BIOINFORMATICS & PHARMACEUTICAL AUTOMATION

Coffey Analysis

Discovery Automation: The Benefits of 21 CFR Part 11 Compliance — Even if the FDA Never Asks for Your Records

Patrick Coffey, Ph.D.

pjcoffey@coffeyanalysis.com

1628 EUCLID AVENUE
BERKELEY CA 94709 USA
+1 (510) 849 0817

Introduction

With its release in 1997 of 21 CFR Part 11, the FDA specified its requirements for accepting electronic records in lieu of paper records. If you are in the process of automating a drug discovery process, you should be aware of the requirements of the Part 11.

Most companies have focused their Part 11 compliance efforts on manufacturing, on clinical trials, and on development, because that has been the focus of FDA auditing efforts. But it also makes good sense to design a discovery automation system to comply with Part 11, even though you may consider it unlikely that the FDA will every demand these records:

- The immediate goal of drug discovery is a patent filing. Compliance with Part 11 will ensure that you have full support for patents. You will be able to supply exact records supporting any filing or later patent dispute. The principal investigator, time of discovery, and methodology will be identified beyond any doubt.
- The requirements of Part 11 are all based on common sense and good workflow. They are really just a codification of good laboratory practice (GLP). Part 11 may be a regulatory process, but it is also a set of guidelines for improving the security and efficiency of your discovery-stage workflow.
- If you have the opportunity of considering Part 11 compliance during the system design process, the implementation requirements are not that onerous. It is much easier to build the system with Part 11 in mind than to deal with Part 11 remediation later.
- The dividing lines between the discovery and development stages are not clear-cut, and drug candidates often cycle between the two stages. If your discovery process is Part 11-compliant, the transition to Part 11-compliant development will be simpler.
- The FDA may decide in the future to further codify the requirements for drug discovery records. If it does so, it is likely that the requirements will follow the same standards as Part 11.

In general, it is a good idea to design any new automation system to comply with the regulations of Part 11. As this paper will show, it is not that difficult to design your system with Part 11 compliance in mind.

This paper describes an approach to designing a drug discovery system that is Part 11-compliant. The ideas described have been implemented in Coffey Analysis's Automation Explorer™. You may find this paper useful in reviewing your current systems, in designing a new system, or in evaluating the approaches suggested by software and hardware vendors.

Part 11 describes the FDA's requirements for acceptable electronic records but does not require a particular implementation. In addition, the FDA objects to describing software as "complying with Part 11". The FDA views the company and its principal investigators as complying (or not) with Part 11. While this paper refers to software as "complying with Part 11" , please view that as shorthand for "software that assists the company and its investigators in complying with 21 CFR

Part 11", and please view the ideas on implementation as those of Coffey Analysis rather than the requirements of the FDA.

Goals behind Part 11

When electronic records are submitted, the FDA wants to be able to ensure the following:

- That the information is complete – that each record can be tracked to its source and that related records are connected.
- That the time of information entry and of any modification is noted.
- That the information has not been altered in a manner that obscures the original information (much as is done with laboratory notebooks, where corrections are signed and where incorrect information is never deleted, merely amended).
- That only authorized personnel can access the system, and the person creating, modifying, or reviewing any information is identified and personally attests to the validity of the process.
- That all experimental protocols are clearly documented and identified.
- That all personnel are trained in the protocols and that documentation is available.
- That the system has been validated to operate correctly.
- That information can be viewed in either electronic or human-readable format and that records be available "in a timely fashion".
- That the system be available for inspection, including both hardware and software.

Is there anything in this list that you would not want in your automation system?

A Typical Automation System

A typical discovery-stage automation system will include

- Manual processes (pipetting, weighing, etc.) with protocols defined and documented for the processes.
- Automated processes that employ laboratory instruments, with protocols defined and documented for the processes.
- Software that controls the individual instruments and generates data and reports. This is usually commercial software supplied by the instruments' manufacturers.
- Decisions and assessments made by scientists based on information generated by the processes, often based upon external third-party software (e.g., protein sequencing, computer modeling).
- A central relational database that is used to store the information generated in the processes.
- Automation software, often called LIMS (Laboratory Information Management Software) that pulls all of this together. This software manages the following:

