system (green code) before being registered in the database.
A few examples will help clarify the above.

Help functions (permitted)

A typical function assigned to JScript is control of the validity of dates PRIOR to these being sent to the server: seeing that this can be done, it seems pointless to accept any combination of day/month/year chosen by the user by means of suitable drop-down menus. As is common knowledge, February only has 28 days in normal years, 29 days in leap years but never 30 or 31 (Fig. 19)
And yet, there are instructions in JScript capable of checking the chosen date.
In the event of the date not being valid, the user is warned even before he attempts to send it to the central servers.
But these instructions are read and executed by the user's PC and, should the PC itself be "out of phase", JScript could give inaccurate results or none at all.
Therefore, to safeguard the integrity of the database, the dates in input must be checked once again. In the event of their not passing the validity checks executed by the central servers (green code), they will simply be discarded. The user will be aware of this because the date entered by him will once again be requested by the e-CRF.
In other cases, when malfunctioning is particularly severe, the user might not even be able to complete a form. After a phone call to the Help Desk, it is often possible to formulate a correct diagnosis and "cure" the PC, thereby restoring its operating function.

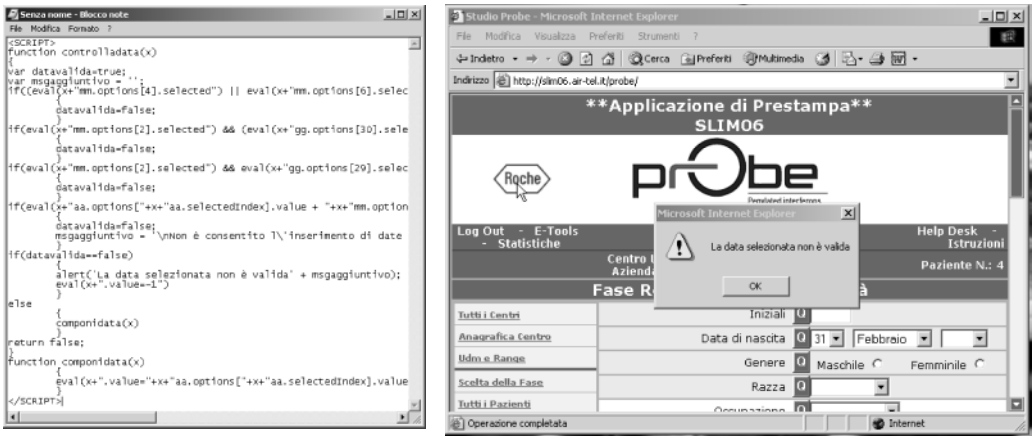


Figure 19.

Control functions (forbidden)

In most studies there are specific restrictions (set by the Protocol) regarding, for example, the age of patients eligible for participation or other relevant aspects.

During our first experience, one control that used JScript instructions was related to the birth date of patients scheduled for enrolment. A series of JScript instructions instantly matched the birth date entered with the date of enrolment, and verified that the age was compatible with the study protocol. If the patient's age proved to be compatible with the enrolment protocol, the date was accepted for transmission to the central database; if the age was not compatible, the user was alerted with a warning message with further data entry for that patient being denied and exclusion of the patient from the study.

But one day, when analysing the data collected, we realized that a patient, whose birth date would have put him at age 9 (a minimum age of 18 is required by the protocol) had been included.

Thanks to data traceability, it was possible to easily trace not only the Investigator who had "by-passed" the control but also the machine used (thanks to the IP address, which belonged to a batch used by his Hospital Centre). After clarifying the situation with the Investigator (it was a question of an input error: the patient was in fact 29 and therefore eligible for enrolment) we set out to discover how the JScript control could have failed.

We noted that the PC used had a curious malfunctioning: it executed the JScript instructions in an unreliable way, missing some and executing others incorrectly.

We were thus able to reproduce the error (which had not been picked up during any of the previous test phases since we had always used correctly configured, operative PCs for the tests).
The conclusion reached was the one mentioned at the beginning of this chapter: on account of not being able to validate all the PCs that could be used to complete an e-CRF, all validity controls (such as that regarding the age of the patients in relation to inclusion/exclusion criteria) were re-routed to the central servers, under the assumption that not all user PCs used were necessarily 100% reliable.

USER CONNECTIVITY (NARROW-BAND)

Especially in the case of new technologies, the USA are considered the source of the most advanced products and projects, already "tried out" by numerous individuals (the use of the Internet in the USA is much more widespread than it is in Europe).
What a surprise it was to learn that, on numerous and different occasions, the e-CRFs developed by USA-based companies and supplied to European branches (and, in particular, Italian branches) did not pass muster and were, thus, not used with much enthusiasm.
The reason soon made itself apparent: in the USA, broad band connections are widely used and, as a consequence, USA programmers worked under the assumption that all their potential users were equipped with this type of connection.
Since broadband means high speed data transmission, this induced the above mentioned programmers and designers to "enrich" their e-CRFs with images, sounds and even film clips which, though appealing to users, nevertheless proved to be extremely "cumbersome" (i.e. requiring enormous files).
The problem in Italy (and in other European Countries of course) is that the use

	Analogic 19.2 bit	**ISDN 64 bit**	**ISDN 128 bit**	**ADSL 512 bit**	**ADSL 2000 bit**
1 image enriched page	150 sec.*	50 sec.*	25 sec.*	6 sec.*	1 sec.
1 text-only page	15 sec.	5 sec.	2.5 sec.	<1 sec.	<1 sec.
* waiting times not suitable for on-line use of an e-CRF					

Table 3
Loading times for pages of 35 or 350 Kb on Internet connection from analogic modems to ADSL

of narrow bands is still widespread. In particular, standard old phone cables, analog modems, capable of transmitting at ridiculously low speeds (19.2 kbit) are the only connection resource available in many locations.
The unequivocal result of the tests performed showed that, in Italy, when using an e-CRF designed for broadband use, the waiting times are much longer than just those few seconds usually accepted for the completion of on-line tasks.
Fortunately, Clinical Research does not require images or gadgets, but is based on pure, fairly light, data transmission (numbers or short texts): nothing more suitable for our still, somewhat, minimal phone lines. Furthermore, consider that the problem is made even worse by the use of a secure protocol (HTTPS), since the encrypting procedure always increases the size of files to be transmitted. Anyone wishing to provide a product for use in Europe, must pay careful attention to this aspect, reducing data transmission times to a minimum and foregoing (for, at least, a few more years) everything else.

AN INTERFACE FOR INEXPERT USERS

Another important aspect that could become a problem if not considered and solved is that e-CRF users are doctors, not IT experts.
We are not insinuating that these users are not up to the task but simply pointing out that their prime concern is healing the sick, not using a PC or a web application.
Since they are well aware of this, they, naturally, expect the interface to be as user-friendly as possible: this means that though familiar with obscure medical terms and procedures, they prefer not to get involved in the technicalities of information systems.

Avoiding nightmarish form fill-ins

Anyone who has used the Internet and filled in an on-line form is likely to have had the frustrating experience of filling in almost all the fields, pressing the send button only to see the appearance of brief message that says "Please note, not all the fields have been filled in" (some of you will have guessed that this is a JScript code…).
Following receipt of this message, the form is re-presented as such and, if the missing data are not available at that moment, it will be necessary to repeat the entire entry procedure at some other time. Forms of this type deserve to be nick-named "nightmare" forms and are a valid reason for e-CRF failure.
This is clearly explained by the fact that, when an Investigator sits down at his PC to enter data, he is often interrupted or called to attend more pressing (and often more relevant) matters.
Heaven forbid if the data entered up to that moment had not been stored by the system and requested for the second time: we would have lost an Investigator and, in all probability, an Experimental Centre (at least as regards the use of the e-CRF).

Guiding users in a Friendly way

PC users know full well that their requirements change rapidly as they gradually learn how to use a program. A whole series of warnings and suggestions, which the user initially found useful, subsequently become a real form of persecution from which users gladly rid themselves by clicking on the wording "do not show this message in the future".
The first e-CRFs were full of suggestions, controls and compulsory routes, at times, in excess of protocol requirements (the aim was to "guide" inexpert users). We had to move quickly in order to remove everything that did not prove to be strictly necessary, thereby restoring the e-CRFs with their primary function: i.e. collecting the data provided by the Investigator in a secure, reliable way.
Although we were tempted (and even requested by the client) to enhance the tool with additional logic (a sort of "artificial intelligence" or elementary "first aid" technique) we decided to resist, leaving the practitioner with full responsibility for the data supplied, including the right to make (and register) mistakes.
Despite getting rid of the superfluous, an e-CRF must, however, contain all those elements of logic that guarantee a consistent flow of information and the immediate highlighting of potential errors. It must also guide the user along the way in order to (at least) minimize mistakes.

Mandatory data?

When it is a question of guiding fill-in, the concept of "mandatory" data (extremely useful from a practical point of view) inevitably crops up.
In the broadest sense of the term, mandatory means

"the Investigator MUST enter this data".