◊ Prints unique identifiers for all objects (e.g. plates, vials) and tracks

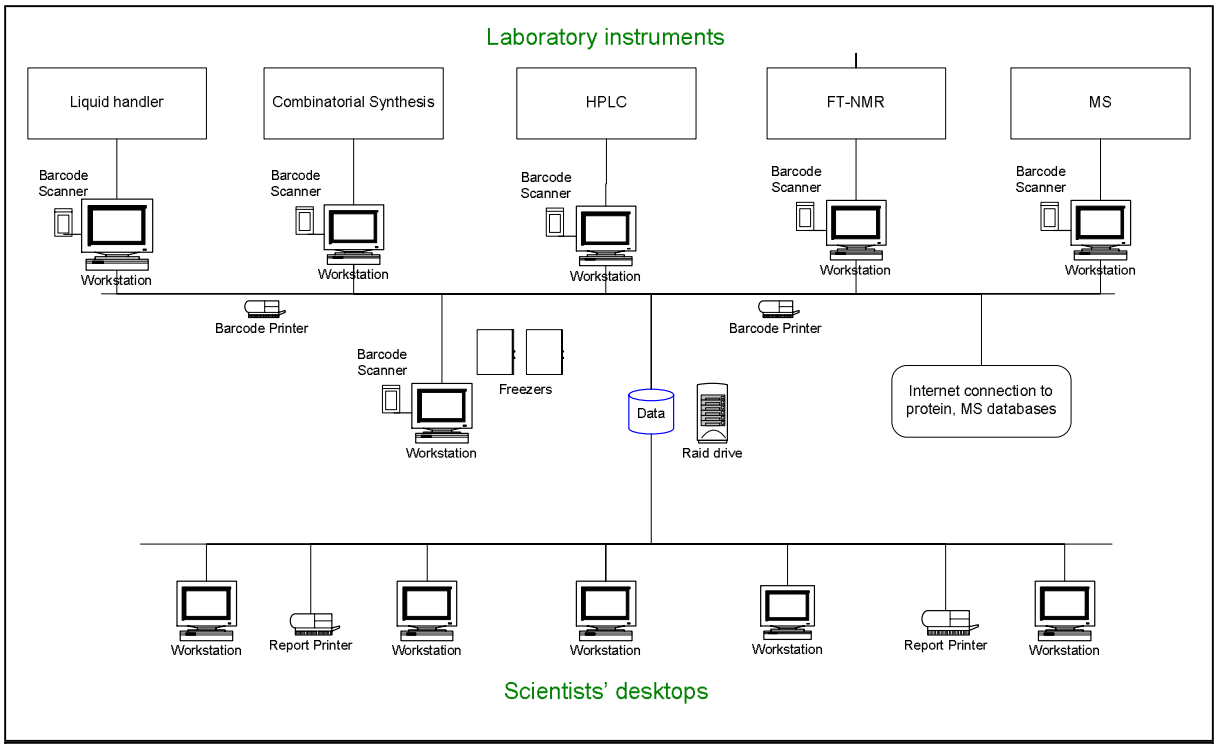


Figure 1. A typical discovery-stage automation system includes several laboratory instruments, freezers or other storage devices, and a central database for storing all the data. The instruments each have a workstation, usually supplied by the instrument manufacturer. Using automation software, the users initialize processes to be run on the instruments or to be run manually. The automation software then retrieves the results of the processes from the reports generated by the individual instruments and stores the results in the central database. The automation software provides visualization and reporting tools that can be run either from the instruments or from the scientists' desktops.

their storage and custody.

◊ Allows operators to initiate manual and automated processes.

◊ Updates the central database with the results of manual processes.

◊ Insofar as possible, sets up the input files for the automated processes, thereby minimizing data entry errors.

◊ Imports the information generated by the automated processes, and enters the results into the central database.

◊ Provides visualization and reporting tools to connect all the information in an intelligible manner.

Open or Closed?

Part 11 differentiates between two types of systems:

- Closed systems, where system access is controlled by persons responsible for the content of the records.
- Open systems, which include everything else.

The requirements for open systems are more stringent than those for closed sys-

tems in one aspect: In addition to all of the controls required for closed systems, open systems require "procedures to ensure the authenticity, integrity, and, as appropriate, the confidentiality of electronic records."

Is your system open or closed?

At first glance, it may seem that it is closed. Only your personnel work on the system, and they all have passwords to ensure that they have the proper permissions. But in practice almost all systems are open:

- Does anyone ever use someone else's password when running your system? If so, your system is open – system access has not been controlled.
- Are you using public internet databases or programs as part of your system? In that case, people at that web site, whom you do not know, have access to part of your system. Your system is open.

For a discovery system, where support of patent filings, it is very important that you are able to authenticate the operator who performed each experiment. If you treat your system as an open system, you will be able to provide this authentication

So this paper proceeds with the assumption that you have decided that your system is an open system. Fortunately, it is not that difficult to comply with the added requirements of an open system, and an open system in fact provides a much better security model.