But the coercive power of a PC is minimal if not associated with some form of "obstacle" or "punishment" ("sanctions make law").
And yet, in a well-structured e-CRF, "mandatory" means that it will not be possible to continue a sequence if that particular data is missing (for this reason, it is sometimes preferable to call this field 'blocking' data).
If, for example, the minimum identification data (or GCP codes) are not specified, it will not be possible to enter any clinical data for that patient, because no one would know to which patient they were actually referring; similarly, if date of birth has not yet been specified, no real experimental data will be accepted, because, according to Protocol specifications, the patient might not be eligible because of his age.
This tells us that a field can only be "mandatory" if entered within sequentially ordered forms. However, this does not always hold true: take, for example, forms related to Adverse Events, which must always be available in order to pick up possible warnings or in the worst case scenario, in the form "Deceased".
There is no way that the clinician can be forced to complete it, even when all fields are indicated as mandatory.

Or better, there is no IT/telematic method: naturally, intervention by the Monitor (who is notified, by the system, about the above mentioned incompleteness) remains.
In view of the fact that he has greater coercive powers, he can invite the Investigator to complete all the data or, at least, send them to him in some sort of valid form.
In actual fact, although the IT "mandatory concept" is a useful tool to ensure the coherent flow of sequential data, it does not guarantee completeness: even in this instance, the role of the Monitor continues to be of fundamental importance.

EXPECTATIONS

The confusion that still surrounds e-CRFs and their use is further increased by expectations. Some are just fine, others so unreasonable that they often generate misunderstandings and severe production problems.
Anyone who has tried "playing around" with a graphics program (e.g. in order to retouch images) will have noticed the innumerable keys available: lighter, darker; with more colour or less colour; smaller or larger and so forth.
A friend of mine (a well-known art director of an advertising agency) once looked at the screen and said: "there are loads of keys but the most important one is still missing"
"Which one?" I asked.
The one which says "make it even better"…he answered.

Unreasonable expectations

This joke shows that buttons can implicate widely differing actions or concepts: every graphic program understands what is meant by "lighter" but there is not one in the whole world that knows what "better" means.
Those who assume that an e-CRF is capable, come what may, of preventing the entry of wrong values can be equated to the graphic artist in the above mentioned example. After all, this was never expected of traditional e-Crfs on quadruplicate carbon paper, on which anyone could write anything that entered their head, including the shopping list or a friend's phone number.
There are undoubtably errors which, on account of their enormity, are recognizable by following simple algorithms (e.g. laboratory values that are incompatible with life, such as 5 red blood cells per cubic millimetre) and a good e-CRF can (and must) prevent their entry by mean of plausibility checks.
But many of the errors made, even though seemingly obvious to a clinician, are extremely difficult for an information system to evaluate (there is a good reason why everyone still prefers a human doctor to a highly powerful artificial medical intelligence system). I remember the comment made by a clinician who, after examining a page specifying a therapeutic programme for a drug (out of over 450) and randomly selecting a drug, a packaging and a dosage programme said: "But this dosage is absurd for this drug!". He was truly disap-

pointed that the e-CRF had actually accepted the sequence entered.
While I was about to explain that it was thanks to his medical knowledge plus years of specialist studies and clinical experience that he had been able to "pickup" on this absurdity, a colleague of his came to my rescue in the best possible way by saying: "In my opinion, there are cases in which this could, in fact, be plausible".
This obviously resulted in a heated discussion about the medical issue.
The observation previously made about the e-CRF (which, obviously, did not undergo any changes insofar as that particular point was concerned) became irrelevant.
But this example is an excellent one and takes us back to our real objectives.
The job of an e-CRF is to collect data in a secure, reliable way, possibly highlighting any errors or inconsistencies.
However, under no circumstances, should an e-CRF come into conflict with a doctor about issues of merit (such as, for example, the plausibility of a therapeutic programme).
More likely than not, the very fact that technology continues to amaze us with its "miracles", results in the conviction that it is capable of doing anything and more besides.
On the contrary, from this point of view, the experience acquired with e-CRFs is extremely reassuring. As yet, nothing is able to replace the clinical eye or the ensuing decision made by the clinician.

Realistic expectations

On the other side, what an IT system dedicated to data collection can do is to generate indicative algorithms based on data entered previously. This has proved to be extremely useful and highly appreciated in the use of new drugs in "complex" fields, as is sometimes the case with programmes for compassionate use.
It is, however, necessary to specify that these algorithms must exist and be complete prior to the development of a program: otherwise, their implementation is impossible.
The production of valid, accepted algorithms is the most difficult part, regardless of whether they are generated via the Internet or provided as a sequence of paper photocopies.
Other possible expectations are summarized in the aforementioned table which lists the advantages of e-CRFs over the paper method and, as can be seen, we are dealing with extremely important matters. (Table 1).

SECURITY AND MANAGEMENT OF PASSWORDS

Despite claiming that the distribution of individual certificates could prove to be too complicated and costly and that the method based on the use of UserIDs and Passwords, together with encrypted transmission, constitute a system that is more than acceptable for the transmission of confidential data, it is advisable to take a look at the real conditions related to use of this method by the Investigator, both in Italy and in Europe.

It will, immediately, become apparent that there is an enormous difference between the desires and expectations of the Sponsor (who is, at least theoretically, particularly careful about these aspects) and the actual user: the former desires an overdose of security technology while the latter still tends to write his UserID and Password on a post-it note, which he then sticks on the department computer (we should not be overly surprised…This action is the same as leaving a "paper" CRF lying around in the Doctors' Staff Room).

In view of the fact that security is a sort of chain whose overall strength lies in the weakest of its links, some steps need to be taken:

- Provide information about IT security in general as well as about choice criteria and management of the Password (see Appendix 1, further on in the book).
- Break up the "chain" into parallel fragments, so that the breakage of one of its links does not result in the break-down of the entire system but only a part of it, namely, the area presided over by the person who leaves his UserID and Password lying around: in this way, only the person (and his co-workers) who has been careless will suffer.
- Only provide access at management level to individuals (Sponsors, CROs) who have attended a short course on IT security and who have understood its basic principles.
- Leave important administrative actions to Data specialists, restricting access to a physically identified workstation (e.g. only PCs present on the local network of system administrators).

For more information about Password security, refer to appendix 1.

ASSISTANCE

Human assistance

An extremely important aspect related to the provision of assistance to investigators concerns the need to always allow for "human" interaction: we, therefore, maintain that the temptation to use "call answerers", so popular at large Help Desks (with thousands of users every day), should be avoided (namely, those who tell us which telephone button to press to do something, guiding us through long and complicated procedures and ending the conversation with a vague answer or an invitation to call back - !?!?!?).

Even though well programmed, they are undesirable because they are dealing with people who are already sick and tired of fighting with PCs and Automated

Systems, and who wish to converse with a human being, who not only understands but who is also possibly able to solve their problem.
At least, during these first years of e-CRFs use, the quality of assistance is possibly the most important aspect for the successful outcome of a telematically conducted Clinical Study.

Full-range Assistance

When it is understood that a doctor, approaching an e-CRF for the first time, finds himself facing a situation that is totally new to him whereby the unknown element is not represented so much by the e-CRF itself, as by the fact that he has to use a PC and navigation tools on the Internet, then one realizes that requests for assistance will be of a global nature, thus requiring answers which should be just as global.
By this term I mean that, if the user cannot see the page, it will be necessary to make him perform (and help him while he performs them) a series of tests to check the following, in order:

- Correct functioning of his PC
- The existence and working order of a connection
- The possible presence of filters or a proxy that, unbeknownst to the doctor, has been activated by the IT Department of the Centre
- Correct configuration of his browser
- Correct configuration of his e-mail box and of the software managing it

All of this, naturally, after having checked that the central system is "up-and-running" as it is 99.9% of the time.
As can be noted, it is often a question of distributed system elements that do not fall under the direct control of the technology provider.
Nevertheless, the successful outcome of a Study using an e-CRF is based on acceptance of the new tool by the user and the answer given ("Look, your problem has nothing to do with our e-CRF. It is a local problem that should be solved by you"), even though probably true, is the worst one he could receive.
This results in a lack of interest in the project and a request to receive the paper version which will be sent, as usual, to the CRO.
Conversely, courteous, efficient service that helps a user solve a problem that he might have (and probably has) encountered when using other Internet sites or mails, wins the Investigator over, giving him that small amount of self-confidence which, together with his clinical experience, make him a user who, though possibly not enthusiastic, is nevertheless collaborative and highly valuable.

A source of improvement

At times, however, requests for assistance reveal possible misunderstandings and provide input in terms of changing this or that point which is, evidently, not sufficiently clear to the user.

When dealing with specific aspects of the e-CRF (and not with the system as a whole), requests are a precious source of improvement of the application itself in terms of its interface; instructions that might seem obvious to the programmer are not so obvious to the user and must therefore be highlighted. At times, some of the information provided is interpreted in the wrong way and must, therefore, be reformulated.

Furthermore, requests for assistance, when tending to refer to the same point, might indicate a real weakness, resulting in the need to change the supply platform, with improvements and advantages that will also be extended to other users, projects and studies.

FREQUENTLY ASKED QUESTIONS (FAQ)

Here are some of the questions that are asked most frequently during investigator meetings involving a study that plans to use an e-CRF.

These questions, referring strictly to Italy, give a good indication of expectations, actual IT skills and local user know-how.