"A Timely Fashion"

The FDA requires that you be able to provide requested information "in a timely fashion". Whether or not the FDA ever asks for your records, this is still important — in fact, it is the ultimate goal behind automation software.

Information falls into different categories:

- Much of the information that will be available directly through the user interface of your automation software – whether compound X was purified, on what date, by whom, etc. This is the information that you regularly obtain for your own purposes.
- Some requests may require information that you do not usually need– show all samples purified on a particular HPLC between certain dates, for example. Answering these requests might require custom queries against the central database. These might require a few hours to produce.
- Other requests – show all changes to certain database records, with before/after snapshots – might require the use of third-party tools like Entegra™ (see "Tracking Changes" below).
- A request for training or documentation compliance – e.g., show that all operators were notified of an update to a particular synthetic protocol, for example – might require you to dig up backup email tapes and print the appropriate emails. This might require a few days.

This is fine – all of these responses represent response in "a timely fashion". You should design the user interface only to provide the information that you need for your own purposes. You will not want to

clutter the user interface with buttons to produce everything the FDA (or anyone else) might conceivably ask for. Extra buttons and controls in the user interface will only confuse the normal user. Just make sure that all information is either in the database and can be extracted upon request, or that it is in a known location (like email backup tapes).

Choice of a database system

You will want to use a sophisticated database system with a strong security model. At present, the leading contenders in this category are Oracle, Microsoft SQL Server, and IBM's DB2. The choice really depends on your current use of one or more of these products and which third-party tools you might want to employ for document control and database logging. Any of these database systems will be more than adequate for the job.

Do not use Microsoft Access or other low-end database systems. While these are fine products for single-user needs, Access has a number of disadvantages that make it unsuitable for use as a central automation database:

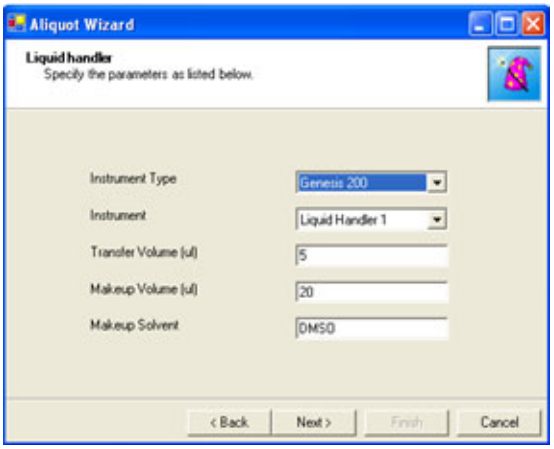
- It does not have a strong security system, which will make it difficult to implement the security and authentication portions of Part 11.
- It is "single-threaded". This means that if one user has started an Access operation, another user will not be able to start an operation until the first operation completes.
- It is limited in database size, currently to 2GB. If you store instrument data files in the database, as this paper recommends, you will quickly run out of space.

What should go into the database?

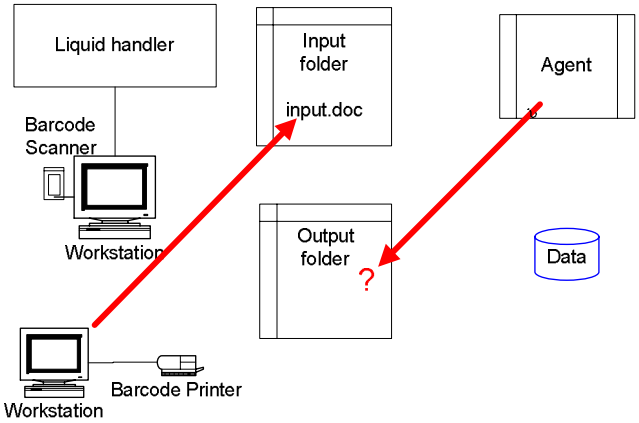
Everything that happened should be included, but only those things that really did happen. Make sure your system is "closed-loop", i.e. that only data actually generated by the instruments or attested to by the operators is included. For example, just because someone sets up an automated liquid handler to aliquot one plate into several daughter plates, your should not assume that the daughter plates were actually produced — perhaps the operator dropped the plate on the floor before the process completed. Wait for the report to be generated by the liquid handler before entering the records into the database. Similarly, make sure your system is "fail-safe", i.e. that it does not include partial data, that it rolls back transactions in the event of a crash or error, and that it resumes operation in the event of a shut-down or crash.

The database should include the records showing procedurally what happened, but should also include all the raw data files from the instruments and all the protocol description documents. If you try to save these documents outside the database, you may find that they have become disconnected or lost by the time you want the information. Disk space is almost free now, and it is just a good idea to store everything where you will be able to find it later. By storing all ancillary files in the database, you will also be able to use those files to support a patent position, since you can demonstrate that the instrument file has not been altered after its storage date, which is not the case with

Importing Data



1. The operator goes to any workstation and starts the Aliquot Wizard intending to aliquot a mother plate into two daughter plates. The Aliquot Wizard asks the operator a few questions (top) and then prepares an input file for the liquid handler's input folder (input.doc in the diagram) and also prints barcode labels for the daughter plates. The Agent then begins looking for the output files from the instrument in the Output folder.



2. The operator goes to the liquid handler with the mother plate and the empty daughter plates after applying the new barcode labels. Using the input.doc file, the operator starts operation of the liquid handler. When the liquid handler completes its operation, it writes the results in its output folder (output.doc in the diagram). The Agent finds the file, extracts the information it needs, and stores the information in the central database.

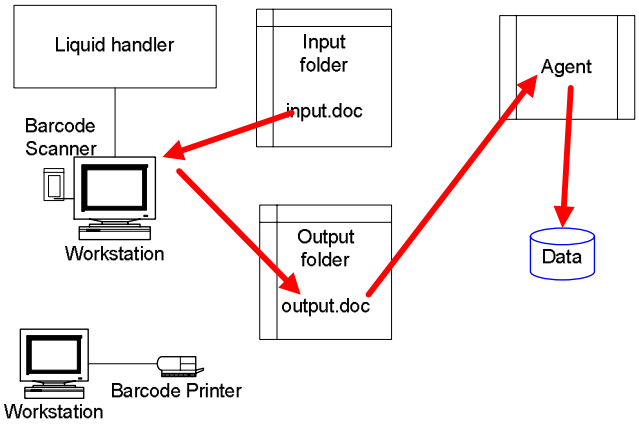


Figure 2. Importing data using a Wizard to reduce input errors. The import is fail-safe (data is not entered until the instrument's reports are generated) and fail-safe (in the event of a crash or other failure, the data is rolled back and entered later).

files that are stored externally.

Eliminating Errors with Wizards

When an operator is starting an automated operation, it is important to give him or her as few opportunities for error as possible. The use of step-by-step "wizards" that ask all the necessary questions and then prepare the input file for the instrument will help to minimize errors. (Figure 2).

How Does the Information Get into the Database?

How can the system be made both closed-loop and fail-safe?

Closed-loop: When a new process is initialized, a wizard sets up the input files for the instrument that will run the process, prints barcode labels for any new plates or vials that will be generated, and then starts up a separate software process (an "agent") that does nothing but wait for the files generated by the instrument to appear. Nothing should go into the database until those files are found. So the operation is closed-loop — information goes into the database only when the instrument confirms that the operation has been completed. If the operator forgets to initiates the process or drops the sample plate on the floor, the database is not corrupted with things that did not happen.

Fail-safe: The database system uses a technology called "transaction processing" that keeps incomplete or garbled information from being imported. This is the same technology used in ATM machines to make sure that your bank account is not corrupted. If the ATM's cash dispenser jams before your money is dispensed, the withdrawal from your account is automatically rolled back. The database import works the same way — if the system crashes before all the data is imported from the instrument, it rolls back the entire data import process and tries again when the system is brought back up. The files from the instrument are still there (they are not removed until the transaction is completed), and the software agent that is looking for them will find them waiting the next time it looks. (Figure 2).

Protocols, Documentation, and Training

Software is only part of Part 11. It is important that you have in place a program of documentation for all procedures. For each process, you must document exactly the procedures that must be followed in a *protocol*.

These protocols must be under a revision control system (Figure 3). That means that if you change a protocol after you put it into practice, the old protocol must be preserved so that it is possible to see the con-

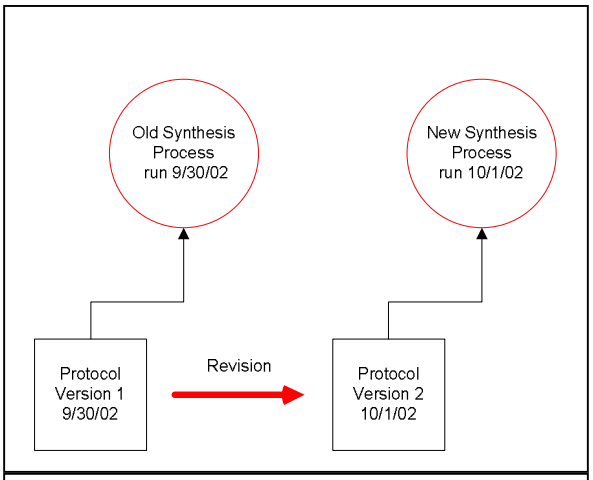


Figure 3. When a protocol is revised, the old versions of the protocol must be preserved, since they describe the procedures under which pre-existing processes were run.

ditions under which prior experiments were run. Each application of a protocol must reference not only the protocol but the particular revision of that protocol.