1. **What are the minimum requirements that my PC should have in order to use electronic case report forms?**

   The most complex (and processor time-consuming) part of an e-CRF is encrypting: this requires processing power and we, therefore, suggest using, at least, a Pentium II with 300 Mhz or an equivalent, with at least 64 mb of RAM (conversely, no specific spaces on the hard-disk are required). Your PC, correctly configured and up-and-running, must have a browser (e.g. Explorer or Netscape) for Internet exploring purposes. The browser must also implement a safe transmission capability with 128 bit strength (all systems currently installed do this).

2. **Do I need to wait for your technician to install the CD-ROM or can I install it myself?**

   One of the advantages of an e-CRF is that no installation whatsoever is necessary.

3. **Must I always use the same PC?**

   No. You can connect up to and work on any PC capable of navigating on the Internet. You can use the PC at the Centre to do part of the work and then view the data on your PC at home, or vice-versa. Even a smart phone can be used, although this is not a particularly practical option (for the meantime) due to the size of the screen and connection speed.

4. **Can I save the Password?**

   If by this you mean that you won't have to re-enter it to access the e-CRF, the answer is: it's better not to. This answer applies to those cases in which the PC is also used by other individuals (on your own PC, possibly protected in turn by a Password, you can do this without running too many risks) and this applies to any Password that allows access to confidential information. Saving a Password is very convenient but the result is that anyone sitting at the PC can access confidential information belonging to the person who has saved the Password.

5. **Can we use a common PWD? There are only 3 of us working on the e-CRF...**

   We strongly advise that you read the short print-out on security and Passwords carefully. The answer is: NO!

   If only one common Password were used, it would not be possible to attribute "unequivocal paternity" to the data entered (traceability would be weakened). Never give your Password to anyone else, as it could be used by someone pretending to be you.

6. **Whom should I ask and how much will it cost me to open a new account for a new operator at my Centre?**
   To avoid interfering with traceability and to increase security, you will not be asked to pay an additional amount for the Password (user) managed; this policy could actually prove to be a sort of false saving.
7. **Why can't I change data?**
   As with traditional e-CRFs on carbon paper, although data can be changed it is, however, necessary to keep track of these changes. This is why an e-CRF includes various options (depending on the type of platform required): the most common is notification by means of a request mail to the CRO.
8. **Will I be the only one that can see my patients' data or will other Centres also be able to see it?**
   Unless established otherwise by protocol (and, in large projects, by the Statute), each Centre has the right to see, at any moment whatsoever, all the data related to patients enrolled at that centre, including general situation, statistics and graphics (if any).
   Only the Sponsor, the CRO (and the technological partner) have the right to examine the entire clinical database. As regards "final" ownership of data, nothing changes with respect to classic collection systems: this is protected by the agreement made by the investigator and the Sponsor.
9. **Since an e-CRF produces a descriptive statistic why doesn't it also produce an inferential one?**
   Because the former is standard, simple and not related to the experimental context.
   Conversely, the inferential statistic (which indicates the probability of observing a certain result in a certain population "not included in the study") is fairly complex and requires methodological choices that depend on the experimental context and which only an experienced statistician will be able to identify.
10. **I don't see any benefits in terms of Safety: if something goes wrong will I still have to send a fax to the competent authorities?**
    Although this still holds true today, things are changing rapidly and, as far as the handling of electronic data is concerned, the Ministry of Health is much further ahead of the game than it seems to be.
11. **The law is not clear and I want to make sure that I am protected. Does this mean that I have to keep a paper copy of the clinical files?**
    This is certainly a possibility. The use of an e-CRF does, at least, prevent tons of clinical files in paper format travelling around the country in order to provide the central database with the few milligrams of information that they contain (i.e.: the weight of the ink).

BIBLIOGRAPHY

1. **Pellegrini Umberto**
Dove ci porta la microelettronica. *LibreriaClup, Milano, 2002*

2. **Cassidy John**
Dot.Con: How America Lost Its Mind and Money in the Internet Era. *Harper Collin Publisher, USA, 2002*

3. **Philip J. Kaplan**
F'd Companies: Spectacular Dot-Com Flameouts. *Simon e Shuster, NY, 2002*

4. **Kevin Kelley**
New Rules for the New Economy (10 radical strategies for a Connected World) *Viking Penguin, 1998*

5. **De Rosa Andrea**
La sfida delle e-CRF (Abbandonare la carta per la grande Rete). *AboutPharma, n. 23, Novembre 2004, pp 34-35.*

6. **Simson Garfinkel**
Web Security, Privacy & Commerce (Second Edition). *O 'Reilly, USA, 2002*