To be compliant with Part 11, you must institute a program that trains personnel in the application of each protocol for which they might be responsible. Notifications must be given to all operators for any updates to the protocols. Make sure that the records of the training sessions and the notifications are preserved.

Documentation for the protocols must be readily available, and all operators must be familiar with the location and contents of the protocols.

Protocols, documentation, and training need not be computerized, but doing so can make compliance a good deal easier. Here are some suggestions:

- Make sure that all automation records reference the protocol revision under which the process was conducted (Figure 3).
- Make sure that your automation software either includes its own protocol revision control system or references a reliable external system.
- Email all authorized operators for every revision of a protocol, attaching the revised protocol to the email. Make it a policy that they acknowledge receipt of the attached protocol. This email notification may be able to be automated as part of the revision control system. Keep backup records of all emails.
- Keep training records on a spreadsheet (or better still in a table within the central database), noting attendees, dates, and subject matter.
- Make all the documentation available online, so that there is never any question about its location. This also ensures that only authorized operators can view the documentation, which is one interpretation of the Part 11 regulations.

Timestamps

One of the FDA's requirements is that the system should use "secure, computer-generated, time-stamped audit trails". This is relatively easy to do. Just add a *timestamp* field to each record when designing the database system. Your database system will automatically enter the time when each record was created or modified.

Whose clock?

When time-stamping records, where do you get the time? Different records in the database may have been generated on different computers, so just using the clock on each computer without a synchronization effort will not work. Computer clocks can drift over a period of a few days. For patent filings, reliable time and date information is critical.

Your system should either reference a single server clock from all workstations or have the server periodically set the time for all workstations.. This is a good practice in general. But how do you set the server time accurately?

The National Institute of Standards (NIST) provides internet access to its clock. There are a number of third-party programs that will synchronize any com-

puter's clock to the NIST clock. One that is free is available at http://www.40tude.com/time/lite/index.htm.

Tracking changes

One of the most difficult requirements of Part 11 is to ensure that "record changes shall not obscure previously recorded information." The timestamp will show you *when* a record was modified but will not show the previous content of the record, nor will it show you who modified it. If the record has been deleted, the timestamp will not help at all. It is relatively easy to enforce rules on modifying or deleting data from within your own user interface and to track the before/after snapshots of any changes. It is impossible however to prevent someone who has administrator privileges for that database from modifying or deleting records from outside your software. In that case, you will be able to spot that the record was changed from the timestamp (if the record still exists), but will need to find a way to determine the before/after details on changes and deletions.

Fortunately, third-party tools are emerging that are designed to do exactly this, priced from around $5,000 (Figure 4). A product named Entegra™ from Lumigent is designed to solve this problem and is described at http://www.lumigent.com/Solutions/papers.htm. If you use such a tool, your own automation software can be made substantially simpler, because it no longer needs to track these modifications – the third party tool takes care of that.

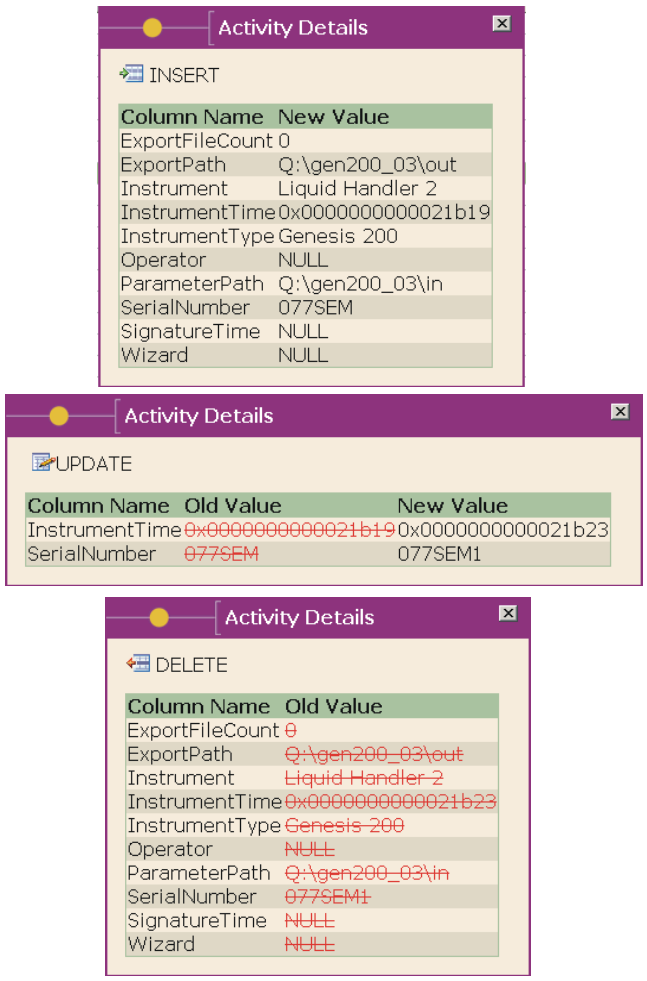


Figure 4. Lumigent's Entegra records data changes automatically. This figure shows Entegra's display of the insertion of a new record to the Instruments table, its subsequent update by changing the instrument's serial number (note that the change to the timestamp of the record in the InstrumentTime field was also picked up), and its ultimate deletion.

Access to the System

You need to ensure that only authorized personnel are able to access your software. Some of this control comes from the password security provided by your operating system. Follow the usual security precautions – each user should have an individual password, passwords should be complex, passwords should be changed regularly, etc. Your systems administrator should already be doing this.

A more complex question is how to ensure that people are only doing tasks for which they are trained and authorized. There are a number of different approaches to this problem, but the one I suggest you consider is "application role security". When users log onto your automation software, they will do so in a particular role and will need to know the password for that role. The roles themselves are defined by the database security system, but what choices you allow the different roles can be controlled directly from within your software. For example, someone with an "Operator" role might not be allowed to assign protein sequences, and only someone with a "Manager" role might be able to change a protocol. Unless a user knows the role password, he or she cannot take on the tasks specific to that role. Again, change role passwords often and make them complex.

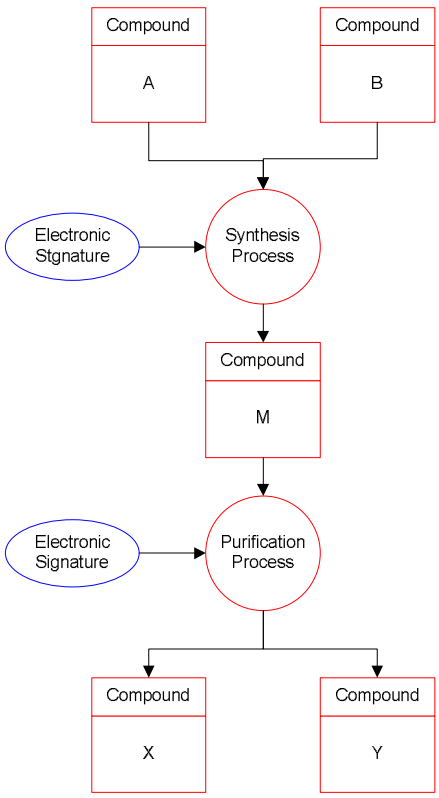


Figure 5. A simplified diagram showing two processes: a synthesis process that takes compounds A and B and produces a mixture M, and a purification process that purifies M into the reactants X and Y. For each of the processes, a digital signature is attached.