7. **Pfleeger Charles P.**
Security in Computing (Second Edition). *Englewwood Cliffs, NJ, Prentice Hall, 1996*

8. **Amoroso Edward**
Fundamentals of Computer Security Technology. *Englewwood Cliffs, NJ, Prentice Hall, 1994*

9. **Denning Doroty E.R.**
Cryptography and Data Security. *Addison-Wesley, 1983.*

10. **Kaufman Charles, Perlman Radia and Speciner Mike**
Network Security: Private Communications in a Public World. *Englewood Cliffs, MJ: Prentice Hall, 1995*

11. **OpenSSL Project**
http://www.openssl.org/ - http://www.thawte.com/
A world wide known Certification Authority

12. **Secure your web server**
http://www.thawte.com/ssl-encryption/digital-certificates.html

13. **FDA, Definizioni e Guidelines**
http://www.fda.gov/cber/gdlns/esubapp.htm

PASSWORD MANAGEMENT FOR END-USERS

What are UserIDs and Passwords?

For years, it was possible to use computers without having to memorize codes or Passwords. Nowdays, you cannot even take one step or visit a site and…for heaven's sake, they have just saddled us with a new ID and another Password to remember.

It is pointless to say that this change has a lot to do with the wide use of the Internet.

In fact, as long as everyone uses their own computer without allowing access to anyone else, the use of UserIDs and Passwords is redundant. But since computers began navigating on the Internet, it has become important to notify the system of the identity of the individual sitting at the console: in this way, the system itself is able to remember our preferences and, above all, permit or deny access to this or that resource.

It is, in fact, for this very reason that Microsoft Windows (from NT to XP and Vista) asks anyone logging into the computer for a UserID and password. Similar functions are also present in Macintosh, Unix or Linux systems.

The bottom line is that being recognized (identified) has become mandatory.

UserID and Password are the two little words necessary for this purpose.

The UserID corresponds to what we once referred to as an alias (or a Nickname): since the use of one's whole name and surname might prove to be too long (Andrea Maria De Rosa), we prefer to use an alias (Andy or Andrea35) to notify the system of our user identity.

However, the UserID is public; in the sense that anyone can have access to it, since it is clearly indicated on all documents.

Therefore, anyone who knows that my UserID is "Andy", could try to use it, thereby attempting to "forge my identity". Fortunately, the system is not content with the UserID alone for identification purposes: it also asks the user Andy to enter a Password which, at least in theory, only the real "Andy" will know.

So, the Password corresponds to a key/word, known only to the original Andy (possibly because he made it up) and, consequently, if entered together with the UserId (i.e. Andy), will be capable of convincing the system that the user logging in is really Andrea Maria De Rosa.

Losing or giving others our Password means almost giving permission to "impersonate" us (i.e. pretending to be us) when using IT systems.

The aim of the instructions that follow is not only to help us understand how to best choose our own Password, but also which Passwords should never be used, being too simple to guess.

Bad Passwords allow doors to be opened

A bad Password is any Password that can easily be guessed (or decoded).
A term that many have come to know over the last few years (with the help of the press and movies) is "hacker": i.e. a person equipped with considerable IT skills, capable of forcing the electronic barriers of banks or other companies.
Often, however, reality goes far behind from movie-type scenarios.
Hackers create software which automatically generates thousand of possible Password and tests them against the system to force it.
Not only: one can find software containing thousands of unoriginal, "ready-to-use" Passwords: these Passwords can be tried out in a very fast sequence.
The diffusion of broad band connections, in a sense, makes things even worse: hackers will have plenty of band (speed) to test hundreds or even thousands of password in a very short time.
So, where does this take us?
It takes us to the need to generate, and use, original Passwords, in order to maintain the security of the systems at the highest level (starting from our e-mails, to make a useful example).

Passwords to avoid

We will start with Passwords that should (read must) be avoided (Table 1).
Some examples are one's own name, the name of one's partner or pets; other inappropriate Passwords are those names written backwards or names taken from cartoons, video games, film titles, book titles, etc.
From this viewpoint, even words that are too short should be considered inadequate since they are easy to decode (never use a Password consisting of only one letter!). Furthermore, one's own telephone number or even one's favourite food or drink have a low security threshold because they are easily traceable to the person using them.

Understanding the importance of Passwords, many Web Services Providers suggest what type of password to use for different services helping the most appropriate choice in the selection of access words; they do so even rejecting Passwords that show a very low security profile.
In fact, other Web Services force the choice of the password requesting the entry of both letters and numbers or the entry of special characters like a dollar symbol ($) or an asterisk (*).