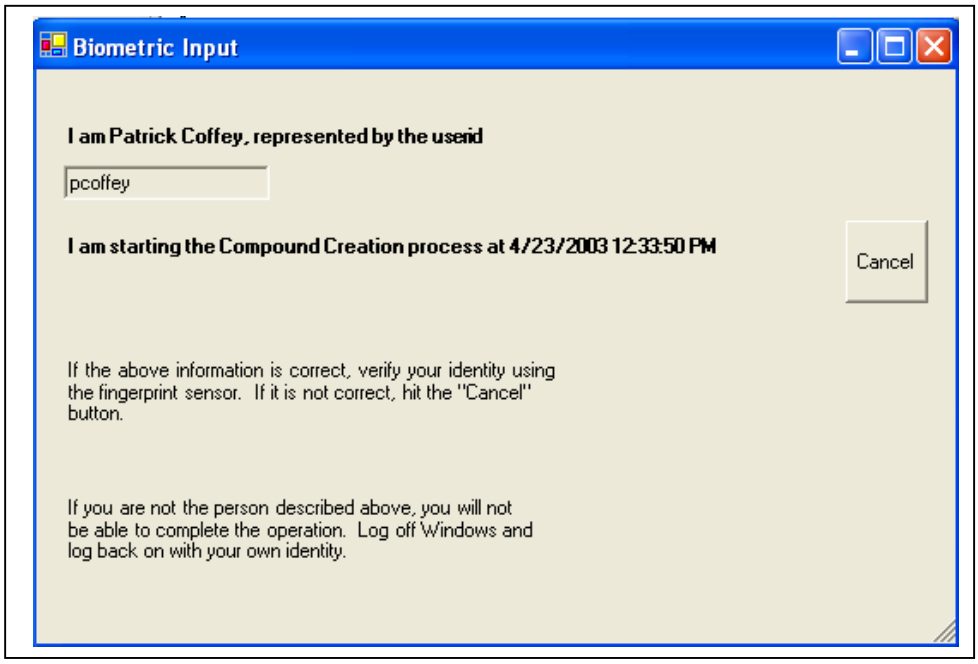
Who did it? – Electronic Signatures

"Persons who use open systems … shall employ procedures and control designed to ensure the authenticity, integrity, and as appropriate, the confidentiality of electronic records .."

For discovery purposes, confidentiality is not necessary, so I will not discuss that any further. For each process, the operator or scientist must sign the record. The FDA wants to know who is responsible and then be able to check that that person is authorized and trained for the process. In a patent dispute, you might also need unambiguous proof that the person whose name is associated with an experiment or process is the person actually responsible.

How can you create an electronic signature? When a user begins a process (say a

compound creation operation), he or she can be presented with a form that looks something like the example below.



(Figure 6) If Patrick Coffey must authenticate himself using a fingerprint, there is no risk of another user imitating him. The FDA does not require biometric signatures for Part 11 compliance, but makes it much simpler to comply with its requirements for authentication if you do choose to use biometrics signatures.

When the user agrees that this is correct, this information is stored in a table in the database. Once the process is completed, the signature is linked to the process.

There is a problem that needs to be addressed however: How sure are we that the person who says he is Patrick Coffey really is that person? We know that someone who knew Patrick Coffey's password started this software, but he might have gone to lunch and someone else might now be running it. Furthermore, Patrick Coffey might not have been the person who actually even logged onto the computer — he might have given his password to someone else. *Solution: Biometric input is the easiest way to solve this problem. Low-cost (around $150) fingerprint authentication devices are now available that simply plug into a USB port or that are built into a computer mouse or a keyboard*

Two similar terms are in use: digital signatures and electronic signatures. Digital signatures are incorporated directly into a document and employ a cryptographic technique for making sure a document has not been modified after it was signed.

Figure 6. Low-cost biometric authentication devices are now available from a number of sources. The picture shows the Digital Persona U.are.U Pro.

Validation Using the User Interface

The user interface can be designed to assist in validating the system. At Coffey Analysis, we use grids for each of the important chemical entities — chemical compounds, proteins, oligonucleotides, etc. -- in the system and a grid that shows the Processes that connect the chemical entities. The example below shows the chain that connects two compounds through three processes — *Data Import* from a spreadsheet, *Transfer* to a reaction flask, and *Synthesis* to form a reaction mixture M.

The paradigm is simple:

- Each compound in a separate vial appears as a separate line in the Compounds grid.
- Selecting a line in the Compounds grid displays all processes that connect to that compound.
- Selecting a line in the Processes grid displays all compounds connected to that process.

Compounds

Compound	Vial	Weight	Volume
A	453001	48.6	128.3
B	453002	65.1	139.6

Processes

Process	Operator	Date and Time	Confirmed

1. The operator uses a filter (a way to select a subset of all the data) to fill the Compounds grid with the compounds recently imported. In this case there are two, A and B, each in separate barcoded vials 453001 and 453002. The vial number, weight and volume of each compound is displayed in the Compounds grid.

Compounds

Compound	Vial	Weight	Volume
A	453001	48.6	128.3
B	453002	65.1	139.6

Processes

Process	Operator	Date and Time	Confirmed
Data Import	JDoe	10/1/2002 9:34:03	True

2. Selecting compound A shows all the processes that that compound in that vial are associated with — in this case, just Data Import (the import of information about the compound from a spreadsheet).

Compounds

Compound	Vial	Weight	Volume
A	453001	0	0
B	453002	0	0

Processes

Process	Operator	Date and Time	Confirmed

3. The operator performs a process that transfers the contents of both vials 453001 and 453002 to a reaction flask with the label 453003 and adds 500 microliters of dilution solvent. When the operator runs the same filter as in step 1, the same compounds appear as in step 1, but with no weight or volume! What happened?