INAPPROPRIATE PASSWORDS
One's own name, the name of one wife/husband, girl-friend/boyfriend
The name of one's pets or children
The name of friends or work colleagues
The name of your favourite comic book superhero
The name of your boss
The name of any other person
The name of the operating system that you are using
The name of the computer that you are using
Your telephone number
Parts of your social welfare card, the code shown on your employment card or any other document in your posses-sion
Any birth date
Any information that can easily be obtained from your bank account
Names (even fantasy names) taken from books or films
Foreign dictionary words
Passwords all consisting of the same letter
Names of places or nouns
Passwords created using the sequence of keys on your keyboard (as asdfg or qwert).
Any of the above, written backwards
Passwords consisting of only one letter

Table 1
Passwords easy to guess

Good passwords keep the door closed

A good Password is one that it is difficult to guess or discover.
The best Passwords are difficult to discover because:

- They contain both capital and small letters (and caps matters, because passwords are case-sensitive, and 's' is totally different from 'S').
- They contain numbers and/or punctuation characters.
- They can include some control characters and/or spaces.
- They are easy to remember meaning that you don't have to write them down somewhere
- They are longer than five or six characters (best choice: 8 characters)
- They can be written quickly so that no-one can guess at them from the way your fingers move on the keyboard

Here are two suggestions about how to create valid Passwords:

1. combine two short words, separating them with special characters or numbers
   – cotton6&thread or screws%4fingers
2. use the acronyms of sentences that are easy to remember
   – ATCARN (Air-tel CRFs are really nice)
   - ILATP (I Love Air-Tel Programmers).

Excessively long words don't help

Although some operators prefer to use Passwords consisting of more than 8 letters, these rarely improve security: if a Password consisting of 8 letters is chosen from among random letters and numbers, it becomes virtually impregnable, since it would be necessary to generate and try out at least 36^8 (2,821,109,907,456) possible combinations: at a pace of 1000 Passwords per second, it would take 89 years of uninterrupted attempts to exhaust all the possible combinations.
Using a good Password only takes us half way there in terms of the necessary security measures: the second, fundamental, step is to keep one's password a secret

Do not annotate passwords as such

In the film "War Games", a young high school student gets hold of the Password of the computer used by the school's secretary and changes his grades. He is suddenly an A student.
But how on earth did he manage to pull off this incredible hacking scoop? Simply by reading the secretary's post-it notes which she had left in her desk drawer with the access Password clearly written down. Believe it or not, this is a true story and applies to hundreds of other cases.
Hence the suggestion: DO NOT WRITE YOUR PASSWORD DOWN, just remember it. And, at this point, someone might say that we seem to be slightly exaggerating.

How will we manage? First you tell us that a Password should be complicated and include signs, numbers, small letters, capital letters and whatever else necessary and then you tell us not to write it down. How on earth will we be able to remember it after not using it for a few months?
And, above all, what about the fact that an average Internet user has from 3 to 15 passwords? How can the poor guy possibly remember 15 different cryptograms?
This explains why most users use "Pippo" or their own name – they simply try to make things easier and ensure that it will be easy to remember their passwords.
But one can certainly do better. Here's how.

How to annotate and manage numerous Passwords

The first point concerns the fact that, notwithstanding rare exceptions, it is possible to use the same Password for different UserIDs: in fact, in principle, it would be sufficient to use only one personal Password to manage access to one's own PC, access to the Internet, access to one's e-mail, etc.
Unfortunately, this involves several problems such as:

- Generally speaking, the administrator of a system could be able to access the usernames and Passwords of his Users: by only using one Password for all services, the system administrator of one service could also access all the other services that we use.
- Many IT systems place restrictions on the Passwords (or UserIDs) that can be used: for example, many Internet sites request that the symbols "%" or "&" are entered (thus helping hackers to guess possible access words since they know that at least one of the characters will be one of those imposed) or that the password contains at least seven letters or a number, etc.

On the other hand, if I use a separate Password for each different service, I will have to remember each one of them, meaning that noting them down will be inevitable.

How not to note down Passwords

The worst way to note down Passwords is to do it in such a way that anyone stealing (or glancing at) the piece of paper on which it has been noted down instantly knows:

1. to what service it refers (e.g. Mail)
2. what UserID is used (e.g. a.derosa@airtel.it)
3. the Password itself (e.g. "pippo")

At this point, anyone who has read the note can download and read all my mail (to say the least).