Figure 7. If the user interface can show the connections between chemical entities and processes, it can be much easier to validate the system. As each process is completed, it is easy to say whether the process is properly attached to all the objects and whether the mass and volume transfers have been correctly accounted for.

Validation Using the User Interface (cont).

Compounds

Compound	Vial	Weight	Volume
A	453001	0	0
B	453002	0	0

Processes

Process	Operator	Date and Time	Confirmed
Data Import	JDoe	10/1/2002 9:34:03	True
Transfer	JDoe	10/1/2002 9:36:21	True

4. Clicking on compound A shows two processes now — Both *Data Import* and *Transfer*. The "Confirmed" column in the Processes grid shows that the electronic signature is attached.

Compounds

Compound	Vial	Weight	Volume
A	453001	0	0
A	453003	48.6	767.9

Processes

Process	Operator	Date and Time	Confirmed
Data Import	JDoe	10/1/2002 9:34:03	True
Transfer	JDoe	10/1/2002 9:36:21	True

5. Clicking on the Transfer process shows all the compounds that are connected to that process. In this case, compound A has been transferred from vial 453001 to reaction flask 453003. Compound A appears twice, once in the original vial and once in the flask. Compound B does not appear because it is not connected to this particular process -- there is a similar process transferring compound B to the flask. The volume shown is the sum of the volumes in vials 453001 and 453002 and the 500 microliter dilution volume.

Compounds

Compound	Vial	Weight	Volume
A	453003	0	767.9
B	453003	0	767.9
M	453003	113.7	767.9

Processes

Process	Operator	Date and Time	Confirmed
Transfer	JDoe	10/1/2002 9:36:21	True
Synthesis	JDoe	10/1/2002 10:45:1	True

6. The operator then ran a synthesis process on the contents of the flask. Selecting compound A in vial 453003 shows the Transfer and Synthesis processes, but not the Data Import process — that was associated with Vial 453001, not Vial 453003. At the completion of the process, compounds A and B have "disappeared" — no weight remains. But a new compound — an uncharacterized reaction mixture M — has appeared with the combined weight of A and B. Further processes might be used to characterize or purify M, and the chain of compound/process interactions would continue.

There primary use is electronic documents (say an NDA filing) rather than in database records. Biometric electronic signatures can be stored in the database in the same fashion as any other record and can be time-stamped. If it is suspected that the any of the associated records have been modified, the timestamp of the signature can be compared with the date within the signature document itself or with the timestamps of the associated processes. If these times do not correlate, there is a problem. Secondly, changes to the digital signature record are tracked in the same fashion as changes to any other record in the database. A digital signature record should never be modified or deleted once it is generated, and any changes or deletions would be immediately apparent by using a third-party audit tool such as Lumigent's Entegra.

Validation

Validation of the system is perhaps the

most important part of the process. A very good basic book on software development including validation (and that also applies to hardware and documentation validation) is "Under Pressure and On Time", Ed Sullivan, Microsoft Press, 2001. While the topic is too complex to handle in detail here, here are some key points:

- Validation includes all components of your system – hardware, software, and documentation.
- Validation is a continuing process – you must validate all components of the system as they change, and must periodically validate the hardware portions of your system.
- You will need a central reporting system to record bugs, hardware problems, service calls, suggestions for improvements, test results, etc. Everyone who is using the system should be encouraged to report problems or suggestions here, where they can be addressed by the project management team. There are a number of commercial products available (do a Google™ search on "issue management"), and you should purchase one of these rather than attempt to write your own. By using an issue management system, you can make sure that all issues and their resolutions are recorded. If you need to support your validation, it is all there.
- Once you are in production with your software, don't change it on the fly. Go to a release schedule. Each release should be validated to ensure that all subsystems that have been changed still function properly and that the documentation has been suitably modified to reflect the changes. Only when the release has been validated should you transfer it to production, and then you must notify and train your staff on the features of the new release.
- Try to build your user interface so that it simplifies validation (Figure 7). This makes system test much simpler.

Conclusion

The FDA's requirements under 21 CFR Part 11 are stringent, but they are based on common-sense needs to authenticate and review research efforts. Complying with them is in most cases just good practice. Using proper design techniques and appropriate third-party tools can make compliance much easier. Following the Part 11 guidelines in discovery-stage automation projects will certify your records for patent filings, will smooth the transition to development, and will very likely make it easier to comply with any future FDA regulations on discovery.