Separating information

A decidedly more efficient, secure system involves two different steps:

1. choose at least 2 (but 3 would be better) different Passwords, in growing order of complexity. This will provide us with a super-secret Password, to be used for all critical services, where security and privacy are fundamental (home-banking, mail, Clinical Studies, on-line purchases, etc.). Then we will have Password2, to be used for less critical services (access to Group Discussions, registration in mailing lists, sports or leisure-time associations, etc.). Finally, a Password3 for all situations in which it is not possible to guarantee the secrecy of the Password (Password3 is similar to a group Password).
2. Jot down the Password and the Password alone without explaining what it has to do with.
   For example:
   - `1: sfGth34&5`
   - `2: KaRTs21`
   - `3: pippo`
3. note down all the remaining information on another support or in another place with respect to 2).
   For example:
   - **Internet Access**
     `UserID:andy498`
     `Pwd:1`
   - **Mail at airontelematica.com**
     `UserId: a.derosa@airontelematica.com`
     `Pwd:1`
   - **Clinical Studies**
     `Address: clinic.aritel.it`
     `UserID:aderosa`
     `Pwd: 1`
   - **Forum on gliders**
     `Address: www.f3k.it`
     `UserId: andy`
     `Pwd: 2`
   - **Chess server**
     `Address: www.chess.com`
     `UserID: andy341`
     `Pwd: 2`
   - **Children's school** (pwd shared with the relevant child)
     `Address: www.isc.com`
     `UserID:derosa21`
     `Pwd: 3`

Managing only 3 Passwords, instead of 15 or 50, is definitely simpler. Furthermore, after a while, we will be able to remember all 3 of them by heart (until then, when necessary, we will probably have to take a peek at the note on which we have jotted down the Passwords).
Conversely, to remember the remaining information, we will have to look at the second note. However, since this does not contain any Password, we can keep it near the PC without having to always hide it. Another advantage of this method is that when, for example, we wish to change the Password, all we will have to do is to note down the new Password in the place of the old one without having to change the note containing all the remaining information (which remain the same, regardless of the Password used).

Password variations

Another Password management strategy is to have a basic Password which can be changed for each different service: e.g. the Password "kapiza" can be changed numerous times depending on its use: for e-mail use it can be changed to "kapizb", for access to Internet sites it can be further changed to "kapizc" etc.

Password rotation

Many people use the Password rotation method to solve security problems.
This method consists in changing one's Password every 2-3 months. This solution, per se, does not do much to increase security. Furthermore, it is necessary to add that the use of this strategy imposes the need to remember all the numerous Passwords used by heart, otherwise the idea of continuously changing one's Password would only create problems in the use of services.

A Password-managing software

The most radical solution (suitable for those who have lots of Passwords and who are familiar with IT) is the use of special software-registers (which can be accessed, needless to say, by means of a Password: the only one that we will have to remember by heart).
The Password Keeper strategy is currently used in the most recent versions of Netscape Navigator and Microsoft Internet Explorer, by means of Gator software, or by means of numerous other types of autonomous software, created especially for this type of management (e.g. SplashID).

Changing type of character

It is also possible to create a personal Password archive using a simple electronic text sheet and then modifying the content entered, replacing it with special characters such as Wingding or Marlet (found on every PC), as long as one remembers what type of character was used to begin with.

Restriction on Password sharing

Be careful about sharing your Password with others.
By analogy, giving your Password to someone is like giving them the keys to your house or office. This means that we have to know the person really well because he could enter our house and literally "clean us out".
When sharing a Password with someone, it is important that they understand the responsibility involved.

Changing/eliminating Passwords when a user no longer needs them

If one shares the keys to one's house, one runs the risk of having the keys copied. To avoid this, one generally uses keys that are impossible to duplicate. However, when the people using the keys no longer need to use those accesses, it is safer to change the locks. Changing locks can be expensive both in terms of time and money. In IT, these two inconveniences do not exist meaning that when one or more users are no longer enabled to access specific subareas or use specific PCs, their Passwords must be eliminated or changed.

The above article is a free derivation from S. Garfinkel's book (6) to which we refer you for further information on this topic..

APPENDIX 2

e-CRF

Top-notch technology, Know-how and 100% Reliability for Discerning Sponsors and Cro's

- Full compliance with all FDA, EMEA and Privacy dispositions
- Web-based accessibility: runs from anywhere
- Pure HTML pages: runs on any PC or Device
- Fast and user-friendly
- Accepts in-depth data when needed (step-up levels)
- SAE/ADE modular reporting system
- Sms gateway to alert clinicians and patients
- Plausibility/consistence tools for easy 'data cleaning'
- Phone and immediate on-line assistance to Investigators
- Monitoring and query-generating tools
- Data safeguard (multi-backup strategy on hw/sw-protected server)
- Standard data export formatting (.mdb, .xls) for analysis

AIR TEL

10 YEARS

WEB-BASED TECHNOLOGY FOR CLINICAL RESEARCH - www.air-tel.it

+39 0260830041 – Milano - Italy